A multi-modal video analysis approach for car park fire detection

Steven Verstockt a,b,* , Sofie Van Hoecke b, Tarek Beji c, Bart Merci c, Benedict Gouverneur d, e, A. Enis Cetin e, f, Pieterjan De Potter a, Rik Van de Walle a

1. Introduction

Video processing techniques for automatic flame and smoke detection have become a hot topic in computer vision during the last decade [1]. The different vision-based fire and smoke detectors, as proposed in the literature, led to a large amount of information produced by the TOF cameras, i.e. depth and sensor. TOF imaging takes advantage of the different kinds of sensor. Nevertheless, IR has its own specific limitations, such as thermal reflections, IR-blocking and thermal-distance problems. Furthermore, the purchase price of IR cameras is still very high. Recently, as an alternative for IR sensors, time-of-flight (TOF) imaging sensors are started to be used as a way to improve everyday video analysis tasks. TOF cameras are a relatively new innovation, capable of providing three-dimensional image data from a single sensor. TOF imaging takes advantage of the different kinds of information produced by the TOF cameras, i.e. depth and time-of-flight imaging.

In this paper a novel multi-modal flame and smoke detector is proposed for the detection of fire in large open spaces such as car parks. The flame detector is based on the visual and amplitude image of a time-of-flight camera. Using this multi-modal information, flames can be detected very accurately by visual flame feature analysis and amplitude disorder detection. In order to detect the low-cost flame related features, moving objects in visual images are analyzed over time. If an object possesses high probability for each of the flame characteristics, it is labeled as candidate flame region. Simultaneously, the amplitude disorder is also investigated. Also labeled as candidate flame regions are regions with high accumulative amplitude differences and high values in all detail images of the amplitude image's discrete wavelet transform. Finally, when there is overlap of at least one of the visual and amplitude candidate flame regions, fire alarm is raised. The smoke detector, on the other hand, focuses on global changes in the depth images of the time-of-flight camera, which do not have significant impact on the amplitude images. It was found that this behavior is unique for smoke. Experiments show that the proposed detectors improve the accuracy of fire detection in car parks. The flame detector has an average flame detection rate of 93%, with hardly any false positive detection, and the smoke detection rate of the TOF based smoke detector is 88%.

*Corresponding author at: ELIS Department—Multimedia Lab, Ghent University—IBBT, Gaston Crommenlaan 8, bus 201, 9050 Ledeberg-Ghent, Belgium. Tel.: +32 9 33 14914; fax: +32 9 33 14896.
E-mail addresses: steven.verstockt@ugent.be, sofie.van.hoecke@howest.be (S. Van Hoecke), tarek.beji@ugent.be (T. Beji), bart.merci@ugent.be (B. Merci), benedict.gouverneur@xenics.com (B. Gouverneur), cetin@bilkent.edu.tr (A.E. Cetin), pieterjan.depotter@ugent.be (P. De Potter), rik.vandewalle@ugent.be (R. Van de Walle).

Available online 13 August 2012

Keywords:
Video fire detection
Time-of-flight imaging
Multi-modal video analysis
Flame detection
Smoke detection
Video surveillance

© 2012 Elsevier Ltd. All rights reserved.
amplitude information. The ability to describe scenes using a depth map and an amplitude image, provides new opportunities in different applications, including visual monitoring (object detection, tracking, recognition and image understanding), human computer interaction (e.g. gaming) and video surveillance. As the TOF sensor cost is decreasing, it is even expected that this number of applications will increase significantly in the near future.

The possibilities of TOF based fire detection have not yet been investigated. As such, the TOF based flame and smoke detection method presented in this paper, are the first attempts in this direction. Preliminary experiments, of which some exemplary TOF flame images are given in Fig. 1, already show that the combination of amplitude, depth and visual information can be a win–win. Thus, TOF cameras seem to have great potential for fire detection. Flames, for example, generate many measurement artifacts, i.e. a kind of TOF noise, which most likely can be attributed to the emitted infrared (IR) light of the flames themselves. Contrarily to ordinary moving objects, such as people, the amplitude of flames changes very rapidly in time and shows a high degree of (spatial) disorder. Our experiments revealed that the combination of these amplitude characteristics is unique for flames, so that they are an appropriate feature for TOF-based fire detection. For smoke, similar characteristics are found, as discussed below.

The remainder of this paper is organized as follows. Section 2 gives a brief overview of the related work in TOF video analysis. Section 3 describes the working principle and the advantages/disadvantages of TOF imaging. Subsequently, Section 4 proposes the novel TOF based flame detection algorithm. Flames are detected by looking for regions which possess high probability for each of the visual flame characteristics and which have high temporal and spatial amplitude disorder. Finally, Section 5 presents the novel multi-modal smoke detector. Smoke is detected as global changes in the depth images, which do not have significant impact on the amplitude images. Experimental results are given in Section 6 and 7 summarizes the conclusions and directions for future work.

2. Related work: Time-of-flight based video analysis

To the best of our knowledge, the TOF-based flame and smoke detection algorithms, described below, are the first attempt to develop a fire detection system based on the use of a TOF depth sensor. Nevertheless, the use of TOF cameras for video analysis is not new. Recently, as an alternative for IR and visual sensors, TOF imaging sensors are started to be used as a way to improve everyday video analysis tasks. The results of these first approaches seem very promising and ensure the feasibility of TOF imaging in other domains, such as fire detection.

So far, TOF imaging devices are used for:

- Video surveillance: Hügli and Zamofing [5] explore a remedy to shadows and illumination problems in ‘conventional’ video surveillance by using range cameras. Tanner et al. [6] and Bevilacqua et al. [7] propose a TOF-based improvement for the detection, tracking and counting of people. Similarly, Grassi et al. [8] fuse TOF and IR images to detect pedestrians and to classify them according to their moving direction and relative speed. Tombari et al. [9] detect graffiti by looking for stationary changes of brightness that do not correspond to depth changes.
Image/video segmentation: In Bleiweiss et al. [10], state that the fusion of depth and color images results in significant improvements in segmentation of challenging sequences. For example, in tracking algorithms [11], segmentation is performed easily using a combination of a visual and a depth classifier, which is shown to be more functional in cluttered scenes.

(Featureless) gaming: TOF imaging also increases the gaming immersion, as with this technology people can play video games using their body as controllers. This is achieved by means of markerless motion capture, i.e. tracking and gesture recognition, using a single depth sensor. The sensor smoothly projects the player's movements onto the gaming character. Recently, several companies, e.g. Omek Interactive and Soft-kinetic, started to provide commercially available TOF technology for gesture based video gaming. Furthermore, Microsoft also focuses on this new way of gaming with its recently launched TOF-like Kinect.

Human Computer Interaction: TOF cameras also pave the way to new types of interfaces that make use of gesture recognition [12] or the user's head pose and facial features [13]. These novel interfaces can be used in a lot of systems, e.g. view control in 3D simulation programs, video conferencing, support systems for the disabled and interactive tabletops [14], which increase the attractiveness of board games.

Face detection/recognition: Hansen et al. [15] improve the performance of face detection by using both depth and gray scale images; Meers et al. [13] generate accurate TOF-based 3D face prints, suitable for face recognition with minimal data and search times.

Other applications: e-health (e.g. fall detection [16]), interactive shopping and automotive applications (e.g. driving assistance and safety functions such as collision avoidance [17,18]).

Based on the analysis of this state-of-the-art in TOF video analysis and the state-of-the-art in VFD, discussed in earlier work [19,20], a multi-modal visual-TOF flame detector and smoke detector are presented in Sections 4 and 5 respectively. Before going more into detail on each of these detectors, we first give a brief introduction on the working principle of a TOF camera and present its advantages and limitations. A basic knowledge of the TOF working principle will facilitate the understanding of the underlying ideas behind the detectors.

3. Working principle of time-of-flight camera

In order to measure the depth for every pixel in the image, the TOF camera is surrounded by infrared LEDs which illuminate the scene with a frequency modulated IR signal. This signal is reflected on the scene, and the camera measures the time $\Delta t$ needed by the signal to go back and forth. If the emitter and the receiver are punctual and located at the same place, then $\Delta t$ provides a measure for the depth of each pixel: $d = \frac{c \Delta t}{2}$, where $c$ is the signal's speed ($c \approx 3 \times 10^8$ m/s for light). Simultaneously, the camera measures the strength of the reflected IR signal, i.e. its amplitude, indicating the accuracy of the distance estimation [12].

As the depth and amplitude information is obtained using the same sensor, the depth map (Fig. 3a) and the amplitude image (Fig. 3b) are registered, i.e. they are aligned onto each other (Fig. 3c). Compared to other multi-modal detectors, no additional processing is required for correspondence matching, which is one of the strengths of the TOF sensor. Other advantages of TOF imaging are:

- Not sensitive to light changes/shadows: the TOF camera uses its own (invisible) light, which simplifies moving object detection substantially.

![Fig. 2. Working principle of TOF imaging: Modulated light is emitted from IR LEDs on the sensor. Light is reflected on the object and captured by the sensor. The time between emission and reception and the measured amplitude is used to generate the depth and the intensity images.](image)

![Fig. 3. Correspondence matching between (a) TOF depth map and (b) amplitude image; (c) registration check.](image)
Minimal amount of post-processing, giving more time to application-processing for real-time detection.

The depth map, the information of which represents the physical properties of object location and shape, can help dividing the objects during occlusion or partial overlapping [21].

Low price compared to other IR-based video surveillance cameras in use today.

In general, it can be concluded that TOF data compensates for disadvantages/weaknesses present in other data, such as noise and other problematic artifacts [10]. However, TOF imaging also has disadvantages:

- **Low spatial resolution:** The average commercially available TOF camera has a rather low QCIF resolution (176-by-144 pixels). However, as with traditional imaging technology, resolution is steadily increases with each newly released model.

- **Measurement artifacts:** Objects that are too near-by can be poorly illuminated, leading to low quality depth measurements (Fig. 4a). Significant motion can also cause corrupt depth/amplitude data, because the scene may change during consecutive range acquisitions. The sensor also has a limited 'non-ambiguity range' before the signals get out of phase (Fig. 4b). In small rooms, this is no problem, but in large rooms this can raise problems.

- **Need for active illumination:** This increases power consumption and physical size, complicates thermal dissipation, but perhaps most importantly, limits the useful operating distance of the cameras. However, as the proposed detectors mainly focus on the IR emitted by the flames themselves, this active illumination can (probably) be switched off.

4. Multi-modal time-of-flight based flame detection

The multi-modal time-of-flight based flame detector as proposed, takes advantage of the different kinds of information represented by visual and TOF imaging sensors. It is noteworthy that, in order to perform this multi-modal analysis, the visual and TOF sensor need to be registered, i.e. they need to be aligned onto each other. Some types of TOF cameras, e.g. the OptriCam [22], already offer both TOF sensing and RGB capabilities and their visual and TOF images are already registered, but the majority of TOF cameras do not have these RGB capabilities yet. As such, the calculation of the visual-TOF transformation parameters is necessary for the visual-TOF registration. For the detectors as proposed, the registration between the sensors is performed using the silhouette-based image registration method described in [4,23]. The reader is referred to these references for more details.

Experiments (Fig. 1 and 4b) revealed that only the TOF amplitude images can be used for flame detection in large open spaces, such as car parks. As TOF depth images are only 'reliable' for indoor detection within the range of the TOF camera, they cannot be used in this kind of environments. In outdoor situations or outside the range of the TOF camera (distance \(d > 10\) m), the depth estimation fails. In order to cope with this problem one could think of only using the TOF amplitude information. However, relying on the amplitude alone can cause many false detections. A better approach is thus to combine the TOF amplitude detection with a visual flame detector, which is done in the visual-TOF flame detector presented in this section.

A general scheme of the visual-TOF based flame detector is shown in Fig. 5. The algorithm consists of three stages. The first two stages are processed simultaneously: the low-cost visual flame detection and the amplitude disorder detection. The final stage, i.e. the region overlap detection, investigates the overlap between the resulting candidate flame regions from the prior stages. If there is overlap, the fire alarm is raised. Each of the stages mentioned is described next.

Based on the working principle and the advantages/limitations of the TOF sensor, several TOF related fire characteristics have been derived and evaluated experimentally. The flame detector, discussed next, is mainly based on these experiments.
4.1. Low-cost visual flame detection

The low-cost visual flame detection (Fig. 6) starts with a dynamic background (BG) subtraction [24,25]. Moving objects are extracted by subtracting the estimated background from the video frames. The estimated background consists of everything in the scene that remains constant over time. Next, in order to avoid unnecessary computational work and in order to decrease the number of false alarms caused by noisy objects, the temporal filtering removes objects that are not detected in multiple successive frames. Each of the remaining foreground (FG) objects in the video images is then further analyzed using a set of visual flame features, namely the bounding box disorder $BBD$ (Eq. 1), the principal orientation disorder $POD$ (Eq. 2) and the flame color rate $FCR$ (Eq. 3). In case of a fire object, the selected features vary considerably in time. Due to this high degree of disorder, extrema analysis is chosen as technique to easily distinguish between flames and other objects. It is related to the number of local maxima and minima within the set of data points. Only objects with a high ‘global flame probability’ $P_{flame}$ (Eq. (4)) receive ‘label 1’ in the candidate flames image $Flames_{visual}^n$. Other objects get ‘label 0’. For more detailed information concerning the visual flame detection the reader is referred to previous work [19,26]. Fig. 7 shows some exemplary ‘visual’ candidate flame regions from our car park ‘car fire’ tests.

$$BBD = \left| \text{extrema}(BB_{\text{width}}) \right| + \left| \text{extrema}(BB_{\text{height}}) \right|$$  \hspace{1cm} (1)

where $BB_{\text{width}}$ and $BB_{\text{height}}$ are the width and height of the bounding box around the object region.

$$POD = \left| \text{extrema}(s_{1:0}) \right| \div N/2$$  \hspace{1cm} (2)

where $s$ equals the angle between the x-axis and the major axis of the ellipse that has the same second-moments as the object region.

$$FCR = \frac{\#_{R-Y}(\Omega)}{\#_{\text{pixels}}(\Omega)}$$  \hspace{1cm} (3)

which is the ratio of the number of pixels $\#_{R-Y}(\Omega)$ within the red–yellow range $(R \geq G > B)$ and the total number of pixels $\#_{\text{pixels}}(\Omega)$ within the object region $\Omega$.

$$P_{flame} = \frac{BBD + POD + FCR}{3}$$  \hspace{1cm} (4)

4.2. Amplitude disorder detection

For the amplitude disorder detection, the accumulated frame difference $AFD_{amp}^n$ is calculated first. It is the sum of the differences between the current amplitude frame $F_{amp}^n$ and the previous and subsequent amplitude frames (Eq. (5)). Rounding the absolute frame differences (using the $\lfloor \ldots + 0.5 \rfloor$ round operation), the use of $AFD_{amp}^n$ already allows distinguishing rapidly changing flames from more slowly moving ordinary objects. The $\lfloor \ldots + 0.5 \rfloor$ operation can be interpreted as thresholding. It provides the closest integer to the absolute frame differences. If the fractional part is greater than or equal to 0.5, the absolute frame differences are rounded up ($\lfloor \ldots + 0.5 \rfloor$) to a larger integer. If not, they are rounded down to a lower integer.

$$AFD_{amp}^n = \lfloor |F_{amp,n} - F_{amp,n+1}| + |F_{amp,n} - F_{amp,n-1}| + 0.5 \rfloor$$  \hspace{1cm} (5)

As high $AFD_{amp}$ values also occur at the boundary pixels of ordinary moving objects that are close to the TOF sensor, this feature is not enough by itself to guarantee accurate flame detection. In order to distinguish flame pixels from such boundary pixels, the discrete wavelet transform (DWT) [27] of the amplitude image is...
also investigated. Experiments (Fig. 8) revealed that flame regions are uniquely characterized by high values in the horizontal $H$, vertical $V$ and diagonal $D$ detail images of the DWT. In case of fire, many of the flame pixels reach the maximum $H$, $V$, and/or $D$ value of 1. As such, multiplying $\max(H_D)$, $\max(V_D)$ and $\max(D_D)$ is equal to 1 for (most) flame objects. Ordinary ‘close’ moving objects do not have this characteristic. For this reason, an $AFD_{n\alpha}$ region $\Omega$ with high accumulated amplitude differences is only labeled as candidate
flame region (‘label 1’) in $F_{\text{amp}}^n$ if $\max(H_\Omega) \times \max(V_\Omega) \times \max(D_\Omega) = 1$, where $\max()$ denotes the maximum operator, and $\times$ is the product operator. Other regions receive ‘label 0’ Eq. (6) and (7).

$DWT_\Omega^{\text{detail}} = \begin{cases} 1 & \text{if } \max(H_\Omega) \times \max(V_\Omega) \times \max(D_\Omega) = 1 \\ 0 & \text{otherwise} \end{cases}$  (6)

$F_{\text{amp}}^n = \begin{cases} 1 & \text{where } AFD_\Omega^{\text{amp}} > 0 \text{ and } DWT_\Omega^{\text{detail}} = 1 \\ 0 & \text{otherwise} \end{cases}$  (7)

Next, morphological closing with a $3 \times 3$ structuring element connects neighboring candidate flame pixels (with ‘label 1’ in $F_{\text{amp}}^n$). Subsequently, a morphological opening filters out isolated candidate flame pixels using the same structuring element. The resulting connected flame pixel group(s) of $F_{\text{amp}}^n$ form(s) the amplitude candidate flame region(s). An example of the amplitude disorder detection is given in Fig. 9.

### 4.3. Region overlap detection

This final stage investigates the overlap between the visual ($F_{\text{visual}}^n$) and the amplitude ($F_{\text{amp}}^n$) candidate flame region(s), performing a logical AND operation to $F_{\text{visual}}^n$ and $F_{\text{amp}}^n$. If the resulting binary image contains one or more ‘common’ pixels, i.e. pixels with a value of 1, the fire alarm is raised. In Fig. 10, an example of this region overlap detection is shown.

In an ideal case, the detection criteria, i.e. the number of ‘common’ pixels, should depend on the environmental characteristics, such as the size of the room and the camera position. However, no method yet exists to automatically relate these characteristics to the detection criteria. Related to this, it is also important to discuss the noise sensitivity of the proposed algorithm. First of all, isolated candidate flame pixels are filtered out using a $3 \times 3$ morphological opening filter (as is mentioned on Section 4.2). This already removes the ‘single-pixel’ noise from each of the image modalities. Furthermore, one ‘common’ (multi-modal) pixel will only give a fire alarm if it is noise in both image modalities, i.e. if it is incorrectly detected as candidate flame region in the amplitude and the visual image. The likelihood of such simultaneous false positive detection in both images is low.

Before discussing the preliminary experimental results of the proposed flame detector (Section 6), the following section introduces our novel multi-modal TOF smoke detector.

### 5. Multi-modal time-of-flight based smoke detection

By further analyzing the TOF video sequences of our preliminary fire experiments, it was noticed that smoke causes a kind of global changing in the depth images. The observed phenomenon (Fig. 11) can best be described as if the scene is floating in depth.

---

**Fig. 9.** Amplitude disorder detection: (a) consecutive TOF amplitude images and their (b) morphologically and DWT filtered accumulated depth differences ($F_{\text{amp}}^n$).

**Fig. 10.** Region overlap detection: (c) logical AND of (a) visual and (b) amplitude candidate flame regions.
Furthermore, it was noted that these smoke related 'depth changes' have no significant impact on the amplitude images. Based on this TOF related smoke behavior, the development of a novel TOF based smoke detector was initiated. Although this detector has not yet been fully evaluated, its preliminary results show that a TOF sensor will be able to detect smoke when it appears in its field of view.

A general scheme for the smoke detector is provided in Fig. 12. The algorithm consists of three stages. The first two stages are processed simultaneously: the amplitude based detection of background blocks and the average depth change detection. The final step, namely the block overlap detection, checks the overlap between BG amplitude blocks and moving depth blocks. Overlapping blocks, i.e. blocks with an average depth change which does not cause changes in the amplitude values, are labeled as candidate smoke. If several candidate smoke blocks occur in consecutive images, the fire alarm is raised.

The algorithm performs the smoke detection on a block level, rather than on pixel level. Each input frame \( F_{\text{depth}}^n \) and \( F_{\text{amp}}^n \) is subdivided into \( n \times n \) size blocks, in order to reduce measurement disturbances, i.e. in order to filter out errors and measurements inaccuracies. Depending on the camera resolution and the scene characteristics, an appropriate blocksize must be chosen. In our experiments, blocks of \( 8 \times 8 \) pixels proved to be the most successful. For each depth block \( b_{\text{depth}}^n[i,j] \), an average depth value \( F_{\text{depth}}^n[i,j] \) is computed as the average of all the pixel values \( F_{\text{depth}}^n[x,y] \) in that block (Eq. (8)).

\[
F_{\text{depth}}^n[i,j] = \frac{\sum_{x=i}^{i+n-1} \sum_{y=j}^{j+n-1} F_{\text{depth}}^n[x,y]}{n \times n}
\]

5.1. Amplitude based background detection

Experiments revealed that smoke has no significant impact on the TOF amplitude images. When smoke appears in the field of view of the TOF camera, the amplitude images remain practically unchanged. In the depth images, on the other hand, smoke causes a global change in the depth direction (as discussed below). In order to detect the non-changing part of the amplitude images, i.e. the amplitude BG blocks, we perform a kind of moving object detection. However, instead of looking for blocks with certain level of motion, we look for blocks that do not change significantly. A BG amplitude block is determined by comparing the amplitude values of the block \( b_{\text{amp}}^n[i,j] \) in the current frame \( F_{\text{amp}}^n \) with the values of the corresponding block in the BG model \( B_{\text{amp}}^n \). If the sum \( \sum_{x=i}^{i+n-1} \sum_{y=j}^{j+n-1} |dF_{\text{amp}}^n[i,j]| \) of the absolute differences of the block pixels (Eq. (9)) is lower than the dynamic threshold \( t_{\text{BG}} \), the block is labeled as BG block.
(Eq. (10)). For more information on the dynamic threshold, the reader is referred to [28,29].

\[
dif_{amp}[i,j] = \sum_{x = i-1}^{x = i+1} \sum_{y = j-1}^{y = j+1} |F_{amp}(x,y) - RC_{amp}(x,y)|
\]  

(9)

\[
b_{amp}[i,j] = \begin{cases} 
BG & \text{if } dif_{amp}[i,j] < t_{BG} \\
FG & \text{otherwise}
\end{cases}
\]  

(10)

5.2. Average depth change detection

The average depth change detection performs a temporal analysis of the average depth values (Eq. (8)) of the current block \(b_{depth}[i,j]\) and the previous blocks \(b_{depth}[k,n]\). If the standard deviation of these average depth values exceeds the \(t_{depth}\) threshold of 0.1, the block is labeled as moving depth block.

\[
b_{depth}[i,j] = \begin{cases} 
MOVING & \text{if } \sigma(b_{depth}[i,j]) > t_{depth} \\
NON-MOVING & \text{otherwise}
\end{cases}
\]  

(11)

5.3. Block overlap detection

The block overlap detection, the final step of our TOF based smoke detector, checks the overlap between BG amplitude blocks \(b_{amp}[i,j]\) and moving depth blocks \(b_{depth}[i,j]\). Overlapping blocks, i.e. blocks with an average depth change that does not cause changes in the amplitude values, are labeled as candidate smoke block (Eq. (12)). If several candidate smoke blocks occur in consecutive TOF images, the fire/smoke alarm is raised.

\[
b_{i,j} = \begin{cases} 
SMOKE & \text{if } b_{depth}[i,j] = MOVING \text{ and } b_{amp}[i,j] = BG \\
NON-SMOKE & \text{otherwise}
\end{cases}
\]  

(12)

If necessary, one can also use a visual smoke detector [19] in addition to, or as a complement for, the amplitude and/or depth detection steps. However, as the experimental results in the next section show, the detector as proposed already gives very promising results. As such, investigating the benefit/added value of using an additional visual smoke detector is considered beyond the scope of the present paper.

### 6. Experimental setup and results

The TOF camera used in this work is the Panasonic D-Imager [30]. The D-imager is one of the leading commercial products of its kind. Other appropriate TOF cameras are the CanestaVision from Canesta, the SwissRanger from Mesa Imaging, the PMD [vision] CamCube and the Optricam from Optrima [22]. The technical specifications of the D-Imager are shown in Fig. 13.

The visual camera, the images of which were registered [4,23] onto the TOF images, is a Linksys WVC2300 camera, a standard security camera. The image processing code was written in MATLAB and is sufficiently simple to operate in real-time on a standard desktop or portable personal computer.

In order to illustrate the potential use of the proposed multi-modal flame and smoke detector, several fire and non-fire experiments have been performed in a car park. Besides the fire, also ordinary moving objects occur in each of the fire sequences. An example of one of the fire experiments, i.e. the ‘car fire 1’ test, is shown in Fig. 14. As can be seen in the amplitude images, the amplitude of flames contains a high degree of temporal and spatial disorder. Even between two consecutive frames, these high amplitude differences are noticeable at the boundaries of the flames. The histograms of the depth images, on the other hand, show the depth changing related to the smoke.

In order to evaluate the detection results of the proposed methods in an objective manner, the detection rate metric (Eq. (13)) is used. This metric is comparable to the evaluation methods used by Celik et al. [31] and Toreyin et al. [32]. The detection rate equals the ratio of the number of correctly detected flame/smoke frames, i.e. the detected flame/smoke frames minus the number of falsely detected frames, to the number of frames with ‘fire’, i.e. flames or smoke, in the manually created ground truth (GT).

\[
detection\text{-}rate = \frac{\#\text{detected} - \#\text{false\ detections}}{\#\text{GT fire frames}}
\]  

(13)

6.1. Performance evaluation of multi-modal flame detector

The results in Table 1 show that robust flame detection can be obtained with relatively simple multi-modal visual/TOF image...
The state-of-the-art methods, chosen for comparison to the proposed flame detection algorithm (Method 1), are the flame detection method by Celik and Demirel [31] (Method 2), which uses a rule-based generic color model for flame pixel classification, and the feature-based flame detection method by Borges et al. [33] (Method 3). Compared to these state-of-the-art VFD detection results, with an average detection rate of 87% and an average false positive rate of 3%, the proposed flame detector,

![Fig. 14. Fire experiment (car fire 1): (a) TOF amplitude images and (b) TOF depth images; (c) visual images.](image)

Table 1
Performance evaluation of TOF-visual flame detector. A comparison is made between the proposed multi-modal flame detection method (Method 1), the flame detector by Celik and Demirel [31] (Method 2), which uses a rule-based generic color model for flame pixel classification, and the feature-based flame detection method by Borges et al. [33] (Method 3).

<table>
<thead>
<tr>
<th>Video sequence (distance in m)</th>
<th># Flame frames GROUND TRUTH</th>
<th># Detected flame frames</th>
<th># False positive frames (-false alarms)</th>
<th>Detection rate¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 min videos at 30fps → 1800 frames</td>
<td>1 GT frame/5 frames</td>
<td>1/Ground truth</td>
<td>1/Ground truth</td>
<td>1/Ground truth</td>
</tr>
<tr>
<td>Car fire 1 (13 m)</td>
<td>1280</td>
<td>1194 1088 1103</td>
<td>0 10 14</td>
<td>0.92 0.84 0.85</td>
</tr>
<tr>
<td>Car fire 2 (19 m)</td>
<td>1055</td>
<td>961 942 959</td>
<td>0 22 18</td>
<td>0.91 0.87 0.89</td>
</tr>
<tr>
<td>Paper fire (10 m)</td>
<td>565</td>
<td>521 502 497</td>
<td>0 9 11</td>
<td>0.92 0.87 0.86</td>
</tr>
<tr>
<td>Christmas tree fire (10 m)</td>
<td>630</td>
<td>604 569 582</td>
<td>0 8 5</td>
<td>0.96 0.89 0.92</td>
</tr>
<tr>
<td>Fire fighters (&gt;10 m)</td>
<td>0</td>
<td>2 21 14</td>
<td>2 21 14</td>
<td>– – –</td>
</tr>
<tr>
<td>Moving crowd (7 m)</td>
<td>0</td>
<td>0 17 25</td>
<td>0 17 25</td>
<td>– – –</td>
</tr>
</tbody>
</table>

¹ detection rate = (# detected flame frames – # false alarms) / # GT flame frames.
Table 2
Performance evaluation of TOF based smoke detector. A comparison is made between the proposed multi-modal smoke detection method (Method 1), the contour/wavelet based smoke detector by Toreyn et al. [32] (Method 2) and the feature-based smoke detection method by Xiong et al. [34] (Method 3).

<table>
<thead>
<tr>
<th>Video sequence (distance in m)</th>
<th># Smoke frames</th>
<th>GROUND TRUTH</th>
<th># Detected smoke frames</th>
<th># False positive frames (−false alarms)</th>
<th>Detection rate$^a$ Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 min videos at 30fps = 1800 frames</td>
<td>1 GT frame/5 frames</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Car fire 1 (13 m)</td>
<td>690</td>
<td>623</td>
<td>616</td>
<td>598</td>
<td>18</td>
</tr>
<tr>
<td>Car fire 2 (19 m)</td>
<td>835</td>
<td>744</td>
<td>757</td>
<td>732</td>
<td>0</td>
</tr>
<tr>
<td>Paper fire (10 m)</td>
<td>340</td>
<td>290</td>
<td>309</td>
<td>291</td>
<td>0</td>
</tr>
<tr>
<td>Christmas tree fire (10 m)</td>
<td>870</td>
<td>804</td>
<td>783</td>
<td>772</td>
<td>11</td>
</tr>
<tr>
<td>Fire fighters (&gt;10 m)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Moving crowd (7 m)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

$^a$ detection rate = (# detected smoke frames − # false alarms)/# GT smoke frames.
with its 93% detection rate and hardly any falsely positive detection, performs better for the evaluation test set of fire and non-fire experiments.

In order to test the detection range of the proposed multi-modal detector, the distance between the sensors and the fire/moving objects has also been varied during the experiments. Increasing the distance between the cameras and the fire source does not affect the detection results substantially. For example, the detection rate of ‘car fire 2’ test (at 19 m distance) is 91%, which is almost as good as the 93% of ‘car fire 1’ test (at 13 m distance). In Fig. 15, we show visual screenshots for each of the tested sequences. The multi-modal smoke detector, discussed hereafter, is evaluated on the same set of fire and non-fire experiments.

**Remark on BG estimation models**: It is important to mention that most of the (visual) background estimation models fail to generate a robust background image in case of fires, resulting in fluctuated illuminations. In order to cope with this fire-related illumination problem, the authors investigated the added value of wavelet-based background subtraction methods. The results of this research have recently been presented in [34]. Wavelet based methods are shown to yield better fire detection results than non-wavelet based background subtraction methods. Especially when there are many flame reflections and other fire-related illumination changes, fewer false alarms and missed detections occur in the wavelet-based setups.

### 6.2. Performance evaluation of multi-modal smoke detector

The results in Table 2 show that the multi-modal smoke detector as proposed (Section 5) is able to accurately detect the smoke. The state-of-the-art methods, chosen for comparison to the proposed smoke detection algorithm (Method 1), are the contour/wavelet based smoke detector by Toreyin et al. [32] (Method 2) and the feature-based smoke detection method smoke detection method by Xiong et al. [35] (Method 3). Compared to these state-of-the-art VFD detection results, with an average detection rate of 87% and an average false positive rate of 3%, the proposed flame detector, with its 88% detection rate and 1% false positive detections, achieves practically similar detection results. Depending on the environment...
et al. [33] (Method 3). For the last test, we also compared our own smoke detection results with the contour/wavelet based smoke detector by Toreyin et al. [32] (Method NIST), recently started to tackle this problem and are developing easy to be conducted frequently. International standard organizations have proposed flame and smoke detectors and to facilitate comparison with SOTA alternatives, i.e. the flame detection rate is already clear that the proposed methods also perform well for these more challenging test sequences. However, more objective results are given in Table 3. Again, the proposed methods outperform visual SOTA alternatives, i.e. the flame detection rate is higher and the number of false alarms is lower.

In the first challenging test (Fig. 16a) a person with reflective coat is walking close to an outdoor daytime fire. Only the fire is detected as a flame object. The reflective coat is not detected and does not disturb the fire detection itself. In the ‘car park entrance’ test (Fig. 16b) the headlights of moving cars, entering or leaving the car park, do not generate false alarms. Similarly, the ground reflections of the fire in the ‘outdoor night-time fire’ test (Fig. 16c) are eliminated by combining the visual and TOF detection results. Furthermore, the ‘outdoor car park’ test (Fig. 16d) shows that car signal/turn lights do not cause mis-detections, either. Finally, in the ‘indoor car park fire’ test both smoke and flames are detected. The proposed work is complementary to previous work [19], in which a multi-view localization framework for 3D fire analysis has been proposed. While the previous work focuses on fire analysis by combining multi-view detection results, considered the next step in an intelligent video-based fire detection system, the present paper focuses on increasing the detection performance. The present work can thus be used as ‘input’ in the multi-view localization framework.

Finally it is stressed that, although the paper is more application oriented and TOF images are already widely used in the computer vision field, the paper also addresses some technical/ fundamental problems, such as the multi-modal image processing/registration, fire detection outside the range of the TOF detector and false alarm reduction. The main contribution of the paper is the fusion of (existing) video analysis techniques into a

### Table 3

<table>
<thead>
<tr>
<th>Video sequence (distance in m)</th>
<th># Flame frames GROUND TRUTH</th>
<th># Detected flame frames</th>
<th># False positive frames (false alarms)</th>
<th>Detection rate</th>
<th>Method 1</th>
<th>Method 2</th>
<th>Method 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outdoor daytime fire (13 m) + reflective coat</td>
<td>1685</td>
<td>1594</td>
<td>0</td>
<td>0.95</td>
<td>0.88</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>Car park entrance (9 m) + headlights of moving car</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.91</td>
<td>0.86</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>Outdoor night-time fire (11 m) + fire reflections</td>
<td>1800</td>
<td>1636</td>
<td>0</td>
<td>0.91</td>
<td>0.86</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>Outdoor car park (3 m) + headlights and signal lights</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.93</td>
<td>0.94</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>Indoor car park fire (10 m) + smoke spreading</td>
<td>1560 (flame/smoke frames)</td>
<td>1455</td>
<td>9</td>
<td>0.93</td>
<td>0.94</td>
<td>0.85</td>
<td></td>
</tr>
</tbody>
</table>

* detection rate = (# detected flame frames – # false alarms)/# GT flame frames.
novel algorithm for multi-modal flame and smoke detection. Contrary to many other research approaches, the proposed optimization for the detection of flames/smoke is more based on a breadth than on a depth research. To the best of our knowledge, this has not been done before within the fire context.

Acknowledgments

The research activities as described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT), University College West Flanders, Warrington Fire Ghent, the Car Park Fire Safety project (www.carparkfiresafety.be), Xenics nv, Bilkent University (Department of Electrical and Electronics Engineering), the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientific Research-Flanders (FWO-Flanders G.0060.09), the Belgian Federal Science Policy Office (BFSPO), and the European Union.

References