Smoke Detection Using Simplified Descriptors of Video Information

Gustavo Monte[1], Juan Ignacio Pastore[2,3], Virginia Ballarín[2], Damián Marasco[1], Pablo Liscovsky[1]

[1]Universidad Tecnológica Nacional, Facultad Regional del Neuquén, Argentina

[2]Universidad Nacional de Mar del Plata, Facultad de Ingeniería, Argentina

[3] CONICET, Argentina
Email: gmonte@frn.utn.edu.ar

Abstract— **Automatic visual detection of smoke in confined or open spaces is overriding to issue early warnings that can save lives or prevent irreparable damage. While fire presents a range of characteristic colour, smoke does not present a readily apparent pattern. Changes its shape, does not contain clear edges, presents a chaotic behaviour and colour manifests from white to black, including all nuances. This paper presents an algorithm that efficiently pre-process a frame that extracts the main component of information, decreasing orders of magnitude the source size. From this new structure, algorithms based on the temporal and spatial change of subsets of the new structure are applied. Decision is based on fusion of weak classifiers. The algorithms are described and validated with experimental results of real-time detection for open and confined spaces, considering simplicity and efficiency of the proposed method suitable for embedded systems.**

Keywords—smoke detection; real time; embedded systems; video processing; image representation.

I. Introduction

Video processing is increasingly present in automatic detection and firefighting systems. The availability of low-cost digital cameras coupled with the processing power of today's digital systems, creates the right conditions for developing embedded detection systems. In most real cases the smoke is the precursor of fire, therefore, detection of it is vital to infer the beginning of a fire.

Traditionally, smoke detection systems have been based on ionizing chambers that react with smoke particles or on infrared sensors, emitter and receiver, which detect the presence of particles in suspension. The weak point of these sensors is that detection is not achieved during early stages and they are not suitable for open spaces.

Fire-fighting systems are based on three fundamental aspects:

- Prevention
- Detection
- Extinction

The present work focuses on early detection of smoke in real time. It could be classified between prevention and detection. The main objective is to minimize the computational and economic cost to achieve massive applications. Image-based smoke sensors will replace conventional smoke sensors in short term. As algorithms, these can be executed on existing hardware, such as video surveillance systems, adding only the computational cost of software services associated with smoke detection.

In the literature, there are different approaches to smoke detection, but two lines of research stand out. The first one analyzes the chaotic behavior of smoke, its displacement and growth. In this approach, an estimation of the background is made, the most used method is GMM (Gaussian Mixture Model). This estimation is subtracted from the image and the chaotic behavior of the smoke is analyzed. Some techniques are contour tracking, chaos theory concepts texture analysis [1], and fluids mechanics [2], [3], mainly turbulent flow behavior. These processes are centered on the detection of smoke presence, considering its temporal and spatial characteristics.

The second approach is to observe the distortion of background image caused by presence of smoke, i.e. to track changes of the image produced by the smoke in an indirect way. In [4], [5], wavelets are used to quantify the loss of high frequency content due to the presence of smoke, added to temporal occurrence of background loss. To minimize false detections, sensory fusion is used. The work [6] combines the information of visible and infrared spectrum. In [7], the concept of integral image is used to achieve an accumulative model to estimate the motion orientation of smoke.

The presence of smoke, in its initial state degrades, edges of the background image and high frequency information. This temporal attenuation feature is widely used as a smoke indicator in its initial state since edge detection is a very mature technique. Another characteristic that identifies smoke is its movement, processed by means of optical flow. In [8] detection is employed using Lucas Kanade algorithm, combined with a backpropagation neural network to discriminate between different sources of optical flow. In [9] a very complete description of detection methods of fire and smoke more used are presented, highlighting their strengths and weaknesses.

Detection in open spaces is a challenge because of the scene variability and presence of disturbing factors such as wind, dust, movement, non-controllable and completely variable lighting conditions. Fire detection is relatively simpler than smoke since the range of colors it presents is confined to a subspace that together with spatial, temporal and texture characteristics, achieves a precise detection. Fire has shades between red, orange and yellow. For color detection, models that differentiate chrominance from luminance are used. These models achieve linear classifiers between pixels associated with fire and non-fire [10]. In joint detection of fire and smoke, fusion of descriptors is essential to reduce false positives.

The proposed method in the present work differs, from those in the state of the art, in the source information for algorithms execution. The input is based only on an essential information of a frame. This simplified version contains descriptors that minimize redundant data of a frame and provide information structure for inference of regions that belong to smoke. The frame descriptors generate an output that is orders of magnitude smaller with respect to the image size. Although the camera provides frames as input, it is possible to apply algorithms sequentially to rows on the fly, which is the way that a digital camera supplies pixels. Implemented in this mode, a complete frame is never stored in memory. The descriptors conform a compact structure that point out the type of occurrence and its position in the frame. Therefore, the proposed information structure is adaptable to run on embedded systems, in real time, with low-medium processing power, achieving a cost comparable with conventional detectors.

In the following sections the proposed algorithms, fusion and decision mechanisms together with experimental results are described. Finally, the conclusions and future work.

II. Description of Algorithms

A. Descriptors generation

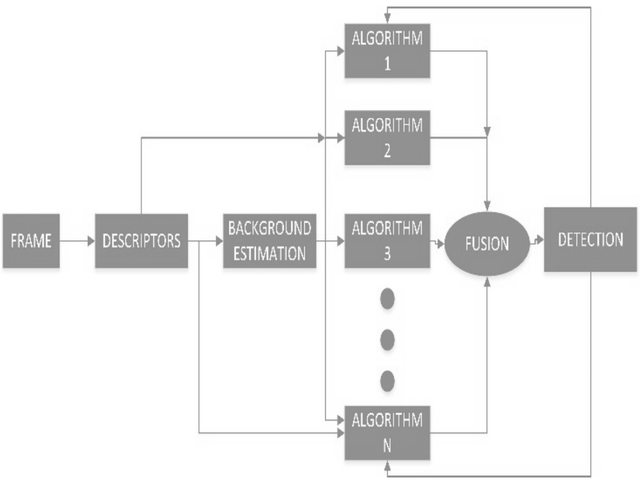

Fig.1. Proposed block diagram for detection. It merges the information of N processes, taking as input information the output of the descriptors and the estimation of the background.

Fig. 1 shows the proposed block diagram. On each frame the descriptors of the scene are obtained. The smoke presence inference processes are fed by the descriptors and background estimation, that is updated with each frame. The information present in an image is inferred by the value of the pixels and their spatial location. *The uniform sampling on a rectangular grid does not provide an efficient structure for extracting information*.

The information present in an image is not only in the value of its pixels and its spatial location, but in the way in which pixels relate to their neighbors. For example, if the value of a pixel can be obtained unambiguously from a linear combination with its neighbors, this pixel carries zero information. Dragging this pixel across all inference processes generates computational cost, memory space, and confusion. Considering applications in embedded systems and the way in which a digital camera communicates with the system, the detection of essential pixels can be computed observing the rows of the image. A pixel is considered essential if its value cannot be inferred from its neighbors. The other important part of the information, as it relates that pixel to the rest, is inferred by observing the path to the next essential pixel by tracking the subspace traversed. The trajectories are simplified to eight subspaces. The algorithms are described in [11], [12] and executed in real time, in our case as the digital camera sends information. Figure 2 shows the simplified signal trajectories.

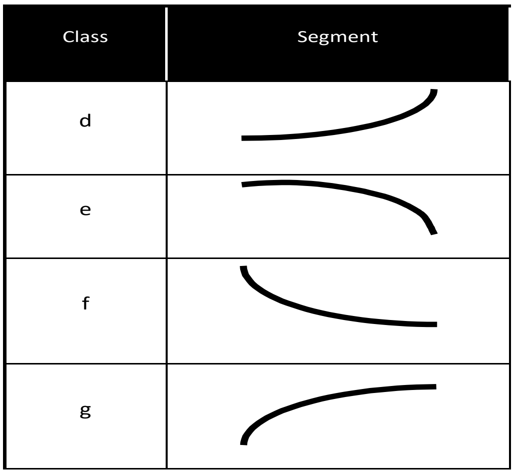

Fig. 2. Simplified trajectories in an image row between essentials pixels.

These algorithms were designed for unidimensional signals. In this application, an image is considered as an array of K unidimensional signals or rows. Since algorithms are applied to rows with a limited size, only segments type “d”, “e”, “f”, “g” are generated.

From essential pixels, the processes of extracting information are improved. For example, the union of a simplified ascending path with a descending one, marks the occurrence of a local maximum, and this maximum is obtained as the camera sends row pixels. The number of essential pixels in as subset of a frame is related with the “*activity*” of this region. If in a frame subset there is no essential pixels, there is nothing to process in that region.

This is detected during the frame acquisition and never stored in memory.

A frame of the image is now represented by:

- Value of the essential pixel.
- Location row, column.
- Path union with previous essential pixel.

The determination of an essential pixel is achieved by linear interpolation and controlled by an error bound. The size in bits of a frame is a function of the image scene and the interpolation error. For example, for a VGA image, 640 x 480 pixels, the descriptor are about 5 to 10 % of the frame size, depending on the scene and the interpolation error.

In Fig. 3, the marking of essential pixels for the occurrence of segment type “g” are shown for an indoor scene. It is important to note that it is not just an image segmentation, but it is inferred how essential pixels relate to the rest. To automatically adjust the sensitivity, the number of descriptors found is counted, and the sensitivity is adjusted until an adequate amount is achieved to detect the occurrence of smoke at an early stage. This amount is a function of the visual field. During sensitivity update state, the decision processes are suspended and the background estimation is updated. This setting is performed only when a minimum of occurrences is detected.

Fig.3.Indoor scene and essential pixels for interpolation error of 5%. Image of 640 x 480 pixels, gray levels.

B. *Processes for smoke inference*

The smoke inference process is performed based on a set of classifiers. Each classifier provides a clue, a necessary but not sufficient condition. A specific characteristic of smoke is detected with each classifier. The classifiers are not only sensitive to smoke but other patterns, and all of them become active in presence of smoke. The outcomes fusion of them generates a reliable detection. Some algorithms, as we will see later are exclusive, i.e. necessary for positive detection. Two processes are not related to the presence of smoke but with the validation of a frame. The first one is the total number of essential pixels of a frame. This value is input to the algorithm for obtaining essential pixels, controlling the error bound of the linear interpolator. In this way, the processes are adapted to the scene and considers variations of illumination and uncontrolled changes. The second process calculates the difference of essential pixels between two consecutive frames. If the change is greater than a threshold, all inference processes adopt a reset state.

It is important to note that these two simple processes, which base their result on essential pixel count, generate an adaptive interpolation scheme depending on the scene and detect considerable changes in a scene to reset the inference processes and start over. These two processes would have required a large computing power from pixels on a rectangular grid.

The algorithms will work on sub images; therefore, the first step is to divide a frame in M x N regions. The information in each subset are the essential pixels and the signal trajectory between them. The proposed algorithms that detect a specific feature of a smoke region are the following:

1. Background mean estimation of regions.
2. Relative change in the average value of the descriptors with respect to the estimated background.
3. Decrease in the high frequency content of the region by observing the decrease of descriptors variance.
4. Change in the number of descriptors in the region.
5. Intra frame persistence.
6. Agglutination of neighbor regions.
7. Spatial distribution of essential pixels inside the region.

Below, the processes are described highlighting the part of the smoke dynamic for which they have sensitivity.

B1. Process 1.

The background mean estimation is obtained by averaging up to k frames using recursion as shown in Eq. (1).

$$\tilde{f}_{i,j,n} = \frac{\tilde{f}_{i,j,n-1} * n + I_{i,j,n}}{n+1} \quad (1)$$

For $n=1, 2, 3,\ldots k$; $i=1, 2, 3,\ldots M$ and $j=1, 2 ,3;\ldots N$

Where $\boldsymbol{f_{i,j,n}}$ is the mean background estimation for region ***i, j*** at time ***n***, and $\boldsymbol{I_{i,j,n}}$ is the mean of descriptors at region ***i, j*** at time ***n***. This estimation is achieved by storing only two frames of descriptors. When ***n*** reaches ***k***, the estimation starts over from a new estimation.

B2. Process 2.

This process computes the change in the average value of the descriptors with respect to the estimated background, relative to the descriptors mean in a region, Eq. (2).

$$df_{i,j} = \sum \left| \ (\boldsymbol{f_{i,j}} - \boldsymbol{I_{i,j}}) \right| / I_{i,j} \quad (2)$$

Where $\boldsymbol{I_{i,j}}$ is the mean of descriptors that belongs to región ***i, j,*** and $\boldsymbol{f_{i,j}}$ is the background mean estimation of region ***i, j***.

Fig. 4 shows the signal generated over 200 frames for this process in smoke and smokeless regions.

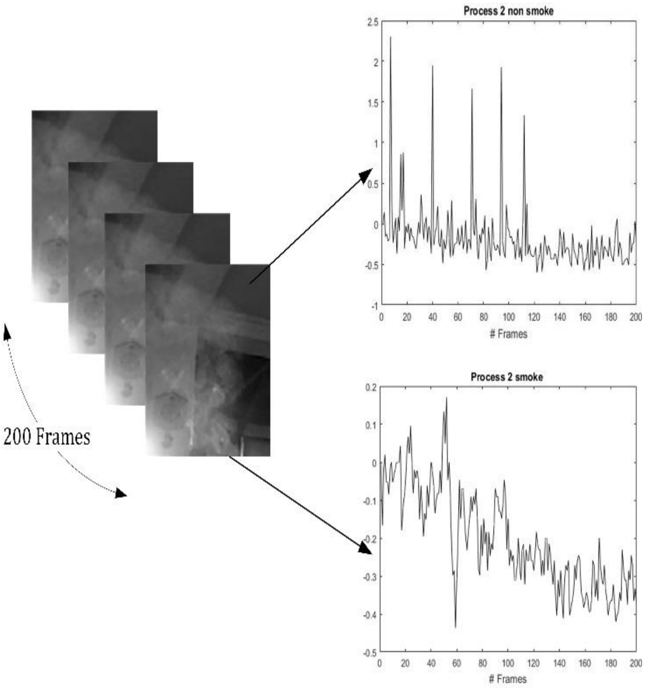

Fig.4. Example of signal evolution generated by process 2 for a smoke and smokeless regions.

B3. Process 3.

The descriptors variance of each region is computed and used as an inference of the high frequency content in the region. The change is computed with respect to the background variance, Eq. (4). This process is only calculated in indoor scenes. It captures the accumulation of smoke that occurs in the areas near the ceiling.

$$dv_{i,j} = VAR\left(f_{i,j}\right) - VAR(I_{i,j}) \quad (4)$$

For $i=1, 2, 3,\ldots,M$ and $j=1, 2 ,3,\ldots,N$.

Fig. 5 shows the time evolution over 45 seconds, 1350 frames at 30 FPS, of the essential pixel variance of a region near the ceiling, which begins to flood with smoke. Note the slow decrease in the average variance value.

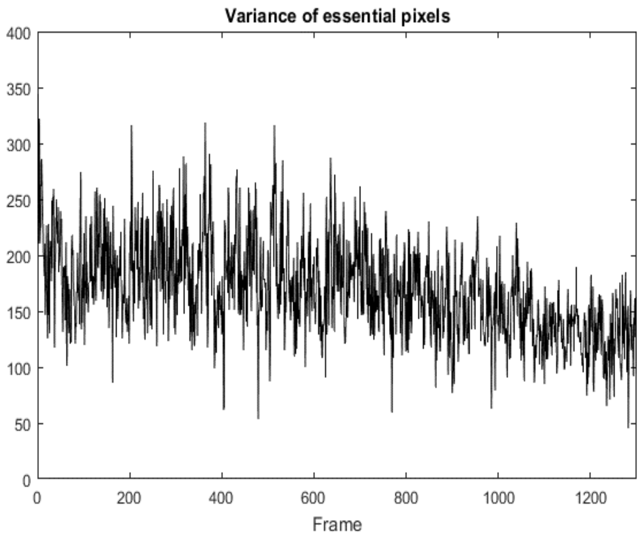

Fig.5. Time evolution for the essential pixel variance in an indoor scene region which evidences the correlation of the decrease of variance with the accumulation of smoke for a region near the ceiling. Elapsed time: 45 seconds at 30 FPS.

B4. Process 4.

The number of descriptors in all regions are calculated. This quantity provides an indicator of the region activity and the difference with respect to the background provides inference of temporal changes, Eq. (5). This process has two important objectives. First, if the number of descriptors in the region is less than a threshold, this region is not considered in the fusion processes. Second, the sum of all frame descriptors provides information to trigger adaptive processes for obtaining essential pixels.

$$dc_{i,j} = \left(pcF_{i,j} - pcI_{i,j}\right) \quad (5)$$

Where $pcF_{i,j}$ is the pixel count for background estimation of region i,j and $pcI_{i,j}$ is the pixel count for current frame. This feedback is important especially in open spaces, since adjusting the interpolation error to obtain essential pixels achieves a number of descriptors bounded at an interval independent of the scene.

B5. Process 5.

The overall smoke inference process is based on the convergence of observations in individual regions. Processes 5 and 6 measure how this information is aggregated to determine the presence of smoke. The regions take three states based on processes outcomes: "Neutral", "Alert" and "Smoke". Regions that exceed a fused threshold established for processes 2,3,4 and 7 take the "Alert" state. Process 5 quantifies the persistence of "Alert" regions as a quotient between the number of frames that the region was on alert and the number of observed frames, Eq. (6).

$$pt_{i,j} = \frac{A_{i,j}}{H} \tag{6}$$

For *i=1, 2, 3,…,M* and *j=1, 2 ,3,…,N.*

Where $A_{i,j}$ is the number of times that region *i,j* was on alert in the last **H** frames.

B6. Process 6.

To quantify spatial agglutination, the number of neighboring regions labeled as "Alert" relative to all neighbors is computed. It is considered eight possible neighboring regions including those located in diagonal, Eq. (7). Fig 6 shows the neighbors for a region *i, j.*

$$pe_{i,j} = \frac{AL_{i,j}}{8} \tag{7}$$

For *i=1, 2, 3,…,M* and *j=1, 2 ,3,…,N.*

Where $\boldsymbol{AL}_{i,j}$ is the number of regions labeled with "Alert" that are neighbors to the region *i, j*.

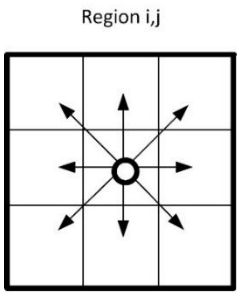

Fig.6. Neighbors considered for process 6. If a neighbor is in “Alert” state takes value one else takes zero.

B7. Process 7.

This process captures the smoke randomness during its time evolution. This is not computed for every region but on “Alert” ones. The region is subdivided into 16 blocks and the number of essential pixels for each one is calculated. Then, each number of pixels is multiplied by a coefficient which is the prime number represented by the sub-block position. The sum identifies that region at that moment. Fig. 7 shows the weighting coefficients, which are the first 16 prime numbers.

Block i,j

1	3	5	7
11	13	17	19
23	29	31	37
41	43	47	53

Fig.7. Weighting coefficients a(k), prime numbers, for the number of essential pixels in each sub-block.

The process outcome is computed as:

$$G_{i,j} = \sum_{k=1}^{16} a(k)\, pc(k) \tag{8}$$

For all regions *i=1, 2, 3,…,M* and *j=1, 2 ,3,…,N in “Alert” state.*

Where *a(k)* are the prime coefficients and *pc(k)* is the pixel count for each sub-block.

The value $\boldsymbol{G_{i,j}}$ captures the motion dynamics of region i, j evaluating the intra-frame progression. This process is sensitive during the beginning of the smoke, and merged with the other processes, can locate the beginning of the incident.

C. Fusion of processes for smoke inference

The inference of smoke presence is obtained by analyzing together the processes results. The regions first adopt a state of “Alert”, and then observe their behavior of spatial and intra-frame distribution to adopt the state of “Smoke”.

The acquired frame is validated for subsequent processes. If the total number of essential pixels that change from one frame to the next exceeds a threshold, all processes are reset by losing their state memory.

Process 1 is the background estimation. For process 2, Eq. (9) must be satisfied. This implies that there is a small change in the mean value of the region.

$$\delta 1 < df_{i,j} < \delta 2 \tag{9}$$

Process 3 is executed only in confined spaces in the upper part of the scene. If the slope is negative and it is bounded and maintained for a period, the region adopts the “Alert” state.

For process 4 the same reasoning as for process 2 is used. The difference between the number of pixels between two consecutive frames should be moderate, Eq. (10).

$$\delta 3 < \left| dc_{i.j.n-1} - dc_{i,j,n} \right| < \delta 4 \tag{10}$$

A region becomes “Alert” if:

$$\delta 1 < df_{i,j} < \delta 2 \quad \text{AND} \quad \delta 3 < \left| dc_{i.j.n-1} - dc_{i,j,n} \right| < \delta 4$$
$$\text{AND} \quad dc_{i,j,n} > \delta 5 \ \text{AND} \ \left| G_{i,j,n-1} - G_{i,j,n} \right| > \delta 6$$

A region takes the “Smoke” state if it is in “Alert” state and satisfies the temporal and spatial requirements that depend on indoor or outdoor scenes and the depth of the view.

III. EXPERIMENTAL RESULTS

Experiments were carried out on own and database videos, [13]. We worked on a Netbeans platform with C programming language. OpenCV library was used for video acquisition. All processes are own algorithms and work on the structure of essential pixels instead of matrices. The results are obtained in real time. We carried out experiment with video in VGA resolution (640 x 480) at 30 frames per second and ¼ VGA (320x240). The image was divided into 20x15 regions to implement the processes. The developed algorithms can be easily ported to low cost embedded systems, which is one of the main objectives of this work. A microcontroller considered suitable for this application is the NXP LPC4337 of the ARM Cortex M family with double asymmetric cores. The core M0 will deal with the digital camera interface, delivering in memory essential pixels structure for each region. The process of obtaining the essential pixels is simple and is performed as the row pixels are acquired. The M4 core will handle the processes of inference and fusion in a bare-metal application. Video in VGA, gray levels, is possible to implement in this platform. Nowhere in the process a complete frame is stored, which lowers memory RAM requirements. A software design philosophy suitable for the inference processes is a cooperative state machine since the natural sequence of events propitiates the development of software through this technique. In the platform developed in personal computers, the algorithms were implemented in this way.

Fig. 8 shows the labeling of "Smoke" regions for a video with the presence of smoke. On the left, a video frame captured and on the right the essential pixels and marked the regions of "smoke". Detection is successful using fusion mechanisms among simple inference processes.

Fig. 8. Marking on a video of "Smoke" regions in real time. Resolution 320x240 pixels. Capture of the program developed in C, platform Netbeans.

IV. CONCLUSIONS

A smoke detection process has been presented that bases its decisions on a representation of a simplified video information structure. The essential pixels represent, better than on a rectangular grid, the information in an image. As the occurrence of these pixels depends on an interpolation error that relates them to their neighbors, this interpolation error controls the amount of information they capture, presenting a simple way to minimize redundancy of the signal directly from acquisition.

It is almost impossible for a single algorithm to capture the essence of smoke pattern behavior. In the present work, simple Boolean logic fused the processes outcomes, with minimum and maximum threshold constraints. As future work is proposed to use Fuzzy Logic for the decision engine. The detection in open spaces is a challenge because of the variability of illumination, movement and the smoke dynamic changes depending on the distance of occurrence. This fact modifies the decision threshold values, but with an adequate fusion, the occurrence of false positives is minimized.

The structure of the proposed information and algorithms are simple to implement in low cost embedded systems for real time detection, which is the main objective of this work.

REFERENCES

[1] Yu Chunyu,et al. “Texture Analysis of Smoke for Real-Time Fire Detection” Computer Science and Engineering, 2009. WCSE '09. Second International Workshop on.

[2] M. A. Akhloufi, R. Lucile, A. Lilia and T. Yves, “Dynamic Fire Modeling in Three-dimensional Space”, First Workshops on Image Processing Theory, Tools and Applications, (2008), November 23-26, Sousse, pp. 1-7.

[3] N. Fujiwara and K. Terada, “Extraction of a smoke region using fractal coding”, IEEE International Symposium on Communications and Information Technology, 2004, ISCIT 2004, Volume 2, 26-29 Oct. 2004, Page(s):659 - 662.

[4] Ugur Toreyin B, Yigithan Dedeoglu, Cetin A. Enis. Wavelet based real-time smoke detection in video. 13th European Signal Processing Conference (EUSIPCO2005).Antalya, Turkey: Curran Associates, 2005.

[5] Li W, Fu B, Xiao L, et al. “A Video Smoke Detection Algorithm Based on Wavelet Energy and Optical Flow Eigen-values”. Journal of Software, pp. 63-70, 2013.

[6] MehdiTorabnezhad, Ali Aghagolzadeh, HadiSeyedarabi, “Visible and IR image fusion algorithm for short range smoke detection,” IEEE International Conference on Robotics and Mechatronics, 2013.

[7] F. Yuan, "A fast accumulative motion orientation model based on integral image for video smoke detection," Pattern Recognition Letters, vol. 29, pp. 925-932, 5/1/ 2008.

[8] Video Fire Smoke Detection Using Motion and Color Features.Yu Chunyu, Fang Jun, Wang Jinjun, Zhang Yongming. Fire Technology. July 2010, Volume 46, Issue 3, pp 651-663

[9] Enis Çetin, Bart Merci, Osman Günay, Behçet Uğur Töreyin, Steven Verstockt. “Methods and techniques for fire detection : signal, image and video processing perspectives”. London, UK : Academic Press.Elsevier, 2016.

[10] B. U. Toreyin, A. E. Centin, "Online detection of fire in video," IEEE. Conf. on Computer Vision and Pattern Recognition, pp. 1 - 5, June 2007.

[11] G. Monte, “Sensor Signal Preprocessing Techniques for Analysis and Prediction” Industrial Electronics, 2008. IECON 2008. 34th Annual Conference of IEEE. pp 1788-1793. ISBN 978-114244-1766-7.

[12] F. Abate, V. Huang, G. Monte, V. Paciello, A. Pietrosanto, “A Comparison Between Sensor Signal Preprocessing Techniques” Sensors Journal, IEEE, 25 July 2014, DOI: 10.1109/JSEN.2014.2341742.

[13] http://signal.ee.bilkent.edu.tr/VisiFire/Demo/SampleClips.html