WAVELET-BASED MULTI-MODAL FIRE DETECTION

Steven Verstockt[1,2], Ioannis Kypraios[3], Pieterjan De Potter[1], Chris Poppe[1], Rik Van de Walle[1]

[1] Department of Electronics and Information Systems, Multimedia Lab, Ghent University – IBBT
Gaston Crommenlaan 8, bus 201, B-9050 Ledeberg-Ghent, Belgium
phone: +32 9 33 14914, fax: +32 9 33 14896, email: steven.verstockt@elis.ugent.be

[2] ELIT Lab, University College West Flanders, Ghent University Association
Graaf Karel de Goedelaan 5, 8500 Kortrijk, Belgium

[3] School of Engineering and Design, University of Sussex
Falmer Brighton, BN1 9QT, United Kingdom

ABSTRACT

Over the last decade, video fire detection is started to be explored as an alternative for traditional fire sensors. Inspection of the several flame and smoke detection algorithms that have been proposed in literature shows that most of them start from simple background subtraction in spatial domain. The influence of the background model, as such, is not yet fully explored. This paper is a first attempt in this direction and investigates the added value of two wavelet-based background subtraction methods for segmenting the input scene during video fire detection. The first of these wavelet based methods focuses on both the high energy and low-pass images of the Discrete Wavelet Transformed input video frames in spatial domain. The second wavelet-based method is a non-linear Difference of Gaussians method, which is illumination invariant and performed in frequency domain. Experimental results show that both wavelet based methods lead to better fire detection results than non-wavelet based background subtraction methods. Especially when there are a lot of flame reflections and other fire-related illumination changes, less false alarms and missed detections occur in the wavelet-based setups.

1. INTRODUCTION

Recently, research on video analysis for fire detection has become a hot topic in computer vision. This has resulted in a large amount of vision-based detection techniques that can be used to detect the fire at an early stage [1]. Based on the numerous advantages of video-based sensors, e.g. fast detection (no transport delay, i.e. particles are detected as soon as they appear in the field of view of the camera), indoor and outdoor detection at a distance, and the ability to provide fire progress information, it is expected that video fire detection (VFD) will become a viable alternative for the more traditional fire sensors.

Although the experimental results in literature show that VFD promises good fire detection and analysis results, we believe that the use of wavelet-based background (BG) subtraction methods can be of added value and can help reducing the number of missed detections and false alarms. The fact that wavelet-based BG subtraction methods have much less problems with illumination changes compared to the non-wavelet based BG subtraction methods which are currently used in VFD, strengthens this idea.

The remainder of this paper is organized as follows. Section 2 lists the related work on VFD. Next, Section 3 gives a global description of the multi-modal flame detectors, which are mainly based on BG subtraction and low-cost visual/IR flame features. Subsequently, Section 4 discusses the non-wavelet and wavelet based BG subtraction techniques which were tested in our evaluation: a simple running average based dynamic BG subtraction, a more advanced MGM based simple mixture modelling, a discrete wavelet transform (DWT) based BG subtraction and an illumination invariant non-linear Difference of Gaussians method (NL-DoG) based method which is performed in the frequency domain. Then, Section 5 reports the objective evaluation results of our preliminary experiments on challenging fire and non-fire video sequences. Finally, Section 6 lists the conclusions.

2. VIDEO FIRE DETECTION

2.1 VFD in visible light

The several vision-based fire and smoke detection algorithms that have been proposed in literature have led to a large amount of VFD algorithms that can be used to detect the presence of fire at an early stage. The majority of these algorithms detect flames or smoke by analyzing vision-related fire features such as color, motion, energy, and spatial and temporal fire disorder. Color was one of the first features used in VFD and is still by far the most popular. The majority of the color-based approaches make use of RGB color space, sometimes in combination with color-related features from other color spaces. Other frequently used fire features are flickering and energy variation [2, 3]. Both focus on the temporal behavior of flames. Fire also has the unique characteristic that it does not remain a steady color, i.e., the flames are composed of several varying colors within a small area. Spatial difference analysis [4] focuses on this feature to eliminate ordinary fire-colored objects with a solid flame color. Also an interesting feature for fire detection is the disorder of smoke and flame regions over time. Frequently used metrics to measure this disorder are randomness of area size and boundary roughness [3, 5]. Finally, motion is also used in most VFD systems as a feature to improve the detection process, i.e., to eliminate the disturbance of stationary non-fire objects. In order to detect possible motion, the moving part in the current video frame is detected by means of a BG subtraction method [2, 3, 4, 5], i.e. the focus of this paper.

2.2 VFD in infrared

Although the trend towards infrared (IR) -based video analysis is noticeable, the amount of research about IR-based fire detection in literature is still limited. Nevertheless, the results from existing work already seem very promising and ensure the feasibility of IR video in fire detection. Owrutsky et al. [6] work in the near infrared spectral range and focus on an increase in the global luminosity, i.e., the sum of the pixel intensities in the frame. Although this fairly simple algorithm seems to produce good results, its limited constraints do raise questions about its applicability in large and open places with varying backgrounds and a lot of ordinary moving objects. Toreyin et al. [7] detect flames in infrared by searching for bright-looking moving objects with rapid time-varying contours. A wavelet domain analysis of the 1D-curve representation of the contours is used to detect the high frequency nature of the boundary of a fire region. In addition, the temporal behavior of the region is analyzed using a Hidden Markov Model. The combination of both temporal and spatial clues seems more appropriate than the luminosity approach and, according to the authors, greatly reduces false alarms. A similar combination of temporal and spatial features is used by Bosch et al. [8]. Hotspots, i.e. candidate flame regions, are detected by automatic histogram-based image thresholding. By analyzing the intensity, signature and orientation of these hot objects, discrimination between flames and other hot objects is made.

In order to detect the moving objects in the cited works, the moving part in the infrared video frames is also segmented by means of a simple non-wavelet based BG subtraction method. As such, it is also interesting to look if wavelet-based BG subtraction methods can improve the detection of fire in IR video to further improve these works.

3. MULTI-MODAL FLAME DETECTION

3.1 Low-cost visual flame detector

The low-cost visual flame detector (Figure 1) starts with one of the BG subtraction methods that are proposed in Section 4 and extracts the moving part, i.e. the foreground (FG) objects, from the BG. Each of the remaining FG objects is further analyzed using a set of visual flame features. In case of a flame object, the selected features, i.e. spatial flame color disorder, principal orientation disorder and bounding box disorder, vary considerably over time. Due to this high degree of disorder, extrema analysis is chosen as a technique to easily distinguish between flames and other non-flame objects. If the number of extrema, i.e. local maxima and minima, is high, the region is labeled as a flame region. For more detailed information the reader is referred to [9].

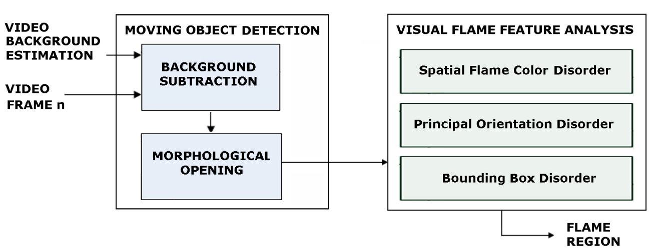

Figure 1: Low-cost visual flame detector

3.2 Low-cost IR flame detector

Similar to the visual flame detector, the thermal long-wave IR (LWIR) detector (Figure 2) starts with a BG subtraction. Then, it automatically extracts hot objects from the foreground thermal images by histogram-based segmentation, which is based on Otsu's method [10]. After this thermal filtering, only the relevant hot objects in the scene remain foreground. These objects are then further analyzed using a set of three LWIR fire features: bounding box disorder, principal orientation disorder, and histogram roughness. The set of features is based on the distinctive geometric, temporal and spatial disorder characteristics of bright flame regions, which are easily detectable in LWIR thermal images. By combining the probabilities of these fast retrievable local flame features we are able to detect the fire at an early stage.

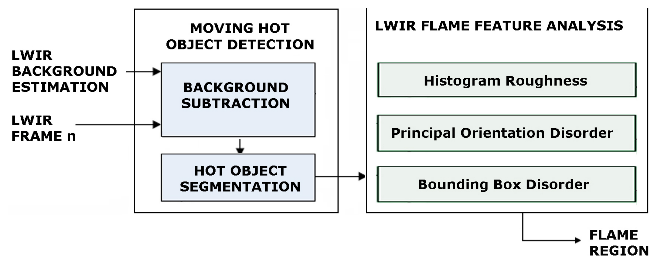

Figure 2: Low-cost LWIR flame detector

Consequently, one can think of combining both flame detectors. This has already been done by the authors in [11]. However, as the focus of this paper is on the influence of the BG model in different multi-modal spectra, and not on the multi-model processing itself, it is (currently) out of the scope of our experimental setup.

4. BACKGROUND SUBTRACTION METHODS

4.1 Running average based background subtraction

Currently, the majority of flame detectors is based on rather simple dynamic background subtraction methods, such as the running average based method which is used in [7, 8]. This type of background models extracts moving objects by subtracting the LWIR/video frames with everything in the scene that remains constant over time, i.e. the estimated background BG_n. This estimation is updated dynamically after each segmentation using (Eq. 1). Only pixels which are labeled as BG in F_n are updated in BG_{n+1} using their pixel value. FG labeled pixels, on the other hand, are not updated, i.e. for these pixels $BG_{n+1} = BG_n$.

$$BG_{n+1}[x,y] = \begin{cases} \alpha BG_n[x,y] + (1-\alpha)F_n[x,y] & \\ & \text{if } F_n[x,y] \rightarrow BG \\ BG_n[x,y] & \\ & \text{if } F_n[x,y] \rightarrow FG \end{cases} \quad (1)$$

where the update parameter α is a time constant that specifies how fast new information supplants old observations. Here α (=0.95) was chosen close to 1 as in the work of Toreyin et al. [7].

4.2 Advanced MGM: simple mixture of models

Mixture of Gaussians Model (MGM) is one of the most popular background subtraction techniques, which can handle highly complex, multi-modal scenes with difficult situations like moving trees and bushes, clutter, noise, and permanent changes of the background. However, although MGM gives good results in many video surveillance applications, the use of the Gaussian models and the update scheme are complex.

To overcome the complexity of the traditional MGM, a simple mixture of models technique (SMM) is proposed by Poppe et al. [12]. The SMM models consist of an average, an upper and lower threshold, a maximum difference with the last background value, and an illumination allowance based on Skellam parameters. In many cases, only performing temporal background subtraction is insufficient, so SMM is extended with spatial information, i.e., fast edge-based image segmentation, to improve the detection results. The experimental results in [12] show that this advanced MGM method is more robust than 'standard' MGM and more recent techniques, resulting in less false positives and negatives. This is also the reason why SMM is selected as one of the non-wavelet based BG subtraction methods in our evaluation. For more detailed information on SMM, the user is referred to the original work.

4.3 Discrete Wavelet Transform based FG extraction

The proposed DWT based FG extraction algorithm is shown in Figure 3. First, the input video frame is transformed using a DWT, which convolves the image with several filter banks. This leads to a multi-resolution decomposition of the image. Given the input image I_n, the decomposition produces four sub-images: the compressed (low-pass) version of the original image C_n, the horizontal detail (high-pass) image H_n, the vertical detail image V_n and the diagonal detail image D_n.

Next, the algorithm is split up into two parts, which can run simultaneously. The first part further analyzes the low-pass C_n image and extracts its moving part using a similar running average based BG subtraction as the one which is described in Section 4.1. Only its input differs: here also the previous extracted foreground FG^c_{n-1} of C_{n-1} is used in combination with the compressed BG model. The second part focuses on the high-pass detail images H_n, V_n and D_n, and combines them into an 'energy' image using (Eq. 2). This kind of energy analysis is also used with success in [3, 4] for flame feature analysis. However, to the best of our knowledge, the application of the DWT in the context of BG subtraction is novel.

$$E_n[x,y] = \sqrt{H_n^2[x,y] + V_n^2[x,y] + D_n^2[x,y]} \qquad (2)$$

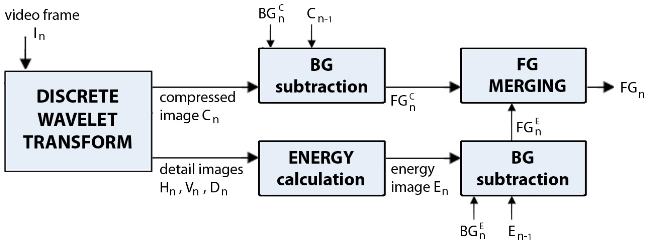

Figure 3: DWT based FG extraction

Subsequently, the moving part of E_n is subtracted with the energy BG model, which is constructed in the same way as the compressed BG model. Finally, both the compressed and energy moving part are merged and filtered. Only objects which have overlapping compressed and energy moving parts, are labeled as FG, i.e. moving object.

4.4 Non linear Difference of Gaussians

The subtraction of two Gaussian concentric kernels with different standard deviation values for a limited time duration forms a new kernel [13] which has an average value of zero turning it useful for wavelet analysis applications. The resulted Difference of Gaussians (DoG) filter can detect edges independent of orientation and when applied produce an edge enhanced image [13]. It closely resembles the Mexican Hat Hermitian (continuous) wavelet and is given by:

$$\Phi^{DoG}(x,y) = \Phi(x,y) \otimes (g_1(x,y,\sigma_1) - g_2(x,y,\sigma_2)) \qquad (3)$$

where $g_1(x,y,\sigma_1)$ and $g_2(x,y,\sigma_2)$ are the two Gaussian kernels with standard deviations σ_1 and σ_2, $\Phi(x,y)$ is the input image (in spatial domain), and $\Phi^{DoG}(x,y)$ is the linearly (linear difference) convolutioned image with the two Gaussian kernels. (Eq. 3) can be re-written as:

$$\Phi^{DoG}(x,y) = (\Phi(x,y) \otimes g_1(x,y,\sigma_1)) - (\Phi(x,y) \otimes g_2(x,y,\sigma_2)) \qquad (4)$$

Now, $\Phi^{DoG}(x,y)$ is given as the subtraction of two smoothed images of the original image $\Phi(x,y)$ with the two Gaussian kernels of different standard deviations σ_1 and σ_2. It is found that the first standard deviation σ_1, with lower value, fixes the high frequency noise in the image and, thus, produces a smoothing effect while the second standard deviation σ_2, relatively greater in magnitude, removes the low frequency content. In effect, the DoG filter forms a type of band-pass filter with lower and upper cut-off frequencies set by the two Gaussian kernels. With appropriate tuning of the standard deviation values σ_1 and σ_2, DoG filter is able to select the discriminative pass-band mid-frequency features as the foreground scene, and stop the low-frequency illumination changes effects and any high-frequency noise in the input image scene [14]. It can be shown that DoG filter approximates best the Laplacian ∇^2 operator (or the two-dimensional second directional derivative of the Gaussian kernels ∇^2 for creating a narrow band-pass differential operator [13]) when the ratio of σ_1/σ_2 is equal to 1.6.

DoG ∇^2 operator creates a non-uniform distribution of energy around the image. Hence, the partially closed areas of the image have more energy relative to other areas. However, this unequal distribution causes the image being highly sensitive to rotation and scale changes of edges. Jamal-Aldin et al. [15] have shown that by applying a non-linear function on top of the DoG ∇^2 operator, the resulted NL-DoG filter allows a more uniform distribution of energy around the closed regions of the image. Again, it is emphasized, that NL-DoG filter is based on the Mexican Hat Hermitian wavelet. In practice, the non-linearity application allows more fine details of the image around the edges to be enhanced. The non-linear function $\aleph$ is applied in the spatial domain of the image. When $\aleph$ is applied on top of the DoG ∇^2 operator the resulting image $\Phi^{NL-DoG}(x,y)$ is given by:

$$\Phi^{NL-DoG}(x,y) = \aleph \cdot \Phi^{DoG}(x,y) \qquad (5)$$

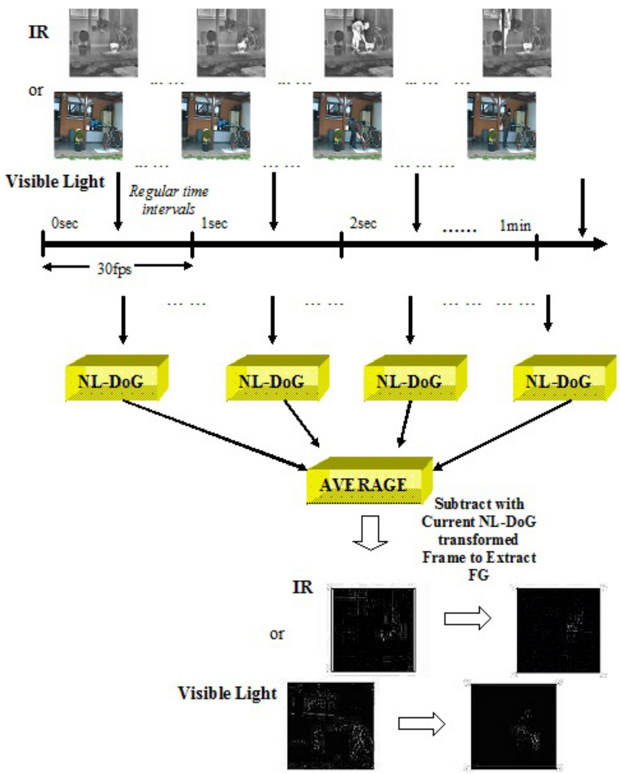

Figure 4: FG extraction in visible/IR test video sequences using TIME algorithm and NL-DoG wavelet-based filter.

Though the NL-DoG filter is formulated in the spatial (time) domain, for calculating its coefficients when applied on the input images we use a Fast Fourier Transform (FFT) operation and, then, calculate them in frequency domain. $\aleph$ is chosen to be a sigmoidal-type function.

Figure 4 shows the NL-DoG filter implementation for FG extraction. We used the previously described Time Intervals with Memory (TIME) algorithm for composing the reference background frame [16]. A background frame is selected at regular time intervals for the whole duration of each test video sequence in visible light or IR. Then, NL-DoG filter is applied on each selected frame in the frequency domain, and all the NL-DoG transformed frames are averaged to compose the background reference frame. We apply NL-DoG filter in the frequency domain on the current test video sequence frame, and we subtract the composed background reference frame for extracting the FG scene [16] (shown in Figure 4 is the spatial domain FG scene).

5. EVALUATION

5.1 Test setup

Figure 5 shows some exemplary visual and LWIR frames of the fire and non-fire realistic video sequences which were captured to test the proposed flame detection algorithm. As can be seen, different types of fires were investigated. The camera which was used to capture the thermal images is the Xenics' GOBI-384 [17], one of the leading commercial products of its kind. The visual camera was an ordinary CCTV camera. The image processing code was written in MATLAB, and is optimized to operate in real-time on a standard desktop or portable personal computer.

Figure 5: Exemplary (non-registered) multi-modal test sequences: pit fire, car park fire and human actions

5.2 Evaluation metrics

In order to objectively evaluate the detection results of the proposed wavelet-based BG subtraction methods, and to compare them to state-of-the-art non-wavelet based moving object detectors, the detection results are evaluated against manually created ground truth (GT) data. It's worth emphasising that this evaluation is done on an object level basis, which is more strict than the more currently used frame-based evaluation techniques. The object-based comparison compares the bounding box (BB) of every detected flame object to all the BBs of the GT flame objects which occur on the same frame. Based on all these comparisons the precision, recall, specificity and accuracy are calculated [18]. The higher each of these metrics, the better the flame detector, and more specific its BG subtraction, performs.

5.3 Preliminary results

Table 1 summarizes the detection results for all the tested sequences. The multi-modal sequences were acquired at 30fps. By comparing the precision, recall, specificity and accuracy for the investigated BG subtraction methods, the added value of wavelet versus non-wavelet based BG subtraction can easily be seen. As the results indicate, the DWT and NL-DOG yields best detection results. They perform especially better than the investigated state-of-the-art non-wavelet based methods when light conditions are bad, such as in the car park fire test. As an additional tool for studying the performance of each method, one can easily draw the receiving operating characteristics (ROC) curves based on the GT data.

By further inspecting the results one can also see that the overall gain of using wavelet based BG subtraction is bigger in the visual than in the thermal domain. This is to be expected, as illumination and light-related problems are visual artifacts, which do not have much influence on the thermal images. Finally, it is important to remark that the precision and recall in the human actions is left blank, as the GT for this sequence is empty. For this sequence, however, it is important to investigate the specificity, i.e. the true negative rate, since this is an indication for the number of objects which are falsely detected as flames. Again, the wavelet-based methods perform better than the non-wavelet based methods.

Table 1: Performance evaluation of BG subtraction methods for visual/LWIR flame detection

sequence *method / range*	precision	recall	specificity	accuracy
outdoor pit fire				
SIMPLE / visual	0.589	0.662	-	0.487
/ IR	0.756	0.824	-	0.678
SMM / visual	0.551	0.894	-	0.522
/ IR	0.835	0.908	-	0.794
DWT / visual	0.813	0.975	-	0.796
/ IR	0.876	0.921	-	0.828
NL-DoG / visual	0.829	0.956	-	0.803
/ IR	0.848	0.937	-	0.815
car park fire				
SIMPLE / visual	0.578	0.529	0.472	0.518
/ IR	0.918	0.443	0.971	0.626
SMM / visual	0.673	0.581	0.524	0.547
/ IR	0.992	0.423	1	0.643
DWT / visual	0.779	0.577	0.756	0.650
/ IR	0.983	0.522	0.985	0.698
NL-DoG / visual	0.815	0.603	0.734	0.687
/ IR	1	0.658	1	0.788
human actions				
SIMPLE / visual	-	-	0.472	0.472
/ IR	-	-	0.595	0.595
SMM / visual	-	-	0.540	0.540
/ IR	-	-	0.663	0.663
DWT / visual	-	-	0.822	0.822
/ IR	-	-	0.956	0.956
NL-DoG / visual	-	-	0.897	0.897
/ IR	-	-	0.874	0.874

6. CONCLUSIONS

This paper investigates the added value of wavelet-based background subtraction in video fire detection. Objective evaluation on a challenging set of multi-modal fire and non-fire video sequences shows that, compared to non-wavelet based BG subtraction methods, the DWT and NL-DoG wavelet-based moving object detectors performs much better, especially when light conditions are bad or illumination changes occur. The first of the investigated wavelet based methods focuses on both the high energy and low-pass images of the DWT input video frames in spatial domain. The second wavelet-based method is the NL-DOG, which is illumination invariant and performed in the frequency domain. In order to draw complete comparisons between the DWT and NL-DOG wavelet-based methods and also between the wavelet and non-wavelet based methods, more test sequences will be analyzed in future work.

REFERENCES

[1] S. Verstockt, B., Merci, B., Sette, P., Lambert and R., Van de Walle,"State of the art in vision-based fire and smoke detection," in *Proc. 14th International Conference on Automatic Fire Detection*, 2009, pp. 285–292.

[2] G. Marbach, M. Loepfe and T. Brupbacher, "An image processing technique for fire detection in video images," *Fire Safety Journal*, vol. 41, 2006, pp. 285–289.

[3] S. Calderara, P. Piccinini, and R. Cucchiara, "Smoke detection in video surveillance: a MoG model in the wavelet domain," in *Proc. 6th international conference on Computer vision systems*, 2008, pp. 119–128.

[4] B.U. Toreyin, Y. Dedeoglu and A.E. Cetin, "Contour based smoke detection in video using wavelets," in *Proc. 14th European Signal Processing Conference*, 2006.

[5] P. V. K. Borges, J. Mayer and E. Izquierdo, "Efficient visual fire detection applied for video retrieval," in *Proc. 16th European Signal Processing Conference*, 2008.

[6] J.C. Owrutsky, D.A. Steinhurst, C.P. Minor, S.L. Rose-Pehrsson, F.W. Williams, and D.T. Gottuk, "Long Wavelength Video Detection of Fire in Ship Compartments," *Fire Safety Journal*, vol. 41, 2006, pp. 315–320.

[7] B.U. Toreyin, R.G. Cinbis, Y. Dedeoglu and A.E. Cetin, "Fire Detection in Infrared Video Using Wavelet Analysis," *SPIE Optical Engineering*, vol. 46, 2007.

[8] I. Bosch, S. Gomez, R. Molina and R. Miralles, "Object discrimination by infrared image processing," in *Proc. IWINAC*, 2009, pp. 30–40.

[9] S. Verstockt, S. Van Hoecke, N. Tilley, B. Merci, B. Sette, P. Lambert, C. Hollemeersch, and R. Van de Walle, "Firecube: a multi-view localization framework for 3d fire analysis," submitted at *Fire Safety Journal*, 2011.

[10] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, 1979, pp. 62–66.

[11] S. Verstockt, A. Vanoosthuyse, S. Van Hoecke, P. Lambert and R. Van de Walle, "Multi-sensor fire detection by fusing visual and non-visual flame features," in *4th International Conference on Image and Signal Processing*, 2010, pp. 333–341.

[12] C. Poppe, G. Martens, S. De Bruyne, P. Lambert and R. Van de Walle, "Robust spatio-temporal multimodal background subtraction for video surveillance," *SPIE Optical Engineering*, vol. 47, 2008.

[13] D. Marr and E. Hildreth, "Theory of edge detection, " in *Proc. Royal Society of London. Series B, Biological Sciences*, vol. 207, 1980, pp. 187–217.

[14] O. Arandjelovi and R. Cipolla, "Achieving illumination invariance using image filters," *Face Recognition, K. Delac and M. Grgic, Ed. InTech Education and Publishing*, 2007, p. 558.

[15] L. S. Jamal-Aldin, R. C. D. Young and C. R. Chatwin, "Non-linearity applied to wavelet pre-processed imagery to improve correlation filter performance," in *Applied Optics*, vol. 36, 1997, pp. 9212–9224.

[16] M. H. Khan, I. Kypraios and U. Khan, "A robust background subtraction algorithm for motion based video scene segmentation in embedded platforms," in *Frontiers in Information Technology (FIT-2009) by ACM*, 2009.

[17] Xenics, "Infrared Solutions: Situational Firefighting," *http://www.xenics.com/*, 2011.

[18] S. Verstockt, A. Vanoosthuyse, B. Merci, N. Tilley, B. Sette, C. Hollemeersch, P. Lambert, and R. Van de Walle, "Performance evaluation framework for vision-based fire detection," in *12th International conference on Fire Science and Engineering*, 2010, pp. 257–268.