

Analyse van multimodale video voor vroegtijdige branddetectie

Multi-modal Video Analysis for Early Fire Detection

Steven Verstockt

Promotoren: prof. dr. ir. R. Van de Walle, dr. ir. S. Van Hoecke, prof. dr. ir. B. Merci
Proefschrift ingediend tot het behalen van de graad van
Doctor in de Ingenieurswetenschappen: Computerwetenschappen

Vakgroep Elektronica en Informatiesystemen
Voorzitter: prof. dr. ir. J. Van Campenhout
Faculteit Ingenieurswetenschappen en Architectuur
Academiejaar 2011 - 2012



ISBN 978-90-8578-468-5
NUR 980, 983
Wettelijk depot: D/2011/10.500/71

Acknowledgements

I thought of that while riding my bike
- Albert Einstein -

Writing this book feels like climbing the ‘Bosberg’ in the Tour of Flanders. After many miles of work, going over rough cobbled roads and steep and numerous climbs, this last ‘secteur pavé’ is physically probably the hardest, but is mentally the one which is most enjoyable. It is also the moment to look back to the start and to muse about all the pittoresk villages which we passed on our road. It was an interesting and joyful etappe, on a personal level as well as on a professional level. As such, it is also time to thank the people that contributed in reaching the finish.

First of all, I would like to thank the ‘tour director’ of this PhD, prof. Rik Van de Walle. He gave me the opportunity to pursue a PhD at the Multimedia Lab of Ghent University. His guidance and directions were very helpful during the four years that I performed this research. Together with my other promoters, dr. ir. Sofie Van Hoecke and prof. Bart Merci, he has pushed me to the finish. So not only Rik, but also Sofie and Bart, have helped me become a better ‘science-clyst’. A big ‘merci’ to each of them.

Special credits also go to my video analysis team mates: Gaëtan Martens, Chris Poppe, Sarah De Bruyne and Pieterjan De Potter. I am convinced that the work I present here in this dissertation would not be the same without their input and support. Also on a personal level, many moments have found their place in a happy corner in my memory. The atmosphere at Multimedia Lab has always been open and friendly, and of course I would also like to thank the other colleagues for any suggestions/remarks that have helped to improve the quality of my articles, and finally, of this book. They were my perfect ‘train’, and I am thankful that I was able to ride in their ‘slipstream’. For the administration of my ‘whereabouts’, I would also like to express a special thanks to Ellen Lammens.

From the beginning of my PhD Tour, my ELIT ‘soigneurs’ of the University College of West Flanders also provided me with the right stimulation and energy boosts I needed to survive. Especially the jokes of Johan Beke and the (m)eedings in the ‘Frituur in de Natuur’, which is our favorite ‘feed zone’, helped in bringing this work to completion. To say it in the words of my favorite soccer team: ‘ELIT, ... mes que un club’!

The collaboration with the team of prof. Bart Merci from the Department of Flow, Heat and Combustion Mechanics (FloHeaCom) was also a very learning experience. As experts in the field of fire engineering, Bart’s and his team member’s comments and reviews have surely helped improving the quality of my work. The many discussions and brainstorming sessions we had surely provided me with better insights in the problems that were faced. The atmosphere during these meetings was always very open and productive, and discussions often continued in a less formal setting after-hours, even leading to suggestions for new publications. Special thanks goes to Tarek Beji and Nele Tilley, with whom I have co-authored some contributions that bridge the gap between fire and multimedia research.

It is also the moment to think about those who I have lost during this Tour: my grandfathers, who will always be ‘presente’ in my life; my cousin Pieter and my colleague Pieter, who went to young; and Martin, who was my first contact with Xenics. That they will never be forgotten!

Of course I also want to thank my parents for their continuous support during the years. Together with my brother, the rest of my family and my friends, they added warmth and color to this Tour. A special hug goes to my best friend, my nephew Korneel. He is the one who made me smile when I couldn’t. Finally, the flowers go to Sylvie, who has conquered a special place in my heart, and supported me throughout. She will be playing the key role in the next race of my life, our ‘Paris-Roubaix’.

Steven Verstockt
December 14th, 2011

Summary

The growing demand for security has given rise to the increased use of video surveillance systems in recent years. Surveillance cameras are rapidly appearing in all sorts of places. This has highlighted various problems such as the fact that it is practically impossible for surveillance operators to keep a constant watch on the video from multiple cameras. Identifying and distilling the limited relevant information is the greatest challenge currently facing the operators of monitoring systems. To quote New Scientist magazine: *“There are too many cameras and too few pairs of eyes to keep track of them. There is need for intelligent video content analysis to support the operators by asking for attention only when unwanted behavior occurs (alarm)”*.

In the last decade, intelligent video surveillance has occupied an important position in the field of computer vision research. A considerable amount of research has been conducted concerning the detection and recognition of moving objects (people, vehicles, etc.). Furthermore, attention is also given to algorithms for tracking these objects through a sequence of images or across multiple overlapping and non-overlapping cameras. Given the considerable research efforts, it is no surprise that a variety of commercial applications for automated video analysis are coming to the market, such as perimeter security systems, traffic applications and systems for tracking and counting people and analyzing their behavior.

Intelligent video processing techniques for the detection and analysis of fire are scarce. However, fire is one of the leading hazards affecting everyday life around the world. To avoid large scale fire and smoke damage, timely and accurate fire detection is essential. The sooner the fire is detected, the better the chances are for survival. However, not only early detection is crucial, but also it is important to have a clear understanding of the fire development and the location. Where did the fire start? What is the size of the fire? What is the direction of smoke propagation? How is the fire growing? The answer to each of these questions plays an important part in safety analysis and fire

fighting/mitigation, and is essential in assessing the risk of escalation. Nevertheless, the majority of the detectors that are currently in use just ring the bell and are not able to model fire evolution, i.e. information about the fire circumstances is rarely available and difficult to measure. The research in this dissertation focuses on both problems and presents several video analysis techniques that have proven to be useful in fast and accurate detection and localization of flames and smoke. The proposed techniques are viable alternatives or complements to the existing fire detection techniques and have proven useful to solve several problems related to the traditional sensors. Those conventional sensors, for example, are generally limited to indoors and are not applicable in large open spaces such as shopping centers, airports and car parks; require a close proximity to the fire; and most of them cannot provide additional information about fire circumstances (location, dimension, etc.). Further limitations of today's fire alarm systems include the fact that it may take a long time for particles to reach the detector, i.e., the transport delay. It is our belief that video analysis can be applied in conditions in which conventional methods fail. The major reason of the (future) success of video fire detection (VFD) is its potential to detect the fire from a distance in large open spaces. VFD also does not have the transport and threshold delay that the traditional (point) sensors suffer from. As soon as smoke or flames occur in one of the camera views, fire can be detected. Finally, VFD cameras can also be used to extract useful non-fire information, such as the presence of people caught in the fire.

VFD mainly focuses on the detection and analysis of smoke and flames in consecutive video images. The research in this domain was started in the late nineties. In the beginning mainly flame detection was investigated. Recently, there is a tendency towards smoke detection. The reason for this can be found in the fact that smoke spreads faster and in most cases will occur much faster in the field of view of the cameras. This, of course, depends on the type of the fire. The majority of the state-of-the-art detection techniques focuses on the color and shape characteristics of the smoke and the flames and their temporal behavior. However, due to the variability of shape, motion, transparency, colors, and patterns of smoke and flames, many of the existing VFD approaches are still vulnerable to false alarms. The research presented in this dissertation tries to optimize this. In order to correctly localize and analyze the fire, accurate detection is needed. Everything hinges on a good detection algorithm or method.

Detection is the first step in almost any intelligent video surveillance system and is often the most difficult one. Due to noise, shadows, illumination changes and other visual artifacts in recorded video sequences, developing

a reliable detection system is a huge challenge. Contrary to many other research approaches, the proposed optimizations for the detection of flames and smoke are more in the breadth than in the depth direction. Instead of dealing with ever more complex visual fire detection algorithms, the focus of our research is on investigating and combining multi-modal information from different types of video sensors. It is our strong belief that combining multi-modal video information leads to higher detection accuracy. Each sensor type has its own specific limitations, which can be compensated by other types of sensors. Originally, due to cost reasons, it was one of our objectives to develop a fire detection system which could operate on the existing CCTV equipment. However, the cost of using multiple video sensors does not outweigh the benefit of multi-modal fire analysis. The fact that manufacturers also ensure a decrease in the sensor cost in the next years, fully opens the door to multi-modal video analysis.

The combined detection in infrared and visual spectral range is not new. The fusion of visible and infrared images has already started to be explored as a way to improve the detection performance in many application domains. Also, in the domain of fire detection some steps are already taken in this direction. When light conditions are bad, e.g., for detection at night, when smoke occurs in the field of view of the camera or when the target's color is similar to the background, IR vision is a fundamental aid. Even other visual-specific fire detection problems, such as fire-like colored objects, do not cause problems in IR. Related to this, it is important to mention that all of the currently existing visual-IR fire detectors focus on flame detection. The reason for this can probably be found in the fact that smoke becomes more and more transparent the further in infrared spectrum. However, one of the multi-modal smoke detectors proposed in this dissertation exactly uses this transparency feature of smoke in long-wave infrared (LWIR) in order to detect it. Also, as the visual perceptibility decreases and the thermal perceptibility increases the further we go in the infrared spectrum, hot objects (like flames) will be best visible and less disturbed by other objects in the LWIR spectral range. As such, we have also chosen the visual and LWIR spectral range for our infrared based multi-modal flame detector.

Although the multi-modal detection of flames and smoke in visual and infrared video already shows good results, we have also investigated the added value of time-of-flight (TOF) based VFD and its combination with visual fire detection. 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 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. The possibilities of TOF based fire detection have not yet been investigated. As such, the TOF based flame detection methods presented in this dissertation, are the first attempts in this direction. Preliminary experiments already show that the combination of amplitude, depth and visual information is a win-win. However, problems arise in outdoor situations, outside the range of the TOF camera and if smoke appears in the field of view of the TOF camera. Under these circumstances the TOF depth map becomes unreliable and cannot be used for accurate flame detection anymore. A solution to this TOF related problem is also proposed in this dissertation.

Though the majority of vision based fire detection systems consists of several cameras monitoring the same scene, the analysis is usually carried out separately on each of the camera's video sequences. In order to actually understand and interpret the fire, however, this single-view processing is not enough. By combining the detection results of each of the single-view cameras and analyzing them together, more accurate detection and localization of smoke and flames can be achieved and valuable fire characteristics are detected at the early stage of the fire. In order to accomplish this valuable fire analysis step, this dissertation also proposes a novel multi-view localization framework which fuses low-cost video fire detection results of multiple cameras. The framework merges the single-view detection results of the multiple cameras by homographic projection onto multiple horizontal and vertical planes, which slice the scene. The crossings of these slices create a 3D grid of virtual sensor points, called the FireCube. Using this grid, information about the location of the fire, its size and its direction of propagation can be instantly extracted from the video data. Subsequently, this information can be used for video based fire forecasting, i.e., the FWO project (from the Research Foundation Flanders) in which this research was carried out. The Department of Flow, Heat and Combustion Mechanics (FloHeaCom) at the Faculty of Engineering and Architecture of Ghent University, which also cooperates in this project, uses similar representations to model the fire circumstances.

Being able to model and forecast the fire can help emergency services to work more efficiently and save lives. However, the calculations with current modeling techniques still take too long and valuable time is often lost. Using the multi-view fire analysis framework, which is able to give real-time information about the state of the environment, these zone model-based predictions

of the future state can (probably) be improved and accelerated. By combining the information about the fire from models and real-time data, an estimate of the fire can be produced that is better than could be obtained from using the model or the data alone. This is the final goal of the video based fire forecasting FWO project, of which this work finalizes the first part. The second part, i.e., linking the modeling and the real-time detection, is performed by our colleagues of the Department of Flow, Heat and Combustion Mechanics, under supervision of Prof. Bart Merci. Important to note is that not only flame and smoke information is needed to efficiently forecast and fight the fire, but also other information about the monitored scene can be of high importance. For example, a broken window or a door which is opened can influence the fire growth. Most of this data can also be delivered by an intelligent video surveillance system. However, this is out of the scope of this dissertation.

This dissertation covers different aspects of an intelligent video based fire detection system. Our first contribution treats the multi-modal processing of visual, infrared and time-of-flight video images, which improves the visual detection of flames and smoke. In order to keep the processing cost low, i.e., to ensure real-time detection, a set of 'low-cost' fire features, which uniquely describe smoke and flames, is selected for each sensor individually. For the visual flame and smoke detection, we started by exploring the state-of-the-art video fire detection building blocks, and selected those who (with small modifications) could be used by our low-cost VFD algorithm. Experiments revealed that the majority of these building blocks were also applicable to other types of video images. As such, the step to TOF and infrared VFD is not that big. Experimentally it was also found that by combining the different types of video data, the number of missed detections and false alarms can be reduced drastically, which results in a significant improvement of video based fire detection. In order to combine the multi-modal detection results, the corresponding objects in the scene need to be aligned, i.e., registered. The goal of registration is to establish geometric correspondence between the multi-sensor images so that they may be transformed, compared, and analyzed in a common reference frame. Because corresponding objects in visual and thermal image may have different sizes, shapes, features, positions and intensities, the fundamental question to address during registration is: what is a good image representation to work with, i.e., what representation will bring out the common information between the two multi-sensor images, while suppressing the non-common information? Our second contribution treats this multi-modal registration question and proposes a novel silhouette based registration method, which (semi-)automatically aligns visual, TOF and infrared images. Our third and last contribution treats methods for video-based fire analysis

which, at a later stage, can also be used for fire forecasting. The main part of this contribution consists of our novel multi-view fire analysis framework which fuses low-cost video fire detection results of multiple cameras into the FireCube. Using the FireCube, the location of the fire, its size, its propagation and its direction can accurately be estimated. The proposed multi-modal detection and multi-view localization techniques have been tested thoroughly on fire and non-fire video sequences and have proven to work. For example, under the car park fire safety project (<http://www.carparkfiresafety.be/>), successful tests were conducted for early detection of car fires in a car park.

To conclude, we hope to have convinced the reader that our research has contributed in the development of an intelligent fire detection system. It is also important to stress that the proposed contributions are not limited to fire detection, but can easily be adapted to other application domains, such as multi-modal object recognition. As such, the results in this thesis are not only of scientific importance for fire detection, but also for video surveillance in general. Based on this and on the fact that the video surveillance market is growing rapidly, it is our belief that the results presented in this dissertation will increase in value in the coming years. Finally, we would like to remark that there still exist a lot of unrealistic expectations concerning the possibilities of intelligent video surveillance. Many people expect that automatic video surveillance will be able to detect and analyze everything without false alarms or missed detections and with only one simple configuration. To be honest, we believe this will never be possible. As such, the techniques proposed in this dissertation must not be seen as ‘the’ ultimate fire detection tools. They must be seen as a complement to the existing techniques. Furthermore, there still exists a wide gap between fire engineering and IT, for example in the field of interfacing and communicating the data. As such, this dissertation must be seen as a first step to bridge the gap between both worlds, which hopefully will be continued in the future.

Samenvatting

De groeiende vraag naar meer veiligheid heeft de laatste jaren aanleiding gegeven tot een exponentieel toenemend gebruik van videobewakingssystemen in ons dagelijks leven. Bewakingscamera's duiken steeds meer en meer op in het straatbeeld en op publieke plaatsen. De hiermee gepaard gaande groei in het aantal camerabeelden heeft ervoor gezorgd dat het onmogelijk is geworden voor menselijke operatoren om de vloedgolf aan camerabeelden te verwerken. Het identificeren en distilleren van de beperkte relevante informatie is de grootste uitdaging waarmee de operatoren vandaag de dag worden geconfronteerd. Daarom is er nood aan intelligente beeldanalysetechnieken (~ video content analyse) als ondersteuning van de observant door deze alleen zijn aandacht te vragen bij afwijkend gedrag (alarm).

In het laatste decennia heeft intelligente videobewaking een belangrijke positie ingenomen binnen het onderzoeksgebied van de computervisie. Al heel wat onderzoek is verricht betreffende de detectie en herkenning van bewegende objecten (mensen, voertuigen, enz.). Daarnaast is er ook al uitgebreid aandacht besteed aan algoritmes voor het volgen van deze objecten over verschillende beelden of tussen verschillende al dan niet overlappende camera's. Gezien de enorme aandacht in de onderzoekswereld, hoeft het niet te verwonderen dat er ook een verscheidenheid aan commerciële toepassingen voor automatische videoanalyse op de markt komt/is. Voorbeelden hiervan zijn perimetrische beveiligingssystemen, verkeerstoepassingen en systemen voor het volgen en tellen van personen en het analyseren van hun gedrag.

Intelligente beeldanalysetechnieken voor de detectie en analyse van brand zijn echter schaars. Brand is nochtans een van de grootste problemen waarmee de wereld dagelijks wordt geconfronteerd. Om de materiële en lichamelijke schade bij een brand te beperken is het van groot belang dat de brand vroegtijdig kan worden gedetecteerd. Daarenboven is informatie over de brandontwikkeling heel nuttig tijdens de brandbestrijding. Tot op heden verloopt de detectie echter vaak te traag en is de brandinfo beperkt en moeilijk meetbaar.

Het onderzoek in dit doctoraat focust zich op beide problemen en bundelt verscheidene videoanalysetechnieken die waardevol zijn gebleken in de zoektocht naar snelle en accurate detectie en lokalisatie van vlammen en rook. De voorgestelde videoanalysetechnieken zijn complementair met de bestaande branddetectietechnieken en lossen problemen op die gerelateerd zijn aan traditionele sensoren. Zo treden klassieke sensoren pas in werking als het vuur of de rook de sensor bereikt en gaat vaak kostbare tijd verloren. Dikwijls zijn deze sensoren ook enkel bedoeld voor binnen en zijn ze ook niet inzetbaar in grote open ruimtes, zoals winkelcentra of luchthavens. Bovendien beperken ze zich vaak enkel tot het genereren van een alarm en geven ze geen extra informatie betreffende de brand (locatie, omvang, enz.). We zijn van mening dat beeldanalyse kan worden ingezet daar waar de andere detectoren tekortschieten. Videogebaseerde detectie kan immers detecteren van op afstand (zowel binnen als buiten), nuttige brandinfo genereren en ook worden ingezet om andere nuttige informatie uit de omgeving te onttrekken. Een voorbeeld hiervan is het detecteren van de aanwezigheid van personen in de ruimte in welke brand is gedetecteerd.

Videogebaseerde branddetectie focust zich hoofdzakelijk op het detecteren en analyseren van rook en vlammen in opeenvolgende videobeelden. Het (algemeen) onderzoek rond videogebaseerde branddetectie is eind jaren negentig gestart. Eerst richtte men zich op vlamdetectie, nu ook steeds meer en meer op rookdetectie. De reden hiervoor kan gezocht worden in het feit dat rook zich sneller verspreidt en meestal ook sneller zichtbaar zal zijn voor de camera. Dit is natuurlijk afhankelijk van het type brand. De meeste van de voorgestelde detectietechnieken focussen op kleur- en vormeigenschappen en het temporeel gedrag van rook en vuur. Na onderzoek is echter gebleken dat vele van deze methodes niet bestand zijn tegen de variaties in rook en vlamkleur, rookdikte en belichting. Wij proberen dit in dit doctoraat te optimaliseren. Voor een goede lokalisatie is het immers belangrijk dat de detectie correct is. Alles staat of valt met de detectie.

Detectie is meestal de eerste stap in een intelligent videobewakingssysteem en is ook vaak de moeilijkste. Ten gevolge van ruis, schaduwen, lichtveranderingen en andere artefacten in de opgenomen videosequenties is het ontwikkelen van een betrouwbaar detectiesysteem een enorme uitdaging. In tegenstelling tot wat gebeurt in het werk van vele andere onderzoekers wordt de voorgestelde optimalisatie voor de detectie van vlammen en rook niet uitgevoerd in de diepte, maar wordt een onderzoek gevoerd in de breedte. De focus ligt niet zo zeer op het krachtiger maken van bestaande visuele detectiealgoritmes, maar op het gebruiken en samenvoegen van verschillende types van informatie en de meerwaarde die hiermee gecreëerd kan worden.

We zijn van oordeel dat door het samenvoegen van multimodale informatie, afkomstig van verschillende typen sensoren, een grotere detectiewinst kan worden geboekt. Elk type sensor heeft immers zijn eigen specifieke problemen, welke gecompenseerd kunnen worden door de andere type sensoren. Niettegenstaande dat, omwille van kostredenen, een van de doelstellingen bij aanvang van dit onderzoek was om een branddetectiesysteem te ontwikkelen dat kan functioneren op de reeds aanwezige (camera)-apparatuur, lijkt de extra kost voor het gebruik van meerdere sensoren niet op te wegen tegen de meerwaarde die multimodale analyse biedt. Het feit dat ook verwacht wordt dat de sensorcost in de toekomst nog drastisch zal dalen, zet de deur naar multimodale videoanalyse volledig open.

Detectie door combinatie van infrarood video en ordinaire video is reeds in heel wat toepassingsdomeinen onderzocht. Ook in het domein van branddetectie zijn er al enkele stappen in deze richting genomen. Waar videogebaseerde systemen te kort schieten, zoals detectie tijdens de nacht, bij sterk variërende belichting of bij hoge rookontwikkeling, is gebleken dat detectie op basis van infrarood video een enorme meerwaarde biedt. Belangrijk hierbij op te merken is dat alle van de beschikbare visuele-IR branddetectoren zich focussen op vlamdetectie. De reden hiervoor kan waarschijnlijk gevonden worden in het feit dat rook onzichtbaar wordt naarmate men verder gaat in het infrarood spectrum. Een van de voorgestelde multimodale rookdetectoren in dit werk maakt net gebruik van deze onzichtbaarheid van rook in lange golf infrarood video om deze te kunnen detecteren. Ook voor visuele-IR vlamdetectie werd gekozen om te werken in de lange golf infrarood spectrale band, daar hete objecten, zoals vlammen, hier het best zichtbaar zullen zijn en het minst verstoord zullen worden door andere objecten.

Niettegenstaande de multimodale detectie van vlammen en rook met behulp van visuele en infrarood video reeds goede resultaten oplevert, wordt in dit werk ook de meerwaarde onderzocht van time-of-flight video en diens combinatie met visuele detectie. Time-of-flight video is een relatief nieuwe technologie, die anno 2010 in de gaming wereld zijn intrede heeft gedaan onder de vorm van lichaamsgebaseerde interactie en sinds kort ook voor andere videoanalyse toepassingen wordt aangewend. Voor branddetectie zijn wij echter de eerste die deze sensor hebben aangewend, en met succes, zoals blijkt uit onze resultaten. Dit type sensor kent momenteel wel nog enkele problemen, van welke zijn gelimiteerd werkgebied (maximaal 10m) mogelijks het meest cruciale is voor detectie in grote open ruimten. In dit werk suggereren wij dan ook een oplossing om dit probleem te omzeilen.

De meeste videobaseerde branddetectiesystemen geven alleen maar aan dat er brand is. Geen verdere informatie is beschikbaar. Waar de brand is, hoe groot die is, de snelheid waarmee hij groeit: daarover geven deze systemen geen info. Nochtans zijn die gegevens heel belangrijk. Hoewel deze systemen meestal bestaan uit meerdere camera's die eenzelfde omgeving monitoren, zal de analyse vaak voor elke sequentie afzonderlijk gebeuren. Door de informatie afkomstig van de verschillende camera's samen te voegen kan echter het detecteren en lokaliseren van rook en vlammen nauwkeuriger gebeuren. De in dit werk voorgestelde multi-view brandanalyse raamwerk combineert info vanuit meerdere standpunten tot een driedimensionale kubus van virtuele sensoren, de FireCube. Met behulp van het FireCube raamwerk zijn we in staat om snel en accuraat informatie te geven over de plaats van de brand, de dikte van de rooklaag en het brandproces. Deze informatie kan vervolgens worden aangewend voor videobaseerde brandvoorspelling, het FWO project (van het Fonds Wetenschappelijk Onderzoek Vlaanderen) waarbinnen dit onderzoek kadert. De universitaire vakgroep 'Mechanica van Strooming, Warmte en Verbranding' waarmee we in dit project samenwerken gebruikt gelijkaardige voorstellingen voor het modelleren van brand.

Als je het verloop van een brand kan (voor)modelleren kunnen de hulpdiensten efficiënter optreden en levens redden. De berekeningen met huidige modelleringstechnieken duren nog steeds te lang en kostbare tijd gaat vaak verloren. Als wij kunnen zeggen op elk moment, in real time, hoe dik een rooklaag is, wat de dimensie is van de vlammen en waar de brand juist zit kunnen de voorspellingen worden geoptimaliseerd. Dit is het finaal doel van het project dat loopt tot 2012, waarvan dit onderzoek het eerste luik finaliseert. Het tweede luik, de koppeling tussen modellering en real-time detectie wordt verricht door de collega's van de vakgroep 'Mechanica van Strooming, Warmte en Verbranding' onder leiding van prof. Bart Merci. Belangrijk om op te merken is dat het niet alleen vlam- en rookinformatie is die belangrijk is om brand te voorspellen en te bestrijden, maar ook details uit de omgeving. Zo hebben een springende ruit of een opengaande deur invloed op de brandontwikkeling. Ook kennis van de structuur van het gebouw kan van nut zijn. Heel wat van deze informatie kan een intelligent videobewakingssysteem ook aanleveren, doch valt dit buiten de context van dit proefschrift.

In dit proefschrift worden verschillende aspecten van een intelligent videobaseerd branddetectiesysteem onderzocht. In een eerste luik ligt de nadruk op de multimodale verwerking van visuele, infrarood en time-of-flight video-beelden, die de louter visuele detectie sterk verbetert. Om de verwerkingskost zo minimaal mogelijk te houden, met het oog op real-time detectie, is er

voor elk van het type sensoren een set 'low-cost' brandkarakteristieken geselecteerd die vuur en vlammen uniek beschrijven. Voor de visuele vlam- en rookdetectie zijn we vertrokken van bouwstenen voor automatische objectherkenning die voor handen zijn in de literatuur en die, mits enkele aanpassingen, voor videogebaseerde branddetectie kunnen worden ingezet. Experimenteel is ook vastgesteld dat vele van deze bouwstenen ook bruikbaar zijn op andere typen videobeelden. De overstap naar TOF en infrarood videogebaseerde branddetectie is dus niet zo groot. Door het samenvoegen van de verschillende typen informatie kunnen het aantal gemiste detecties en het aantal valse alarmen sterk worden gereduceerd, wat resulteert in een significante verbetering van videogebaseerde branddetectie. Om de multimodale detectieresultaten te kunnen combineren, dienen de multimodale beelden wel geregistreerd (\sim gealigneerd) te zijn. Het tweede luik van dit proefschrift focust zich hoofdzakelijk op dit samenvoegen van multimodale data en behandelt een nieuwe silhouet gebaseerde registratiemethode die semi-automatisch visuele, TOF en infrarood beelden kan aligneren. In het derde en tevens laatste luik van dit proefschrift worden methodes voorgesteld om videogebaseerde brandanalyse, en in een latere fase ook brandmodellering, uit te voeren. Ons nieuw multi-view brandanalyse raamwerk voegt door homografische projectie de detectieresultaten van meerdere camera's samen in een 3D-kubus van de omgeving, de FireCube. De kubus is een raster van virtuele videosensoren, de snijpunten van de horizontale en verticale vlakken waarop we projecteren. Zo kunnen we precies bepalen waar de rook en de vlammen zich bevinden. Tenslotte worden ook technieken voorgesteld voor het detecteren van de rookpropagatie. Daarbij analyseren we de lokalisatieresultaten in de tijd en zien we hoe de rook zich verplaatst en groeit. De in dit proefschrift voorgestelde technieken voor multimodale detectie en multi-view lokalisatie van brand zijn uitvoerig getest in de praktijk. Zo werden onder andere succesvolle testen uitgevoerd voor de vroegtijdige detectie van wagenbranden in ondergrondse parkeergarages in kader van het car park fire safety project (<http://www.carparkfiresafety.be/>).

We hopen de lezer overtuigd te hebben dat dit onderzoek een originele bijdrage heeft geleverd tot de ontwikkeling van een intelligent branddetectiesysteem. Ook in andere computervisiesystemen kunnen deze technieken interessant zijn, bijvoorbeeld voor multimodale objectherkenning. Hierdoor zijn de resultaten in dit doctoraat niet enkel interessant voor vroegtijdige branddetectie, maar voor videobewaking in het algemeen. We zijn er dan ook van overtuigd dat de resultaten die in deze verhandeling worden voorgesteld in de komende jaren nog in waarde zullen toenemen. Tot slot willen we ook nog opmerken dat er nog vaak onrealistische verwachtingen bestaan over de mogelijkheden van intelligente videobewaking. Er werd en wordt nog steeds te

veel verwacht dat automatische videoanalyse alles zal detecteren en analyseren, zonder foutmarge en met één enkele configuratie. De voorgestelde technieken in dit doctoraat dienen dan ook niet te worden gezien als de ultieme branddetectietechnieken, maar als een complement op de bestaande technieken. Daarenboven zijn er ook nog heel wat hiaten op vlak van IT en fire engineering, zoals gebruiksvriendelijke en compacte interfaces die de informatie goed in kaart brengen. Dit proefschrift dient dus ook te worden gezien als een eerste stap om de kloof te dichten tussen beide werelden, waaraan hopelijk in de nabije toekomst nog verder wordt gewerkt.

List of abbreviations

AFD	Accumulated Frame Difference
b	Block
BAD	Boundary-Area Disorder
BAR	Boundary Area Roughness
BB	Bounding Box
BBD	Bounding Box Disorder
BG	Background
C	Color image
CCD	Contour Centroid Distance
CCTV	Closed-circuit Television
CD	Chrominance Disorder
CFD	Computational Fluid Dynamics
CIF	Common Intermediate Format
COV	Coverage metric
DLT	Direct Linear Transform
DP	Dynamic Programming
DWT	Discrete Wavelet Transform
E	Energy
ED	Energy Disorder
EP	Energy Profile
F	Frame
FCR	Flame Color Rate
FFT	Fast Fourier Transform
FG	Foreground
FHA	Fire Hazard Analysis
FN	False Negative
FP	False Positive
fps	Frames per Second
fov	Field of View
GT	Ground truth
HMM	Hidden Markov Model
H	Histogram
Hom	Homography
HR	Histogram Roughness

HRR	Heat Release Rate
HSI	Hue, Saturation and Intensity (color model)
HSV	Hue, Saturation and Value (color model)
I	Intensity image
IR	Infrared
L	Luminosity
LWIR	Long-wave Infrared
MGM	Mixture of Gaussians Model
MoG	Mixture of Gaussians
MPEG	Moving Picture Experts Group
MWIR	Mid-wave Infrared
NFPA	National Fire Protection Association
NIR	Near Infrared
NIST	National Institute of Standards and Technology
NWP	Numerical Weather Predictions
P	Probability
POD	Principal Orientation Disorder
PTZ	Pan-Tilt-Zoom
QCIF	Quarter Common Intermediate Format
R	Roughness
RGB	Red, Green, and Blue (color model)
RoS	Rate of Spread
S	Silhouette image
SCA	Silhouette Coverage Analysis
SMM	Simple Mixture of Models
SOTA	State-of-the-art
SWIR	Short-wave Infrared
t	Threshold
T	Temperature
TIC	Thermal Imaging Camera
TN	True Negative
TP	True Positive
TOF	Time-of-Flight
UV	Ultra-violet
VFD	Video Fire Detection
VIFD	Video Image Flame Detection
VISD	Video Image Smoke Detection

Contents

1	Introduction	1
1.1	Context	1
1.2	Conventional fire detection	4
1.2.1	Overview of conventional detector types	5
1.2.2	Limitations of conventional detectors	9
1.3	The need for video fire detection?	10
1.4	Video (Fire) Surveillance	14
1.4.1	Moving object detection	14
1.4.2	Multi-view/multi-modal data fusion	18
1.4.3	Concluding remarks	19
1.5	Research question	19
1.6	System requirements	23
1.7	Outline	24
2	Video fire detection	27
2.1	Introduction	27
2.2	State-of-the art in video fire detection	28
2.2.1	Video fire detection in visible/visual spectral range	28
2.2.2	Video fire detection in infrared spectral range	39
2.2.3	Video object detection using time-of-flight imaging	40
2.3	Visual flame/smoke detection	45
2.3.1	Computational low-cost flame feature analysis	46
2.3.2	Low-cost smoke feature analysis	50
2.3.3	Evaluation of (single-view) smoke and flame detector	51
2.4	Infrared flame detection	54
2.4.1	Selection of the spectral range: SWIR, MWIR or LWIR?	55
2.4.2	Histogram-based hot object segmentation	56
2.4.3	LWIR flame feature analysis	56
2.4.4	Experimental results	59

2.4.5	Performance of LWIR cameras in a real fire environment?	61
2.5	Influence of the background model	64
2.5.1	Running average based background subtraction	65
2.5.2	Advanced MGM: simple mixture of models (SMM)	65
2.5.3	Discrete Wavelet Transform based FG extraction	66
2.5.4	Evaluation	67
2.6	Time-of-flight based fire detection	68
2.6.1	Indoor TOF based flame detector (distance < 10m)	70
2.6.2	TOF based smoke detection	76
2.7	General remarks and future improvements	81
2.7.1	Feature combinations and alternatives	81
2.7.2	Advanced strategies to combine feature weights	84
2.8	Conclusions	84
3	Multi-modal fire detection	89
3.1	Introduction	89
3.2	State-of-the art in multi-modal fire detection	91
3.2.1	Multi-modal video surveillance	93
3.2.2	Multi-modal image registration	95
3.3	Silhouette based multi-modal image registration	97
3.3.1	Visual silhouette extraction	99
3.3.2	Thermal silhouette extraction	105
3.3.3	TOF silhouette extraction	106
3.3.4	Visual and LWIR image registration	108
3.3.5	Experimental results	115
3.4	LWIR-visual flame detection	118
3.4.1	Global flame risk value	119
3.4.2	Experimental results	121
3.5	LWIR-visual smoke detection	121
3.5.1	Phase 1: silhouette coverage analysis	125
3.5.2	Phase 2: disorder analysis of visual silhouette	127
3.5.3	Experimental set-up and results	130
3.6	Visual - TOF flame detection	135
3.6.1	General scheme of visual-TOF flame detector	135
3.6.2	Experimental results of TOF-visual flame detector	136
3.7	Conclusions	138

4	Multi-view fire analysis	141
4.1	Introduction	141
4.2	State-of-the art in video fire analysis	143
4.3	Global description of the framework	146
4.4	Homography-based multi-view plane slicing	148
4.4.1	Homographic projections	148
4.4.2	Multi-view plane slicing	151
4.4.3	Dynamic camera maps	153
4.5	Grid analysis	154
4.6	Clean-up filtering	155
4.6.1	Spatial filtering	156
4.6.2	Temporal filtering	158
4.7	Fire and smoke development analysis	160
4.7.1	Retrieval of fire and smoke characteristics	160
4.8	Video driven fire spread forecasting	162
4.8.1	Experimental setup	164
4.8.2	Single view video fire analysis	165
4.9	Experimental results	168
4.9.1	Experiment 1: smoking machine test (subjective evaluation)	169
4.9.2	Experiment 2: pool fire test (objective evaluation)	170
4.9.3	Experiment 3: influence of the number of cameras on the detection performance (objective evaluation)	174
4.9.4	Experiment 4: video driven fire-spread forecasting	174
4.10	General remarks and future improvements	178
4.10.1	Sensor feedback and sensor fine-tuning	178
4.10.2	Combination with conventional/traditional sensors	178
4.11	Conclusions	179
5	Conclusions	183
5.1	Video fire detection	183
5.2	Multi-modal fire detection	186
5.3	Multi-view fire analysis	187
5.4	Future work	188
5.5	Answer to the research question	190
	Publications	195
	References	199

Chapter 1

Introduction

The focus of this chapter is on the introduction of an intelligent multi-view/multi-modal fire detection system, whose main goal is to minimize the risk of fire, i.e., flames and smoke, in large open spaces. First, an overview of conventional fire detection methods is given and problems/limitations of each of them are pointed out. Based on these limitations, the need for a video-based fire detection (VFD) system is explained. Subsequently, the relevant stages of a general video surveillance framework are highlighted. A basic knowledge of these stages will facilitate the understanding of the research question, which is discussed next. Based on this research question, we give the outline of this dissertation and introduce our main contributions.

1.1 Context

Fire has always been constructive as well as destructive. On the one hand, it has proven to be very useful in many daily activities, e.g., for cooking, heating, signaling and industrial applications. On the other hand, it is still one of the leading hazards affecting everyday life around the world. Recent fires, such as the Kaprun (2000) and Gotthard (2001) tunnel fires, the Black Saturday bushfires in the Australian state of Victoria (2009), the Beijing CCTV fire (2009), the 2010 Shanghai fire and the Texas bushfires (2011), have had big impact on many of our contemporaries' live. Because of its speed and destructive forces, it is one of the most serious threats. An uncontrolled fire can destroy an entire room within a few minutes and completely burn out a building in a couple of hours. What is destroyed by fire is gone forever. To avoid large scale fire and smoke damage, timely and accurate fire detection is essential. The sooner the fire is detected, the better the chances for survival. Taking all this into account, no one will deny that research on fire detection is of high importance.

The history of fire detection dates back to the beginning of the 19th century. Since then, the science and technology behind fire detection continued to improve. The primary purpose for fire detection was, and still is, the reduction of loss of life and property from fire. The role of a fire detection/alarm system is to identify the fire in a timely manner, and to alert the building's occupants and fire emergencies. In order to do this, fire detection systems are designed to detect the unwanted presence of fire by monitoring fire-related environmental changes. Humans are able to easily monitor these changes, as they are able to sense multiple aspects of a fire including the heat, flames, smoke, and odors. This is also the reason why most fire alarm systems have one or more manual alarm activation devices to be used by the person who discovers a fire. Unfortunately, humans can also be unreliable detectors, e.g., when they are not present when a fire starts or when they do not raise an alarm in an effective way. For this reason, automatic fire detectors have started to be developed.

The majority of the automatic fire detection systems used today is meant to imitate one or more of the human senses to detect smoke, heat and light generated by the fire. Thermal detectors are similar to our ability to identify high temperatures, smoke detectors replicate the sense of smell, and flame detectors are our electronic eyes. Unfortunately, these 'conventional' fire alarm systems still pose many problems, e.g., they are generally limited to indoor environments, require a close proximity to the fire, cannot provide additional information about fire circumstances such as size, location, and propagation and are unable to predict the fire spread. A further drawback of those traditional detectors is that they are subject to a transport and threshold delay, i.e., the time for particles to reach and to activate the detector. In order to provide more reliable and faster information, research on video-based fire detection (VFD) has started in the late nineties. 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) and the ability to provide fire progress information, VFD is recently becoming a viable alternative or complement for the more traditional fire sensors. By combining a video-based sensor with other, e.g., conventional, sensors or by fusing the detection results of multiple video sensors, it is even possible to develop an intelligent fire detection system which can accurately perform alarming, monitoring/interfaces and forecasting in case of fire. An overview of such a system is given in Fig. 1.1. In this dissertation we mainly focus on the added value of multiple and different types of video sensors within the system. We investigate the benefit of: multi-modal data fusion of different types of video imagery, multi-view/multi-sensor fire analysis and video-driven fire spread forecasting (\sim sensor-driven prediction).

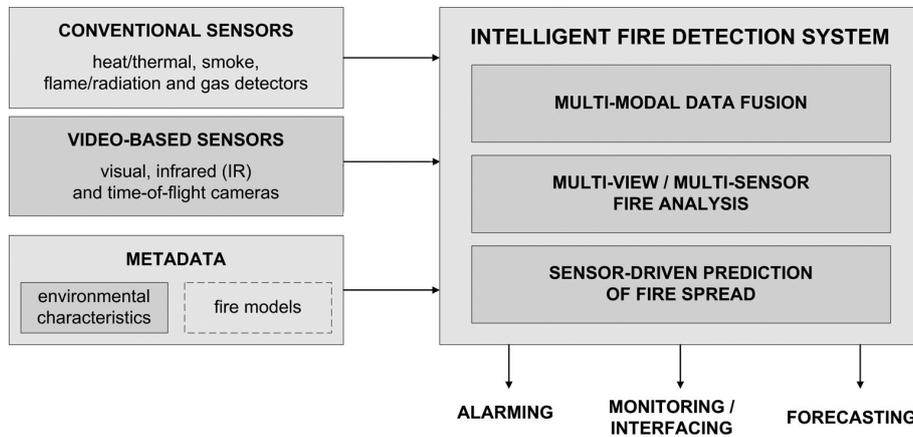


Figure 1.1: General overview of intelligent fire detection system. In this dissertation we mainly focus on the use, i.e., added value, of multiple and different types of video sensors within the system.

Besides the fusion/analysis of different types of (video) sensors, an intelligent fire detection system can also improve its performance by incorporating metadata. This metadata, i.e., side information, can take many forms. Within the scope of this dissertation, we only focus on two kinds of ‘fire’ metadata: environmental characteristics (ground plane, time of day, etc.) and fire models. More details on each of them are given in the chapters throughout this book. We will now further focus on the sensors, and most specifically on VFD.

VFD is closely related to video surveillance. The initial objective of video surveillance is to recognize an event. This could of course mean many things, such as the detection of movement or detection of presence or absence of an object. Today, video surveillance is already used in many application domains as a mechanism to protect people and property, to monitor behavior, to monitor production and so on. The detection of fire, i.e., smoke or flame events, is an example of a mechanism to protect people and property. In order to detect visual smoke and flame characteristics, many VFD algorithms implement and/or extend conventional video analysis building blocks that have already been used successfully in many other video surveillance application domains. Although this already gives satisfactory test results, real-life experiments show that problems can occur, especially when light conditions are bad or smoke production is high. As a solution to these problems, the use of other types of video sensors is started to be explored and data from multiple sensors is fused. This not only leads to better detection results, but also opens the door to the extraction and analysis of valuable fire characteristics, such as fire location, size and growth.

The remainder of this chapter is organized as follows. Section 1.2 gives a brief overview on how fire detection is performed in the ‘conventional’ way and describes its limitations. Furthermore, the ‘large open spaces’ use case, on which we focus our research, is discussed. Related to the limitations of traditional fire detectors, Section 1.3 explains the need for video fire detection. Subsequently, Section 1.4 presents some general video surveillance concepts, which will help to better understand the research described in this dissertation. Special attention is given to the concept of multi-modal and multi-view data fusion, i.e., our ‘broad’ research solutions for the depth-related research problems. Next, Section 1.5 addresses the research question of this dissertation. Finally, Section 1.6 lists the system requirements and Section 1.7 presents the outline of this book.

1.2 Conventional fire detection

In general, fire detection systems to protect and preserve an institution’s buildings, operations and occupants are classified as either automatically actuated, manually actuated, or both. Manual actuation is the oldest method of detection. In its simplest form, a person yelling can provide fire warning. However, since a person’s voice may not always transmit the information throughout the whole building, manual alarms - such as break glass stations, alarm buttons and manual pull stations - are installed along paths of escape. These devices are simple and can be highly reliable when the building is occupied. However, one of their disadvantages is that they require human interaction, and hence will not work when the building is unoccupied. Furthermore, they may also be used for malicious alarm activations. Nonetheless, they are an important component in any fire alarm system.

Automatically actuated systems, on the other hand, do not need human activity. Those systems consist of hardware and software modules which imitate the human observer. Depending on what fire related physical changes they focus on, such systems can vary dramatically in both price and complexity. Through the years, numerous types of automatic fire detectors have been developed each suited to different building types and applications. Without going too much into detail, most of the detectors can be categorized into smoke, heat/temperature, flame/radiation or gas detection. In what follows, we give a brief description on each of these ‘conventional’ detector types and discuss their limitations. For more detailed information, the reader is referred to the fire detection studies of Cote [2], NIST/NASA [3] and the NFPA 72 National Fire Alarm and Signaling Code [4].

1.2.1 Overview of conventional detector types**A. Heat/Thermal detectors**

Heat or thermal detectors are the oldest type of automatic detection device, originating from the mid 19th century with several types still in production today. The most common type of thermal sensors is the fixed temperature heat detector, which activates/triggers the alarm when the sensing mechanism reaches its specific temperature threshold. Usually these kind of sensors contain a fusible metal element which melts above the threshold temperature and causes a short on the initiating circuit. The second most common thermal units are rate-of-rise detectors, which identifies an abnormally fast temperature climb over a short time period. Both kind of heat detectors are 'spot type' detectors, meaning that they are regularly spaced along a ceiling or high on a wall. The third detector type that is often used is the fixed temperature line type detector, which consists of two cables and an insulated sheathing that is designed to breakdown when exposed to heat. The advantage of this type over spot detection is that thermal sensing density can be increased at lower cost.

Thermal detectors are highly reliable and are ideally suited to locations where high sensitivity is required for change in heat and where smoke detectors are found unsuitable for detection of fire. They are also very easy and inexpensive to maintain. One of their major drawbacks, however, is that they do not function until the room temperature has reached a substantial temperature, at which point the fire is well underway and damage is growing exponentially. Subsequently, they are usually not permitted in life safety applications and are also not recommended in locations where there is a desire to identify a fire before substantial flames occur.

B. Smoke detectors

Smoke detectors have gained wide usage during the seventies and eighties in residential and life safety applications. Their primary purpose is to replicate the human sense of smell and to identify a (smoldering) fire in its early stage. The most common smoke detectors are spot type detectors. The majority of them operate on either an ionization or photoelectric principle, with each type having advantages in different applications. Photoelectric smoke detectors interpret the reflection of a built-in light source (infrared LED) to detect the smoke. In order to do this, the light is projected into a smoke sensing chamber inside the detector assembly. In the absence of smoke, the light hits a black background of the chamber and is absorbed. When smoke enters the chamber it reflects the light on to a sensor inside the chamber. This causes the sensor

to indicate an alarm. Ionization detectors, on the other hand, use an ionization chamber and a source of ionizing radiation to detect the smoke. The radioactive source ionizes the air passing through the chamber. As a result, the air chamber becomes conductive permitting current to flow between two charged electrodes. When smoke enters the chamber, it disrupts the flow of current, which triggers the alarm.

For large open spaces such as galleries, warehouses and atria, beam detectors are frequently used instead of the more traditional spot detectors, which are difficult to install and maintain in these areas. A beam detector consists of two components, a light transmitter and a receiver, that operate in line of sight and are mounted at some distance (up to 100m) apart. As smoke migrates between the two components, the transmitted light beam becomes obstructed and the receiver cannot longer see the full beam intensity. This is interpreted as a smoke condition, and the alarm is activated. Another type of smoke detector, which has become widely used in extremely sensitive applications, is the air aspirating detector. Such a detector aspirates air samples from various locations into a tube where the sample is analyzed electro-optically for the existence of smoke. If smoke becomes present in the sample, it is detected and an alarm signal is set.

The key advantage of each of the discussed smoke detectors is their ability to identify a fire (in many cases) before severe damage occurs. They are usually the preferred detection method in life safety and high content value environments. Their disadvantage, however, is that they are usually more expensive to install, when compared to thermal sensors, and are more sensitive to false alarms. As already indicated, each type of smoke detector has advantages in different applications. Ionization detectors, on the one hand, are better at detecting fast, flaming fires than slow, and smoldering fires. Photoelectric smoke detectors, on the other hand, sense smoldering fires better than flaming fires. Hence, neither type of detector is always best, i.e., none of them provides an overall/general solution.

C. Flame/radiation detectors

Flame detectors represent the third major type of automatic detectors, and imitate the human sense of sight. They 'see' the fire by detecting the electromagnetic radiation emitted by the combustion products. Flame detectors are line of sight detectors that operate on either an infrared, ultra-violet or combined principle. Infrared detectors (IR) detect fires when a characteristic flame flicker produced by the fire is received, while ultra-violet detectors (UV) detect fires when any ultra-violet radiation produced by flaming combustion is detected.

The advantage of flame detectors is that they can be used to protect large areas and have rapid response because they do not have to rely on smoke or heat from the fire. Furthermore, they can also be used in open air, unlike the smoke detectors that need a ceiling to function effectively. However, a disadvantage is that flame detectors must be looking directly at the fire source and false alarms may be generated by radiation from other sources such as sunlight and lamps.

D. Gas detectors

Since gases are produced in all stages of combustion, a specific gas signature could also be used for reliable fire detection [5]. Techniques are available now for measuring almost any stable gaseous species produced prior to or during combustion. One such example, and probably one of the most popular types of gas detector, is the carbon monoxide detector. Such a detector is used to notify a threat of potentially hazardous amounts of carbon monoxide gas. Another form of gas detector is an explosive gas detector, which essentially monitors any type of ignitable gas. The most popular of threats for this type of detector is natural gas, due to its high use in kitchens. The response time for gas detectors is in between the response time of flame and smoke detectors. Its sensitivity and false alarm resistance is similar to the one of smoke detectors, as is further discussed in the comparison of the detector types in the next section.

E. Comparison of (conventional) fire detector types

Table 1.1 summarizes the pro and contra of each of the conventional detector types and compares them to video based detectors. The table is mainly based upon tests/data provided in [2, 5–7]. The following criteria are evaluated: the purchase price, the installation cost, the response time, the sensitivity, the false alarm resistance and the suitability for large open spaces and outdoor use. Each of these criteria is evaluated on a scale ranging from -- to ++, indicating the detector type its weaknesses and strengths respectively.

The results show that each of the detector types has advantages and limitations, making each more or less suitable for certain applications/environments. However, although it may seem that none of the detectors scores always best, video based sensors have the best overall performance. Especially for outdoor detection and detection in large open spaces, they seem the most appropriate. This will further be addressed in the next sections, in which we discuss some of the limitations of conventional detectors and emphasize the need for video fire detection.

Table 1.1: Comparison of fire detector types.

Suitable for outdoor use	+	-	-	+	-	+
Suitable for large open spaces	-	-	+	+	+	+
False alarm resistance	+	+	-	+	-	+
Sensitivity	+	-	+	+	+	+
Response time	-	-	+	+	+	+
Installation cost	+	+	+	+	+	+
Purchase price	+	+	+	+	+	-
Detector type	Human detection	Heat/thermal detectors	Smoke detectors	Flame/radiation detectors	Gas detectors	Video-based detectors

1.2.2 Limitations of conventional detectors

As each type of the conventional fire detectors has advantages for different applications, neither type of detector is always best. Furthermore, several limitations are related to each of these traditional sensors. Most of these sensors, for example, are generally limited to indoors and require a close proximity to the fire. Most of them can also not provide additional information about fire circumstances (location, dimension, etc.). Further limitations include the fact that it may take a long time for particles to reach the detector (\sim the transport delay) and the fact that traditional sensors are not able to understand the scene, i.e., to detect/interpret environmental characteristics and changes.

NFPA statistics [8] also show that, despite the advances in traditional fire alarm technology over the last century, losses caused by fire, such as deaths, permanent injuries, property and environment damages still increase. In order to decrease this, timely detection, early fire localization and detection of fire propagation are essential. Until now, however, no fire alarming system exists that is capable of giving this information in real-time. Since fires (frequently) grow at an exponential rate, an increased detection time results in larger fires. As such, a detector must try to be faster than the fire!

When focusing more on the specific use case of our research, i.e., fire detection and analysis in large open spaces, none of the traditional sensors seems somehow appropriate [9]. These spaces, e.g., atria, shopping malls, car parks, stadiums, office buildings, and airports, represent some of the most difficult fire protection challenges. Due to the open nature of these spaces, their large dimensions and their (often) excessive ceiling heights, spot detectors are not an option as the distance heat and products of combustion must travel to reach the detector makes timely detection almost impossible. This is also confirmed by recent studies, such as the study of Kuffner [10] on smoke detectors in spaces with high ceilings. The study shows that, depending on the fire type and size, it can take a long time for smoke to reach the ceiling and that there are limitations in ceiling height (maximum 6m) for smoke to reach. Beam detectors offer already some advantages over the spot detectors, as most often they can detect the fire much faster. Similar to video sensors, beam detectors can also be categorized as ‘volume-sensors’. The range you can monitor with a volume detector is much bigger than with a traditional spot detector. As soon as smoke occurs in the field of view of the beam detector, fire alarm is given. As such, their ‘transport delay’ is also very low. The purchase cost of the newest types of beam detectors, however, is between 800 and 1000 Euro. It is not expected that their cost will decrease as fast as the cost of video technology, since contrarily to video sensors, their practical application is limited to fire detection.

The installation cost, i.e., aligning the imager and the IR/UV emitter, is also high. Furthermore, maintenance costs occur, e.g., due to the fact that beam detectors are sensitive to building movement and, finally, they are also sensitive to nuisance alarms. As such, other detection technology, i.e., video based detection, is needed in order to detect and analyze the fire in large open spaces.

Several research experiments, such as those performed in [11], identified that video fire detection (VFD) systems are an effective detection technology for the protection of large industrial applications, atria and other spaces with high ceilings. VFD can quickly detect a fire by recognizing either smoke or flame anywhere within the field of view of the camera at a great distance. Furthermore, they can provide live video immediately available upon detection. This, for example allows surveillance operators to easily view the protected area, to determine the extent of the fire and to identify the fire location. And finally, VFD cameras can also be used to extract useful fire characteristics, e.g., fire size and smoke layer height, and non-fire information, such as the presence of people caught in the fire. However, despite the many advantages of VFD, it has its own limitations such as illumination-related false alarms. As will be discussed throughout this dissertation, we are able to compensate for these limitations of VFD by using the multi-modal/multi-view functions of the intelligent detection system proposed in Figure 1.1. But first, we further elaborate on the need for VFD.

1.3 The need for video fire detection?

As already mentioned before, ordinary/traditional fire detection systems have several limitations when applied in large open spaces. In order to cope with these limitations, VFD is seen as one of the most promising candidates. The best way to explain the need for video fire detection is to consider the following quotes that we have collected during our research. Each of them gives an answer to the question: why should we use video for fire detection?

- *Yet most fire alarm systems just ring the bells*

Until today, most of the fire alarms have a very limited functionality. They generate an alarm when a threshold is reached and that's it. No specific information about the fire circumstances can be given. By

running intelligent VFD techniques on the images of a video camera much more information concerning the fire can be given, e.g., the flame size, smoke layer height and the fire location. Furthermore, the video itself can also be used by surveillance operators to confirm the alarm or to evaluate the fire risk, which also is not possible when using traditional fire detectors. Perhaps, by combining multiple traditional sensors (\sim sensor networks [12, 13]) using the proposed intelligent detection system, or by combining video and traditional sensors, similar information may perhaps be achieved. However, this is out of the scope of our study.

- *Smoke spreads faster than flames AND where there's smoke, there's fire*

Compared to traditional fire detection, VFD is several steps ahead when distance and large space is concerned. For example, similar to beam detectors, VFD can detect the smoke even without it reaching the sensor. Depending on whether the focus is on flames or smoke characteristics, VFD can be split up into video image flame detection (VIFD) and video image smoke detection (VISD). Both are based on the analysis of color, motion, energy, and disorder information in video. In early research, mainly flame detection was investigated, recently there is a tendency towards smoke detection. Since in most cases smoke occurs much faster in the field of view of the cameras, focusing on smoke often offers a faster detection. Given a large open space, i.e., the proposed use case, smoke can travel in the air several times faster to reach the field of view of the camera making it possible for early detection.

- *We must try to be faster than the fire*

To avoid large scale fire and smoke damage it is important to know the evolution of the smoke and flames and to try to be faster than the fire. As such, besides timely and accurate detection and analysis, also forecasting is very important. Although, this is not an easy task, the experiments in this dissertation show that, contrary to traditional sensors, video processing techniques can not only improve and accelerate the detection, but also provides the analysis and the forecast of the fire.

- *It is not just stay or go*

In order to evaluate the fire risk, an alarm is not enough. Several fire and environment characteristics can help in assessing the level of danger. Where did the fire start? What is the size of the fire? What is the direction of smoke propagation? How is the fire growing? The answer to each of these questions plays an important role in safety analysis and fire fighting/mitigation, and is essential in assessing the risk of escalation. Using VFD and video fire analysis techniques, many of these characteristics can automatically be measured. This gives the opportunity to generate different levels of alarms and to forward a detailed description of the fire event, together with the recorded video data, to the appropriate authorities, e.g., operators and fire fighters.

- *Delivering more than the sum of the points*

First of all, video cameras are volume sensors. Contrarily to most traditional detectors, which are point sensors, a single volume sensor, such as a video camera, will be able to deliver more information about the fire event. Each pixel of the video frame can be seen as an individual detector for the region it monitors. By analyzing the behavior of neighboring pixels in the video frame over time, a video sensor is able to detect flame and smoke related features, e.g., flame flickering and flame size disorder, which can not be detected by point sensors. As such, it is already proven that a video sensor is able to deliver more than the sum of the points. Furthermore, by combining multiple cameras (using the intelligent detection system) it is also possible to automatically distill very useful information about the fire circumstances, such as the location of the fire, the smoke layer height and the flame size.

- *Video surveillance can save your loved ones and your home*

This ‘romantic’ quote is selected to convince people who (still) are against video surveillance in general. It cannot be denied that the opportunities of vision-based surveillance are numerous. There are so many applications where video surveillance can help. This is also the reason why we see an increased use of video surveillance systems in recent years. Surveillance cameras are rapidly appearing in all sort of places such as airports, city centers, road/rail networks and even in individual buildings. Large area monitoring, suspicious activity/package detection and crowd/traffic analysis are some examples of applications

that can ‘Save Your Loved Ones’ in these places. Related to the topic of this dissertation, it should be clear that also VFD and video fire analysis can help to save lives and reduce damage to ‘Your Home’.

- *Small details make big differences*

Not only flame and smoke information is needed to efficiently forecast and fight the fire, but also other information about the monitored scene can be of high importance. For example, a broken window or a door which is opened can influence the fire growth. Most of this data can also be delivered by an intelligent video surveillance system. ‘Scene understanding’, however, is a research topic on its own and is out of the scope of this dissertation.

- *CCTV is everywhere: there are too many cameras and too few pairs of eyes to keep track of them*

The number of surveillance cameras increases day by day. This highlights various problems such as the fact that it is practically impossible for surveillance operators to keep a constant watch on the video from multiple cameras. Identify and distill the limited relevant information is the immense challenge to which those operators are faced today. Hence, there is need for intelligent video content analysis to support them by asking for attention only when unwanted behavior, such as fire, occurs. This is also the reason why recently a considerable amount of research has been conducted concerning the intelligent detection of special events, such as smoke and flames. As many of these VFD algorithms work on ordinary video, they can be incorporated in existing surveillance systems at relatively low additional cost.

- *Early detection saves life and can help shorten the timeline*

Effective response to fire requires accurate and timely information of its evolution. To avoid fire disasters, minimize damage, and save lives, early fire localization and detection of fire propagation are essential. Until now, however, no system exists that is capable of accurately detecting this valuable fire characteristics in real-time. However, it is our strong belief that VFD can help shorten the timeline. For example, in most cases VFD is able to see the fire much faster than traditional fire detectors. As such, a lot of ‘crucial’ time is saved. Furthermore, video

driven fire spread forecasting, which is the final goal of this dissertation, can help in predicting the future state of the fire. This can, together with the video based retrieval of fire characteristics, help fire fighters in making the right decisions without losing lots of time.

1.4 Video (Fire) Surveillance

Video based detection and analysis of fire is closely related to video surveillance. Therefore, it should not be surprising that many VFD and video fire analysis techniques are based on general video surveillance concepts. A basic understanding of these concepts will facilitate the reading of this dissertation. As such, before going more into detail on the detection of smoke and flames, we first briefly discuss the ‘basics’ of a video surveillance system.

An automated video surveillance system can be divided into several stages, as is described in the general framework introduced by Hu et al. [14]. An overview of this framework is shown in Fig. 1.2. Its main task is to simplify the management and analysis of the enormous volume of video data and to detect events requiring attention as they happen. In order to perform this task, the framework can be divided in five different levels of abstraction: environment modeling, motion segmentation, object classification, object tracking and higher level tasks, such as behavior understanding/description, personal identification and fusion of information between multiple sensors. For each level, many algorithms exist and are topic of active research [15]. In this dissertation, we mainly focus on moving object detection, which groups the first three levels, and on the fusion of information from multiple cameras.

1.4.1 Moving object detection

Moving object detection is the first step in nearly every visual surveillance system. Its main aim consists of segmenting regions corresponding to moving objects, i.e., the foreground (FG), from the rest of an image, i.e., the background (BG). Subsequent steps such as tracking and behavior recognition are greatly dependent on it. The process of moving object detection usually involves the first three levels of the general framework, i.e., environment modeling, motion segmentation and object classification, which intersect each other during processing. These levels, extended with higher level fusion strategies, also form the main basis of the work described in this dissertation.

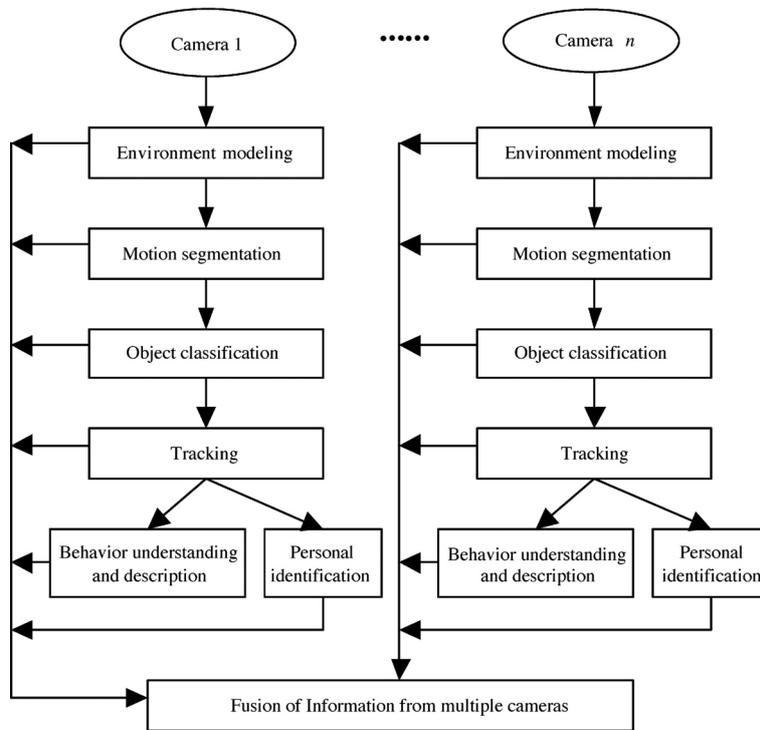


Figure 1.2: General visual surveillance framework (Hu et al. [14]).

A. Environment modeling

Environment models are required to facilitate the localization of moving objects and to evaluate their behavior and trajectories. The combination of video(s) with a model of the environment, also called ‘contextualized video(s)’, allows observers to see the activities in the videos in their proper locations. In this case, spatial relations are presented in the visualization, allowing some cognitive work to be offloaded onto the perceptual system [16]. Based on these facts, it should be obvious that the construction and updating of an environment model is indispensable to video surveillance. In order to create an environment model one can choose between a 2D model in the image plane and a 3D model in real world coordinates. Today, the majority of video surveillance applications still uses 2D models, but for the modeling of indoor scenes a switch to 3D is noticeable. In the video fire analysis experiments presented in this dissertation, both 2D and 3D environment models are used to visualize the fire characteristics of our fire tests.

B. Motion segmentation

The first step towards automated video surveillance is the detection/segmentation of interesting objects in the field of view of the camera. The definition of an interesting object is of course context dependent, but for a general surveillance system, the moving/motion part of the video sequence, such as people and vehicles, is most interesting. Segmentation of this moving/motion part is a fundamental and critical task, as errors made at this abstraction level are difficult to correct at higher levels. When an object is not detected at the lowest level, it cannot be tracked and classified at the higher levels. As such, the more accurate the segmentation at the lowest level, the easier tasks at the higher levels become.

Currently, most segmentation methods use either temporal or spatial video information and can be classified into four major groups: background subtraction, temporal differencing, optical flow and advanced statistical based methods [17, 18].

- **Background subtraction:** is (probably) the most popular method for motion segmentation, especially under those situations with a relatively static background. It detects moving regions in an image by taking the pixel-by-pixel difference between the current frame, i.e., video image, and an object-free model of the background. Usually, this reference BG model is learned, for example by averaging consecutive initialization frames over time. The pixels where the difference is above a threshold are classified as FG, i.e., moving. Although BG subtraction techniques perform well at extracting most of the relevant pixels of moving regions, they are usually sensitive to dynamic changes when, for instance, stationary objects start to move or sudden illumination changes occur.
- **Temporal differencing:** is based on frame differencing and attempts to detect moving regions by making use of the difference of consecutive frames in the video sequence. This method is highly adaptive to dynamic environments, but generally does a poor job of extracting the complete shapes of certain types of moving objects.
- **Optical flow:** is an approximation of the local image motion and specifies how much each image pixel moves between adjacent images. Although it is a complex method, it can achieve success of motion detection in the presence of camera motion or background changing, i.e., it can detect the motion accurately even without knowing the background. However, as most surveillance cameras in our work are static, this com-

putationally complex optical flow approach is not necessary. Furthermore, it cannot be used real-time without specialized hardware.

- **Advanced statistical based methods:** make use of the statistical characteristics of individual pixels. They have been developed to overcome the shortcomings of basic BG subtraction. This is also the reason why they are mainly inspired by the BG subtraction methods in terms of gathering and dynamically updating statistics of the pixels that belong to the background. FG pixels, i.e., the ‘moving’ part of the video frame, is identified by comparing each pixel’s statistics with that of the advanced BG model. Due to its reliability in scenes that contain noise, illumination changes and shadow, this approach is becoming more popular.

For our use case, i.e., the detection of fire using static cameras, (advanced) background subtraction methods are the most interesting. The majority of the proposed algorithms use a dynamic BG model, i.e., a model which is updated continuously based on the FG/BG detection of the current frame. For some of these algorithms we also investigated the added value of advanced wavelet based background subtraction methods. Wavelet-based BG subtraction methods are able to stop low-frequency illumination changes and any high-frequency noise in the input image scene. As such, it is expected that they have much less problems with illumination changes compared to non-wavelet based BG subtraction methods (which are currently used in VFD). Especially when there are a lot of flame reflections and other fire-related illumination changes, less false alarms and missed detections are expected in a wavelet-based setup. This is further discussed/investigated in Section 2.5.

C. Object classification

Typical video scenes contain a variety of objects such as people, vehicles, and natural phenomenon (e.g., rain, snow). To further track these objects and analyze their behavior/activities, it is necessary to correctly distinguish them from other moving objects. This is the task of object classification, i.e., categorizing the type of the detected regions from the motion segmentation stage. At present, there are two main categories of approaches towards moving object classification: shape-based and motion-based classification. For our research, we mainly use the former approach. For example, smoke and flame regions are detected by temporal analysis of several low-cost shape characteristics, such as the size, the boundary and the orientation of the detected region.

1.4.2 Multi-view/multi-modal data fusion

Besides the moving object detection, which groups the first three levels of the general surveillance framework (Fig. 1.2), the fusion of information from multiple cameras is the second aspect of the framework on which we focus in this dissertation. Although many tasks, such as moving object detection and tracking, can be performed using a single camera, multiple cameras can overcome many problems regarding the accurate detection and localization of moving objects [19]. The main advantages of multi-camera systems is that by exploiting the different viewpoints or different image modalities (when using different types of video sensors) they are able to increase the overall field-of-view, improve the accuracy and robustness of detection, handle the occurrence of occlusions and enable 2D/3D positioning of moving objects. These advantages, however, come at a price. Multi-camera systems have to deal with a number of technical barriers, such as the complexity increase and the ‘costs’ related to installation, calibration, object matching, and data fusion [20].

Based on the camera configurations, multi-camera systems can be divided into two categories. The first category, i.e., multi-view camera systems, fuse multiple detection results from different viewpoints in order to improve object detection and localization. The second category, i.e., multi-modal camera systems, intelligently combines different kinds of imagery sensors, e.g., visual, depth and thermal cameras, so that a single view can be provided with enriched information improving detection performance and activity analysis.

A. Multi-view video surveillance

Multi-view camera systems can be divided into systems with disjoint, i.e., spatially non-overlapping, camera views and systems with overlapping camera views [21]. Disjoint views are effective for covering wide field of views. Overlapping views, on the other hand, take advantage of the redundant information coming from different cameras monitoring the same scene to improve the accuracy in the object detection and the estimation of the objects’ position and size [22]. In our work, we mainly focus on the latter group, i.e., systems with overlapping camera views, and fuse the local detection results of multiple cameras into one global coordinate system. In order to do this, we use the homography [23] between the cameras and the coordinate system, i.e., a common technique for multi-view image fusion.

B. Multi-modal video surveillance

Contrarily to the simultaneous analysis of video from different points of view, multi-modal video surveillance focuses on the simultaneous analysis of different types of video which lines of sight are close to each other. Similar to multi-view detection, this multi-modal video processing improves the accuracy in video based object detection. Multi-modal camera systems take advantage of the different kinds of information represented by visual, thermal and/or depth imaging sensors. The combination of these types of imagery yields information about the scene that is rich in color, motion, depth and/or thermal detail. Once registered, such information can be used to successfully detect and analyze activity in the scene with fewer misdetections [24]. Since each type of sensor has its own type of detection limitations, misdetections in one sensor can be corrected by the other sensors. As such, the combination of multi-sensor information is considered a win-win by many authors [19,25] and have started to be actively used to improve the performance of object detection and recognition [26].

1.4.3 Concluding remarks

Because the main purpose of this introductory chapter is to discuss topics directly related to the work in this dissertation, we decided not to include information on the other levels of the general framework. For a general description on object tracking, behavior understanding/description and personal identification, the reader is referred to the work in [14, 15, 17, 18].

1.5 Research question

The central question of this dissertation is:

‘Can we develop an algorithm to timely and accurately detect/analyze the fire in large open public places and can we use the extracted fire characteristics for video driven fire forecasting?’.

In order to facilitate answering this question, it is helpful to break the question down into smaller parts. The main goal of this dissertation is to give an answer to each of these parts. The first part, i.e., **‘can we develop an al-**

gorithm to detect the fire', is already answered by many authors. In the last decade, the literature/research on video-based fire detection is growing rapidly, and this has led to several VIFD and VISD algorithms that can be used to detect the presence of fire at an early stage. However, due to the variability of shape, motion, transparency, colors and patterns of smoke, existing VFD approaches are still vulnerable to missed detections and false alarms. The main cause of both problems is the fact that visual detection is often subject to constraints regarding the scene under investigation, e.g., changing environmental conditions, and the target characteristics. To avoid the disadvantages of using visual sensors alone, we believe that the use of other types of sensors, such as infrared (IR) and time-of-flight (TOF*), can be of added value. Chapter 2 mainly focuses on this aspect.

The second part of the research question, i.e., '**in large open public places**', makes the VFD problem even more difficult. Videos from this kind of surveillance scenes, e.g., car parks, shopping malls and atria, often contain difficulties, such as changing/limited illumination, shadows and noise, which makes the detection error-prone. In order to achieve high accuracy for these kind of scenes, we propose several low-cost multi-modal fire detectors (Chapter 3) which are able to cope with many of the sensor-related 'limitations'. The main benefit of (f)using multi-modal image data is that unreliably extracted parts from one sensor might be reliably extracted from the other sensor. By using the strengths of each medium, fire detection can be done more accurately.

Within the context of a fire, 'some seconds' can make a huge difference. This is also illustrated in Fig. 1.3, which illustrates the timeline of a fire. To avoid large scale fire/smoke damage it is important to **timely** detect the fire. Compared to traditional point sensors, which suffer with a transport delay, volume sensors (like cameras) are able to detect smoke/flames as soon as they occur in the cameras field of view. Realistically, however, a fire detection algorithm in commercially available video surveillance systems has to run in parallel with many other surveillance processes. For many of the state-of-the-art fire detection techniques, real-time operation requires significant computational resources, which cannot always be guaranteed. As such, these methods will have difficulties to timely detect the fire. By focusing on (computational) 'low-cost' fire features, the methods proposed in this dissertation are able to keep the processing cost low, i.e., to run in real-time (~ 25 fps). As such, they are able to ensure real-time detection, even when they are combined in a multi-modal or multi-view setup. This is confirmed by recent experiments which were performed in cooperation with Xenics [27]. As soon as smoke or flames appear in the field of view of the camera, fire alarm is given.

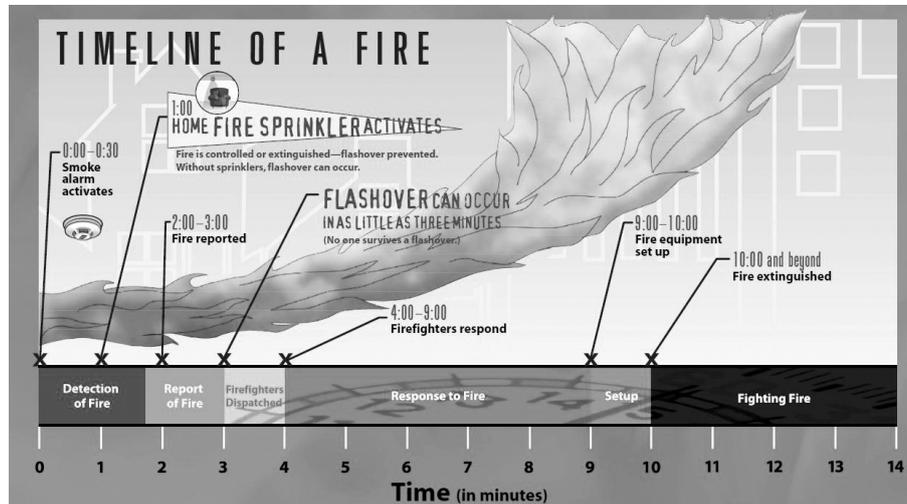


Figure 1.3: Timeline of a fire (image from http://publicsafety.utah.gov/firemarshal/FMdocs/timeline_of_a_fire.pdf).

Related to the ‘timely’ operation is the **accuracy** aspect of the research question. A general definition of accuracy, however, does not exist. It depends on the context, environment and/or application in which the detectors are used. A fire detector is said to be accurate if it detects all the fires, with as few false alarms as possible. Missing a fire is much worse than detecting a non-fire event (false alarm). As such, the recall (i.e., the probability of detecting an item given that it is relevant) is the most important for fire detection. However, although it is (almost) not used as an evaluation metric in video fire literature, precision (i.e., the probability that an item is relevant given that it is detected by the algorithm) is also important. This is mainly because too many false alarms will make the operators less attentive.

** 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 amplitude information. The ability to describe scenes using a depth map and an amplitude image, provides new opportunities in different applications, including video surveillance (object detection, tracking and recognition) and human computer interaction (e.g., gaming). The possibilities of TOF based fire detection have not yet been investigated. As such, the TOF-based detectors which are presented in this dissertation are a first attempt in this direction.*

In general, a fire detection system will be more valuable if it has a higher detection rate (\sim recall) and if its number of false alarms are lower. Secondly, a system its value also depends on what the system is able to do, i.e., the system functionality. For example, the proposed intelligent fire detection system is able to detect, analyze and forecast the fire. Compared to the SOTA systems, which are (mainly) limited to detection, our system is, as such, more ‘valuable’. Lastly, a system its value can also be evaluated based on its **computational efficiency**. In order to evaluate this computational efficiency, one must be able to test the systems using a standard dataset and with standardized evaluation metrics. For video based fire detection, however, both do not exist (yet). This is also the reason why the details on computational efficiency throughout this dissertation are limited.

The fourth aspect of the research question concerns ‘**video fire analysis**’. A study of the literature revealed that the amount of research in this direction is limited. Even today, most video-based fire alarm systems just ring the bells, i.e., they only detect the presence of fire and are not able to model fire evolution. Even though the majority of these systems consist of several cameras monitoring the same scene, the analysis is usually carried out separately on each of the camera’s video sequences. In order to perform more accurate detection and localization of smoke and flames, and to detect valuable fire characteristics at the early stage of the fire, we propose to combine the detection results of each of these single-view cameras and analyze them together into our novel multi-view fire analysis framework (Chapter 4). The main goal of the framework is to provide a more valuable video fire analysis tool than the existing SOTA work, which results are still limited and interpretation of the provided information is not straightforward.

The fifth and last part of the research question is: ‘**can we use the extracted fire characteristics for video driven fire forecasting**’. To the best of our knowledge, we are the first to tackle this subject. The proposed work on video driven fire spread forecasting should be seen as a first step in the direction of an application aiding firefighters in assessing the fire risk more efficiently. It should be able to serve as a solid basis for the extraction of fire characteristics and how this information can be used to estimate the future state of the fire. Initially, real-time estimations of smoke layer height and fire size will be used to accelerate fire models. The main reason for choosing these characteristics is that they can be extracted rather easily from video images. The proposed video driven fire forecasting is a prime example of how video-based detectors will be able to do more than just generate alarms. Detectors can give information about the state of the environment, and using this information, predictions of

the future state can be improved/accelerated. By combining information about the fire from models and real-time data an estimate of the fire can be produced that is better than could be obtained from using the model or the data alone.

Related to the research question, it is also important to have a clear understanding of the requirements to which the proposed VFD and video fire analysis techniques must adhere. As such, we briefly discuss the system requirements in the following section.

1.6 System requirements

As it is important to develop a system which meets the needs expressed by partners in the field [28], the following list of requirements is defined:

- Easy (re-)calibration: automatic registration (\sim alignment) of multi-sensor/multi-view images.
- Low computational cost: the flame and smoke detectors and the fire analysis framework must be able to run in real-time.
- Low number of false alarms and no missed detections.
- Fast warning / alarming with different levels of detection: higher levels of detection should only be activated if the global ‘fire risk value’ is high.
- Sequence/scene independent with low number of thresholds.
- Low purchasing price / installation costs.

Furthermore, the proposed algorithms should be generally applicable, i.e., they should be easy to adapt when the environment changes or when they are applied in a slightly different scene. It is preferred to use algorithms for which the parameters/thresholds are easy to set. This means that either the performance of the algorithm should not rely heavily on the value of the parameter, or we should be able to deduct the parameter value from real world variables. Finally, it is also important to note that we impose the restrictions that the scene is intended to be recorded using static cameras, that the cameras in a multi-view setup overlap and that the lines of sight of the cameras in a multi-modal setup are close to each other.

1.7 Outline

This dissertation is organized as follows. Chapter 2 presents the work done in single-view flame and smoke detection. We present (computational) low-cost detectors based on visual, thermal and time-of-flight imaging. First, we perform a thorough study of related work within the domain of video fire detection. Next, we give an in-depth overview of our single-view detectors and analyze them by including a comparison with related work. Both an objective (by evaluating the correctness of the detections against a manually created ground truth) and subjective (visual examples) evaluation is shown.

In Chapter 3, we present a set of novel multi-modal flame and smoke detection techniques that combine the different kinds of information provided by visual, thermal, and depth sensors. Again, we elaborate on related work, presenting several multi-modal vision applications, with a strong focus on the underlying techniques that can be of use for multi-modal fire detection. The registration of multi-modal images is also briefly touched and important concepts for this dissertation are explained. Subsequently, we present our multi-modal flame and smoke detectors. Next, as in Chapter 2, we compare the proposed algorithms with related work and discuss the performance (in terms of objective and subjective results).

Chapter 4 focuses on multi-view object localization and shows how this is applicable to automatic fire analysis. Our work to solve the fire localization problem is done in the domain of homography based plane slicing. We present a novel framework for video-based fire analysis which, in a later stadium, can also be used for fire forecasting. The framework fuses low-cost video fire detection results of multiple cameras into a grid of virtual sensor points, called the 'FireCube'. Using the FireCube, the location of the fire, its size, its propagation and its direction can accurately be estimated. Finally, Chapter 5 lists the conclusions of this dissertation and points out directions for future work.

Overview publications

The work described in this dissertation has resulted in a number of publications listed in the Science Citation Index: one paper is published in Elsevier's Fire Safety Journal and another paper is accepted for publication in Springer's Machine Vision and Applications (both as a first author). Four other papers are under review with Springer's Multimedia Tools and Applications, Sage's Journal of Fire Sciences, Springer's Fire Technology and Elsevier's Fire Safety Journal respectively. Our work also contributed to a book chapter (as a first author) in Intech's Video Surveillance book. Next to this, the work described in this dissertation contributed to 16 papers as first author and 7 as co-author, which were presented at international conferences. Lastly, our work is also highlighted in an annual report and newsletter of the EGOLF (The European Group of Organizations for Fire testing, Inspection and Certification) and IAFSS (International Association for Fire Safety Science) community. A detailed list of all the publications can be found in the Publications section at the end of this dissertation.

On November 17th 2011, the proposed work also won the Fireforum Award 2011. More info/photos can be found on the Fireforum website: <http://www.fireforumawards.be/>.

Chapter 2

Video fire detection

Based on the underlying techniques of the state-of-the-art (SOTA) algorithms and the results of real-world experiments, we propose a set of novel video fire detection (VFD) algorithms. Our main contribution is the exploration of the added value of infrared (IR) and time-of-flight (TOF) fire detection. The latter one, i.e., TOF fire detection, is the first attempt ever. In order to keep the processing cost low, i.e., to ensure real-time detection, a set of computational ‘low-cost’ fire features, which uniquely describe smoke and flames, is selected for each sensor type individually. Experiments show that this feature based approach gives good results for each of the proposed single sensor detectors.

2.1 Introduction

Video processing techniques for automatic flame and smoke detection have become a hot topic in computer vision during the last decade. Several vision-based 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. Section 2.2 focuses on the SOTA of these VFD algorithms in the visible spectral range. Due to noise, shadows, illumination changes and other visual artifacts, developing a reliable VFD system is, however, shown to be a huge challenge when only using ‘ordinary’ video. As such, the use of other types of sensors started to be explored in the last decade. Instead of dealing with ever more complex visual fire detection algorithms, these new approaches perform fire detection using IR imaging sensors, which operate in short-, mid- or long-wave infrared spectral range. The research in these spectral ranges is also discussed in our SOTA-study. As we believe, and will experimentally show, that a TOF sensor can also be used for fire detection, the SOTA study finishes with an overview on TOF based video surveillance.

Section 2.3 proposes our novel (computational) low-cost visual flame and smoke detection algorithms. In order to ensure real-time detection, the computational cost of both algorithms is kept as low as possible. They both consist of only two building blocks: a moving object detection and a set of ‘low-cost’ fire features, which uniquely describe smoke and flames. By analyzing the values of these features, a fire alarm can be raised. Subsequently, Section 2.4 presents our thermal long-wave infrared (LWIR) flame detector, which mainly reuses most of the ‘visual’ building blocks. Additionally, the moving object detection is extended with a hot object segmentation step to extract the hottest objects out of the set of LWIR foreground (FG) objects. Only these hot FG objects are further analyzed using the set of LWIR flame features. A fire alarm is also raised on the basis of the values of these features.

Our SOTA study of visible and IR fire detectors shows that most detectors start from simple background (BG) subtraction in spatial domain, e.g., frame differencing and running average. Hence, the influence of the BG model in VFD is not yet fully explored. As such, Section 2.5 describes a first attempt in this direction and investigates the added value of a discrete wavelet transform (DWT) based BG subtraction method for VFD.

Next, Section 2.6 proposes our novel TOF based flame detector. To the best of our knowledge this is the first attempt to develop a fire detection system based on the use of a TOF depth sensor. The proposed detector focuses on both the depth and amplitude image of a TOF camera. Using this multi-modal information, experiments have shown that flames can be detected very accurately. At the end of this section, first steps towards TOF-camera based smoke detection are also discussed. Finally, Section 2.8 finishes this chapter and lists the conclusions and suggestions for future work.

2.2 State-of-the art in video fire detection

2.2.1 Video fire detection in visible/visual spectral range

Over the last years, the number of papers about visual fire detection in the computer vision literature is growing exponentially [1]. As is, this relatively new subject in vision research is in full progress and has already produced promising results. A chronological overview of the state-of-the-art, i.e., a collection of frequently referenced papers, is presented in Tables 2.1, 2.2 and 2.3. For each of these papers we investigated the underlying algorithms and checked the appropriate techniques. In the following, we will discuss each of these detection techniques and analyze their use in the listed papers.

Table 2.1: State-of-the-art: underlying techniques (PART1: 2002-2007).

Smoke detection		X	X	X					X	X	X
Flame detection	X			X	X	X	X			X	X
Localization/analysis											
Cleanup post-processing	X										
Training (models, NN, SVM, ...)	X										
Subblocking											
Disorder analysis		X	X	X	X	X		X			X
Dynamic texture/Pattern analysis											
Spatial difference analysis	X							X			
Flicker/Energy (wavelet) analysis	X	X	X		X	X	X	X	X		X
Moving object detection		X	X	X				X	X		X
Color detection	RGB			RGB/HSI	HSV	YUV	RGB	YUV	YCbCr/RGB		
Paper	Phillips [29], 2002	Gomez-Rodriguez [30], 2002	Gomez-Rodriguez [31], 2003	Chen [32], 2004	Liu [33], 2004	Marbach [34], 2006	Toreyin [35], 2006	Toreyin [36], 2006	Celik [37], 2007	Xu [38], 2007	

Table 2.2: State-of-the-art: underlying techniques (PART 2: 2007-2009).

Smoke detection		X	X	X	X	X			X	X
Flame detection	X		X				X	X		
Localization/analysis									X	
Cleanup post-processing	X		X					X		
Training (models, NN, SVM, ...)	X		X	X	X					X
Subblocking				X		X			X	X
Disorder analysis	X	X				X	X		X	
Dynamic texture/Pattern analysis										
Spatial difference analysis								X		
Flicker/Energy (wavelet) analysis		X		X	X			X		X
Moving object detection	X	X	X	X	X	X			X	
Color detection	RGB		RGB	RGB	RGB	RGB	RGB	RGB/HSV	RGB/HSI	
Paper	Celik [39], 2007	Xiong [40], 2007	Lee [41], 2007	Calderara [42], 2008	Piccinini [43], 2008	Yuan [44], 2008	Borges [45], 2008	Qi [46], 2009	Yasmin [47], 2009	Gubbi [48], 2009

A. Color detection

Color detection was one of the first detection techniques used in VFD and is still by far the most popular. The majority of the color-based approaches in VFD makes use of RGB color space, sometimes in combination with HSI/HSV saturation [32, 46, 49, 50]. The main reason for using RGB is the equality in RGB values ($R \approx G \approx B$) of smoke pixels and the easily distinguishable red-yellow range ($R \geq G \gg B$) of flames, as is shown in Fig. 2.1. The major rule-based techniques used to detect the fire colored pixels are Gaussian-smoothed color histograms [29], statistically generated color models [37], and blending functions [42]. Although the test results of color-based fire detection in the referenced work seems promising at first, the variability in color, density, lighting, and background do raise questions about its applicability in real world detection systems, especially for smoke detection. A far more interesting color-based smoke detection mechanism seems the detection of chrominance decrease [36], which is also used by our visual flame detector presented in Section 2.3.2.

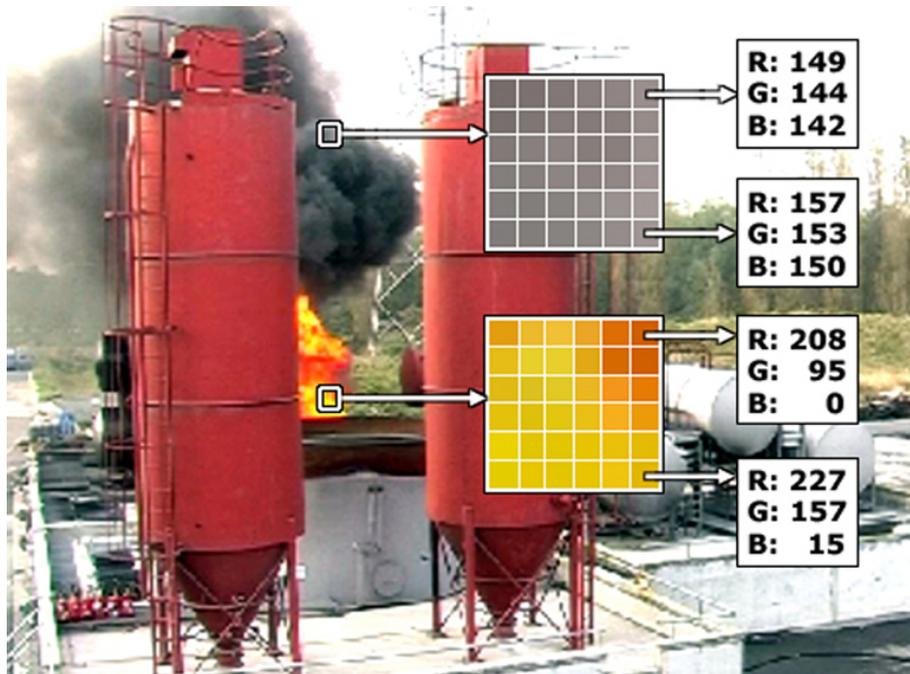


Figure 2.1: Color detection: smoke and flame crop show smoke RGB equality and easily distinguishable red-yellow range ($R \geq G \gg B$) of flames. (sequence from the IBBT ISYSS project [59])

B. Moving object detection

Moving object detection is the second technique that is frequently used as a first step in VFD to eliminate the disturbance of stationary non-smoke objects. In order to detect possible motion, which may be caused from fire, the moving part in the current video frame is detected by means of a motion segmentation algorithm. To determine if the motion is due to smoke or an ordinary moving object, further analysis of moving regions is necessary. The most effective algorithms to perform moving object detection are background (BG) subtraction [35, 36, 38–40, 42–44, 49, 50, 52, 56], temporal differencing [41], and optical flow analysis [30, 31, 51].

In Fig. 2.2, an example is shown of BG subtraction using the dynamic BG model proposed by Collins et al. [60]. This model has already been used in many of the works listed in Tables 2.1, 2.2 and 2.3, and is discussed in detail in Section 2.5. As can be seen, this simple BG subtraction performs well for ordinary moving objects. The moving person can easily be extracted from the background and the small illumination problems on the ceiling can be removed using clean-up post-processing techniques. For flames, however, the background subtraction does not generate appropriate results, as it also detects the reflections of the flames and the illumination change of the scene as part of the moving object(s). This is probably one of the most challenging visual-related problems video fire analysis is faced with in indoor environments.

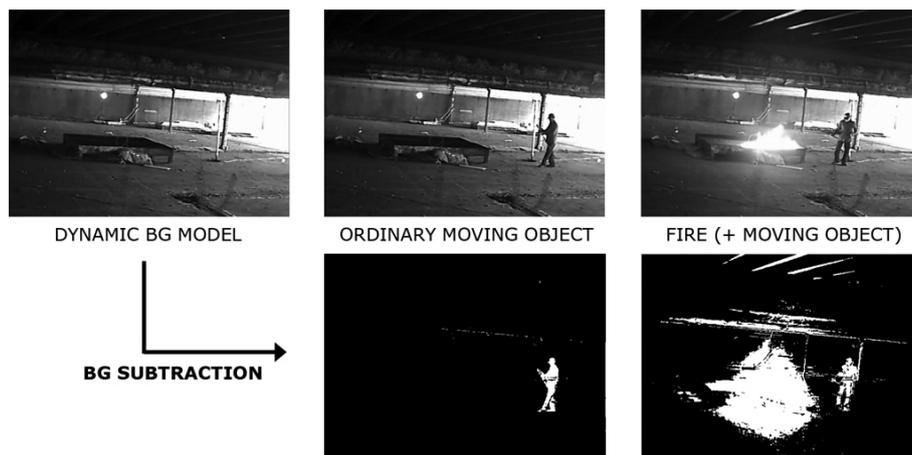


Figure 2.2: Moving object detection: background subtraction using dynamic background model. (*sequence from the CAR PARK FIRE SAFETY project [28]*)

C. Flicker/Energy (wavelet) analysis

Other frequently used fire detection techniques are flicker detection [34, 35, 40, 46, 49, 50, 52] and wavelet-based energy analysis [36, 42, 43, 48, 53, 61]. Both focus on the temporal behavior of flames and smoke. Flickering is the temporal periodicity with which pixels appear and disappear at the edges of turbulent flames. For turbulent flames, the research in [38, 40] shows experimentally that the flicker frequency is around 10Hz and that it is not greatly affected by the burning material and the burner. As such, these works propose to use frequency analysis to differentiate flames from other moving objects. However, due to the time-consuming transformation from the time domain to the frequency domain, and due to the fact that Yang and Wang [62] and Toreyin et al. [36] observed that the flame flicker process is far from periodic, i.e., flames generally oscillate with a frequency in the range of 0.5 - 20Hz, the time domain analysis of the oscillation frequency by Chen et al. [49] seems more appropriate. For smoke, the flicker frequency is even more time-varying. As such, smoke flicker detection does not seem to be a very reliable technique.

More interesting for detecting the temporal behavior of smoke is wavelet based energy analysis. As smoke gradually smoothens the edges in an image, Toreyin et al. [36] found the energy variation between background and current image as a clue to detect the presence of smoke. In order to evaluate the energy variation, they use the Discrete Wavelet Transform (DWT). The DWT is a multi-resolution decomposition of the image obtained by convolving the intensity image with several filter banks. The 9/7 Daubechies-based DWT decomposition, shown in Fig. 2.3, produces four wavelet subimages: the compressed version of the original image C_t , and the horizontal, vertical and diagonal high frequency images H_t , V_t , and D_t . The energy (Eq. 2.1) is evaluated blockwise dividing the image I_t in blocks b_k of arbitrary size, and summing up the square contribution of each high-frequency, i.e., high detail, wavelet subimage:

$$E(b_k, I_t) = \sum_{i,j \in b_k} H_t^2(i, j) + V_t^2(i, j) + D_t^2(i, j). \quad (2.1)$$

As the energy value of a specific block varies significantly over time in the presence of smoke, temporal analysis of the ratio between the current input frame energy and the background energy is used to detect the smoke (Fig. 2.3). This wavelet-based energy analysis is also used by the visual smoke detector proposed in Section 2.3.2.

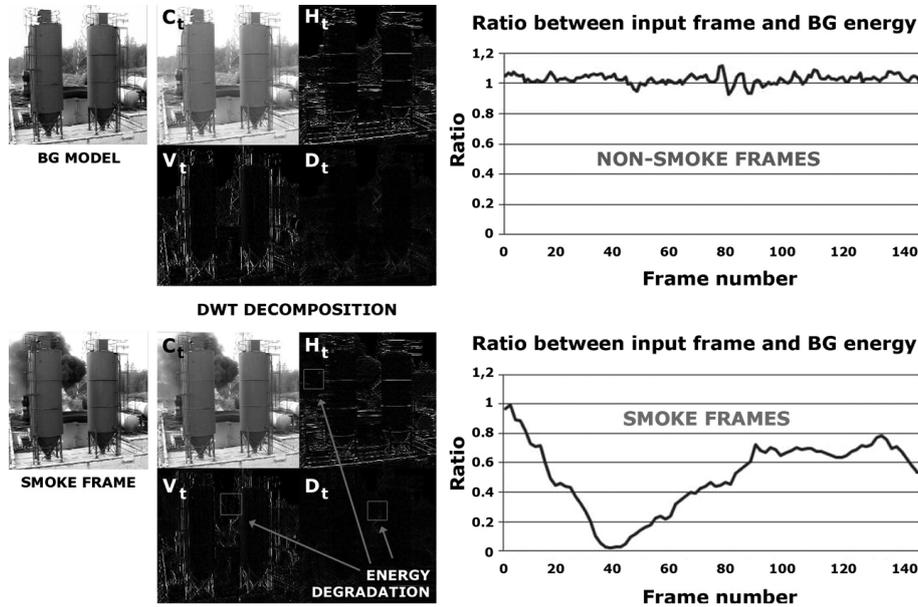


Figure 2.3: DWT based energy analysis: in case of smoke the ratio between the input frame energy and the BG energy decreases and shows a high degree of disorder.

D. Spatial difference analysis

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 [35, 46, 50, 54] focuses on this characteristic. Using range filters [46], variance/histogram analysis [54], or spatial wavelet analysis [35, 50], the spatial color variations in pixel values are analyzed to eliminate ordinary fire-colored objects with a solid flame color. In Fig. 2.4 the concept of spatial difference analysis is further illustrated by means of a histogram based approach, which focuses on the standard deviation of the green color band. It was found by Qi and Ebert [46] that this color band is the most discriminative for recognizing the spatial color variation of flames. This can also be seen by analyzing the histograms: green values vary more than red and blue values. If the standard deviation of the green color band exceeds $t_\sigma = 50$ (\sim Borges [54]), the region is labeled as candidate flame. For smoke detection, on the other hand, experiments revealed that these techniques are not always applicable, because smoke regions often do not show as high spatial color variation as flame regions. Furthermore, textured smoke-colored moving objects are difficult to distinguish from smoke and can cause false detections.

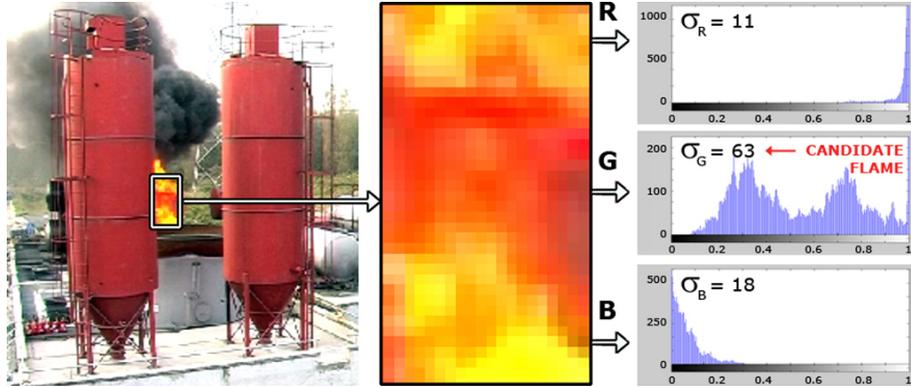


Figure 2.4: Spatial difference analysis: in case of flames the standard deviation σ_G of the green color band of the flame region exceeds $t_\sigma = 50$ (\sim Borges [54]).

E. Dynamic texture and pattern analysis

Dynamic texture and pattern analysis [51, 55, 57] is closely related to spatial difference analysis. Recently, these techniques are also gaining importance in flame and smoke detection. A dynamic texture or pattern, such as smoke, flames, water or leaves, can be simply defined as a texture with motion [63], i.e., a spatially repetitive, time-varying visual pattern that forms an image sequence with a certain temporal stationarity [64]. Although dynamic textures are easily observed by the human eye, they are difficult to discern using computer vision methods as the spatial location and extent of dynamic textures can vary with time and they can be partially transparent. Currently, geometric, model-based, statistical and motion based techniques are used for dynamic texture detection [65]. From all these approaches, the motion based techniques, such as the motion vectors based method in [66], are found the most appropriate [67].

In Fig. 2.5, examples are shown from the dynamic texture detection and segmentation algorithm by Fazekas et al. [64, 67, 68], which input videos were taken from the DynTex dynamic texture databases. [69]. Contours of dynamic texture regions, e.g., fire, water and steam, are shown in the figure. As the results show, the DynTex approach is promising, i.e., the dynamic regions are segmented very well. Due to its high computational cost, this technique is, however, not used by our low-cost algorithms. If future improvements could lower this computational cost, e.g., by using hardware accelerators, we encourage/plan to use this technique.



Figure 2.5: Dynamic texture detection: contours of detected dynamic texture regions are shown in the figure. (Results from *DYNTEX* dynamic texture detection [69])

F. Disorder analysis

Also interesting for fire detection is the disorder analysis of smoke and flame regions over time. Some examples of frequently used metrics are randomness of area size [45, 54], boundary roughness [33, 36, 50, 54], and boundary area disorder [40]. Although those metrics differ in definition, the outcome of each of them is almost identical. In our visible smoke detector (Section 2.3.2), disorder analysis of the Boundary Area Roughness (BAR) is used, which is determined by relating the perimeter of the region to the square root of the area (Fig. 2.6). Another, slightly different technique for disorder analysis is the histogram based orientation accumulation by Yuan [44]. This technique also produces good disorder detection results, but it is computationally more complex than the former methods. Related to the disorder analysis is the growing of smoke and flame regions in the early stage of a fire. In [53, 56], for example, region-of-interest growing is used to detect this fire related characteristic. Compared to disorder metrics, however, growth analysis is less effective in detecting the smoke.



Figure 2.6: Boundary area roughness of consecutive flame regions.

G. Subblocking, training, and clean-up post-processing

Although not directly related to fire characteristics, subblocking, training, and clean-up post-processing are three modules which are commonly used to simplify and improve the detection process.

Subblocking [42–44, 47, 48, 55] reduces measurement disturbances, i.e., filters out errors and measurements inaccuracies. Input images are subdivided in ‘ $n \times n$ ’ blocks, mostly 16x16 pixels, and a block value is computed as the average of all the pixel values in the block. Thereafter, further analysis is performed on block level instead of on pixel level.

Training is used to create background [39, 43, 49] and color models [29, 39, 49, 52, 56, 58], which augment ordinary moving object detection and color-based fire detection, as they use a probabilistic approach instead of a ‘naive’ threshold. Over the last years, more complicated training (\sim classifier) modules, such as Bayesian classifiers [52, 54], neural networks [51, 57], Markov models [50, 55] and support vector machines [48], are also started to be used. Due to their high computational cost, however, they are not covered in this dissertation.

Finally, clean-up post-processing, like median filtering, morphological operations [70], and clustering [58], is mostly used as a final step to remove outliers and to group neighboring elements.

H. Concluding remarks

In order to improve the detection performance, the majority of the referred works use a combination of the discussed fire features. Depending the fire/environmental characteristics, one combination of features will outperform the other and vice versa. Furthermore, it is also important to mention that, although the reported results in the SOTA show that ordinary video promises good fire detection results, our experiments revealed that vision-based detectors still suffer from a significant amount of missed detections and false alarms. The main cause of both problems is the fact that visual detection is often subject to constraints regarding the scene under investigation, e.g., changing environmental conditions, and the target characteristics. To avoid the disadvantages of visual sensors, the use of other types of sensors is started to be explored in the last decade. Instead of dealing with ever more complex visual fire detection algorithms, the majority of these new approaches performs fire detection using infrared imaging sensors.

2.2.2 Video fire detection in infrared spectral range

When light conditions are bad or the color of the target is similar to the background, IR vision is a fundamental aid. Even other visual-specific object detection problems, such as shadows, do not cause problems in IR [71]. Furthermore, due to the fact that IR imaging is heading in the direction of higher resolution, increased sensitivity and higher speed, it is already used successfully as an alternative for ordinary video in many video surveillance applications, e.g., traffic safety, airport security and material inspection. As manufacturers ensure reduction of price in time, it is even expected that this number of IR imaging applications will further increase significantly in the near future [72].

Although the trend towards IR-based video analysis is noticeable, the number of papers about IR-based fire detection is still limited [73–77]. As this is a relatively new subject in vision research, it has still a long way to go. Nevertheless, the results from existing work already seem very promising and ensure the feasibility of IR video in fire detection. Owrutsky et al. [73] work in the near infrared (NIR) spectral range and compare the global luminosity L , which is the sum of the pixel intensities of the current frame, to a reference luminosity L_b and a threshold L_{th} . If there are a number of consecutive frames where L exceeds the persistence criterion $L_b + L_{th}$, the system goes into alarm. Although this fairly simple algorithm seems to produce good results in the reported experiments, its limited constraints do raise questions about its applicability in large open uncontrolled public places.

Toreyin et al. [74] detect flames in IR 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 (HMM). The combination of both spatial and temporal clues seems more appropriate than the luminosity approach and, according to the authors, greatly reduces false alarms caused by ordinary bright moving objects. A similar combination of temporal and spatial features is also used by Bosch et al. [75]. Hotspots, i.e., candidate flame regions, are detected by automatic histogram-based image thresholding. By analyzing the intensity, signature, and orientation of these resulting hot objects' regions, discrimination between flames and other objects is made. The proposed IR-based fire detector (Section 2.4) mainly follows the latter feature-based strategy, but contrary to Bosch et al.'s work a dynamic background subtraction method is used, which is more suitable to cope with the time-varying characteristics of dynamic scenes. Also, by changing the set of features and combining their values into a global fire risk value, a decrease in computational cost is achieved with no negative effect on the detection results.

To conclude the SOTA study on IR fire detection, it is important to mention that IR imaging has its own specific limitations, such as thermal reflections, IR-blocking and thermal-distance problems. In some situations IR based detection will perform better than visible VFD, but under other circumstances, visible VFD can improve IR flame detection. As such, it is our strong belief that only by combining multi-modal video information higher detection accuracy can be achieved under all circumstances. Each sensor type has its own specific limitations, which only can be compensated by other types of sensors. This is further discussed in Chapter 3.

2.2.3 Video object detection using time-of-flight imaging

To the best of our knowledge, the TOF based flame and smoke detection, which is described further, is 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.

A. Time-of-flight based video surveillance

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 already 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 [78] explore a remedy to shadows and illumination problems in ‘conventional’ video surveillance by using range cameras. Tanner et al. [79] and Bevilacqua et al. [80] propose a TOF-based improvement for the detection, tracking and counting of people. Similarly, Grassi et al. [81] fuse TOF and IR images to detect pedestrians and to classify them according to their moving direction and relative speed. Tombari et al. [82] detect graffiti by looking for stationary changes of brightness that do not correspond to changes in depth.
- Image/video segmentation: In [83], Bleiweiss et al. state that the fusion of depth and color images results in significant improvements in segmentation of challenging sequences. For example, in tracking algorithms [84], segmentation is performed easily using a combination of a visual and a depth classifier, which is shown to be more functional in cluttered scenes.

- Face detection/recognition: Hansen et al. [85] improve the performance of face detection by using both depth and gray scale images; Meers et al. [86] generate accurate TOF-based 3D face prints, which are suitable for face recognition with minimal data and search times.
- (Deviceless) gaming: TOF imaging also increases the gaming immersion, as with this technology, people can play video games using their body as controllers. This is done by 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 Softkinetic, 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 camera, called Kinect.
- Human Computer Interaction: TOF cameras also pave the way to new types of interfaces that make use of gesture recognition [87] or the user's head pose and facial features [86]. 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 table-tops [88], which increase the attractiveness of board games.
- Other applications: e-health (e.g., fall detection [89]), interactive shopping and automotive applications (e.g., driving assistance and safety functions such as collision avoidance [90,91]).

B. Time-of-flight imaging: working principle

The working principle of TOF imaging is shown in Fig. 2.7. 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 roundtrip time t of the signal. If the emitter and the receiver are punctual and located at the same place, then Δt allows us to measure the depth of each pixel, as the depth $d = c\Delta t/2$, where c is the signal's speed ($c \simeq 3 * 10^8 m/s$ for light). Simultaneously, the camera also measures the strength of the reflected infrared signal, i.e., its amplitude, which is an indicator about the accuracy of the distance estimation [87].

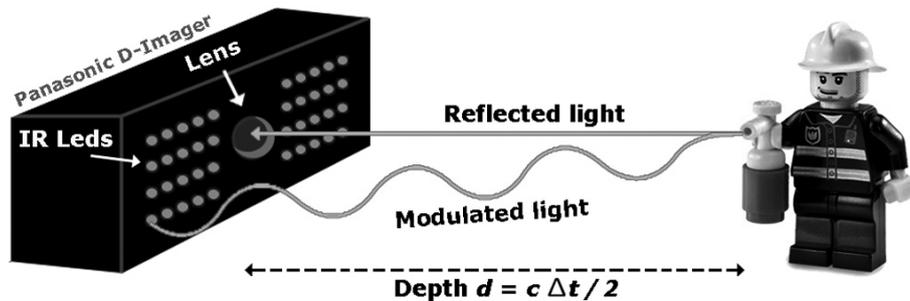


Figure 2.7: Working principle of TOF imaging: Modulated light is emitted from IR LEDs on the sensor. Light is reflected on the object and captured by the sensor. The time between emission and reception and the measured amplitude is used to generate the depth and the intensity images.

As the depth and amplitude information is obtained using the same sensor, the depth map (Fig. 2.8a) and the amplitude image (Fig. 2.8b) are aligned on each other (Fig. 2.8c). 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 a lot.
- Minimal amount of post-processing, giving application-processing more time for real time detection.
- The depth map, of which the information represents the physical properties of object location and shape, can help in dividing the objects during occlusion or partial overlapping [92].
- Low price compared to other IR-based video surveillance cameras.

In general, one can conclude that time-of-flight data compensates for the disadvantages and weaknesses, e.g., noise and other problematic artifacts, present in other data [83]. However, time-of-flight imaging also has its disadvantages:

- Low spatial resolution: The average commercially available TOF camera has a QCIF resolution (176×144 pixels), which is rather low. However, as with traditional imaging technology, resolution is increasing steadily with each new model.

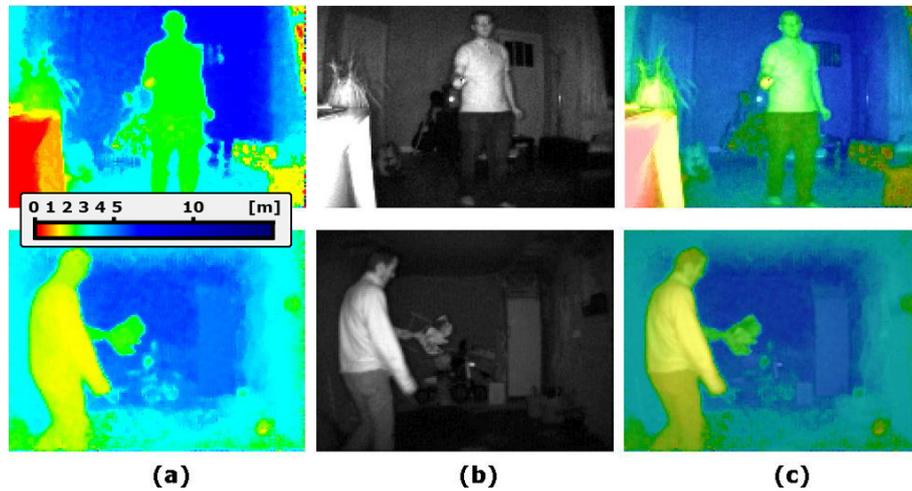


Figure 2.8: Correspondence matching between (a) TOF depth map and (b) amplitude image; (c) registration check. As a side note, this figure also shows a color-depth bar, which clarifies the meaning of the ‘depth colors’ in the depth maps.

- Measurement artifacts: Objects too close can be poorly illuminated leading to low quality range/depth measurements (Fig. 2.9a). Significant motion can also cause corrupt range/amplitude data, because the scene may change during consecutive range acquisitions. The sensor also has a limited ‘non-ambiguity range’ ($< 10m$) before the signals get out of phase (Fig. 2.9b). In small rooms, this is no problem, but in large rooms this can do 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, the proposed detectors can also focus on the IR emitted by the flames themselves. This way, the active illumination can (probably) be switched off.

C. Time-of-flight based fire detection?

Using a depth sensor like TOF camera to detect fire is not immediately intuitive, i.e., it is not obvious to link depth sensors, like TOF cameras, to fire detection. However, these kind of cameras focus on new object characteristics: depth, amplitude and reflectivity. As we expected that these ‘new’ characteristics could possibly be linked to flame characteristics, it seemed worth investigating this new image modalities. This was also the reason why we started

investigating thermal (LW)IR cameras. Furthermore, TOF cameras are not sensitive to light changes/shadows, need minimal amount of post-processing and have a low purchasing cost. Since they are also volume sensors, their information can easily be mapped on the other types of sensors that are investigated in our research.

Based on preliminary experiments with a Panasonic D-Imager [93], of which some exemplary TOF flame images are shown in Fig. 2.10, it is already possible to state that TOF cameras have great potential for flame detection in indoor environments. Flames produce a lot of measurement artifacts (i.e., TOF noise), which most likely can be attributed to the emitted infrared (IR) light of the flames themselves. Contrarily to ordinary objects, like people, the depth of flames changes very fast over time. Furthermore, the amplitude of the boundary pixels of flames shows a high degree of disorder. For smoke, on the other hand, the experiments do not directly show appropriate features. However, further testing (Section 2.6.2) will show that TOF-based smoke detection is also possible. Finally, the experiments show that in outdoor situations, outside the range of the TOF camera, and in case that smoke appears in the field of view of the TOF camera, the TOF depth map becomes unreliable. In order to cope with this problem, a multi-modal visual-TOF flame detector is proposed in the next chapter (Section 3.6).

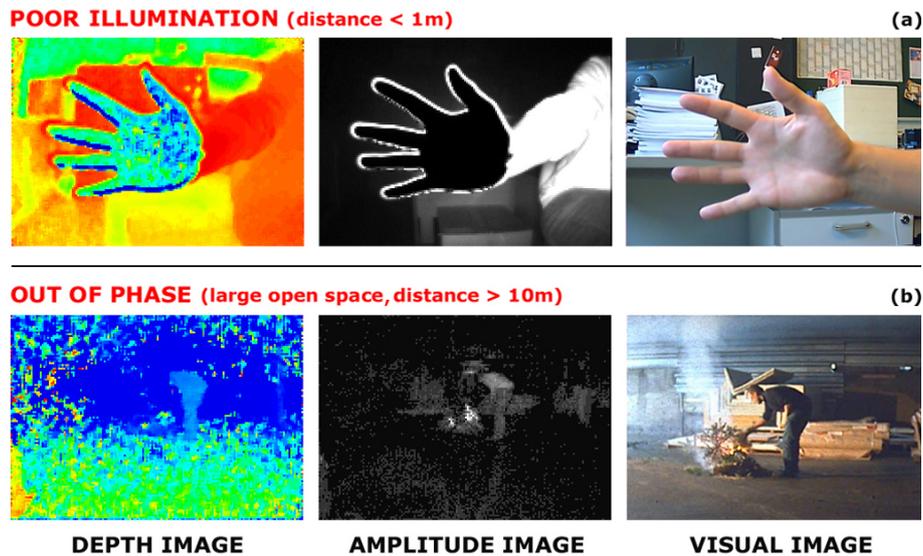


Figure 2.9: Measurement artifacts of TOF sensor: (a) poor illumination and (b) out of phase problem.

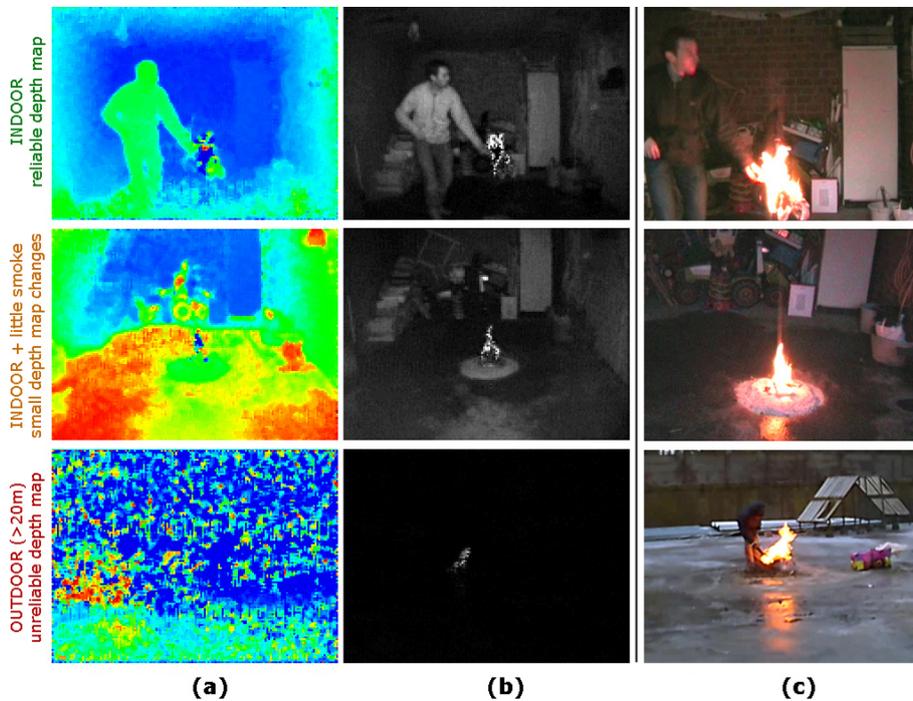


Figure 2.10: Exemplary TOF flame images: (a) depth maps and (b) corresponding amplitude images; (c) ordinary video (not registered).

2.3 Visual flame/smoke detection

Based on our experimental results [94,95] and the discussed state-of-the-art, a computational low-cost flame and smoke detection algorithm (Fig. 2.11) for visual VFD is presented in this Section. In order to keep the processing cost low, i.e., to ensure real-time detection, both detectors consist of only two building blocks: a moving object detection and a set of ‘low-complexity’ fire features, which uniquely describe smoke and flames. By analyzing the values of these features, a fire alarm can be raised. As will be shown, these (computational) low-cost algorithms yield good detection results, which are comparable to, and sometimes better than, the results of the investigated state-of-the-art. Hence, in addition to their low computational cost, they outperform the SOTA algorithms.

2.3.1 Computational low-cost flame feature analysis

The proposed flame detection starts with a *dynamic background (BG) subtraction* [35, 60], which extracts moving objects by subtracting the video frames with everything in the scene that remains constant over time, i.e., the estimated background. Next, to avoid unnecessary computational work and to decrease the number of false alarms caused by noisy objects, the *temporal filtering* removes objects which are not detected over multiple frames. Each of the remaining foreground (FG) objects in the video images is then further analyzed using a set of visual flame features, i.e., the *bounding box disorder*, the *principal orientation disorder* and the *flame color rate*. In the following subsections, detailed information is given on each of these visual flame features and how they are combined into a *global flame risk value*. For more detailed information on the BG subtraction, the reader is referred to Section 2.5.

A. Bounding box disorder (BBD)

Experiments (Fig. 2.12), on a set of fire and non-fire video sequences with varying environmental characteristics, revealed that the bounding box BB of flames varies considerably over time in both directions and that this variation shows a high degree of disorder. As such, the BBD (Eq. 2.2) is chosen as a feature to distinguish between flames and other ‘moving’ objects:

$$BBD = \frac{|\text{extrema}(BB_{1:N}^{\text{width}})| + |\text{extrema}(BB_{1:N}^{\text{height}})|}{N}. \quad (2.2)$$

The BBD is related to the number of extremes, i.e., local maxima and minima, in the set of N consecutive BB^{width} and BB^{height} data points (where a data point is the measured value at frame number t in time). By smoothing these data points using a moving average filter, small differences between consecutive points are filtered out and are not taken into account in the extrema calculation, which increases the strength of the feature. Flames, with a high number of extremes, will have a BBD close to 1, while for more static objects it will be near to 0. Important to mention is that all the flame and smoke features that are discussed in this chapter follow the same ‘feature strategy’, i.e., the feature values of flames/smoke objects will always be close to 1, while for more static objects it will be near to 0.

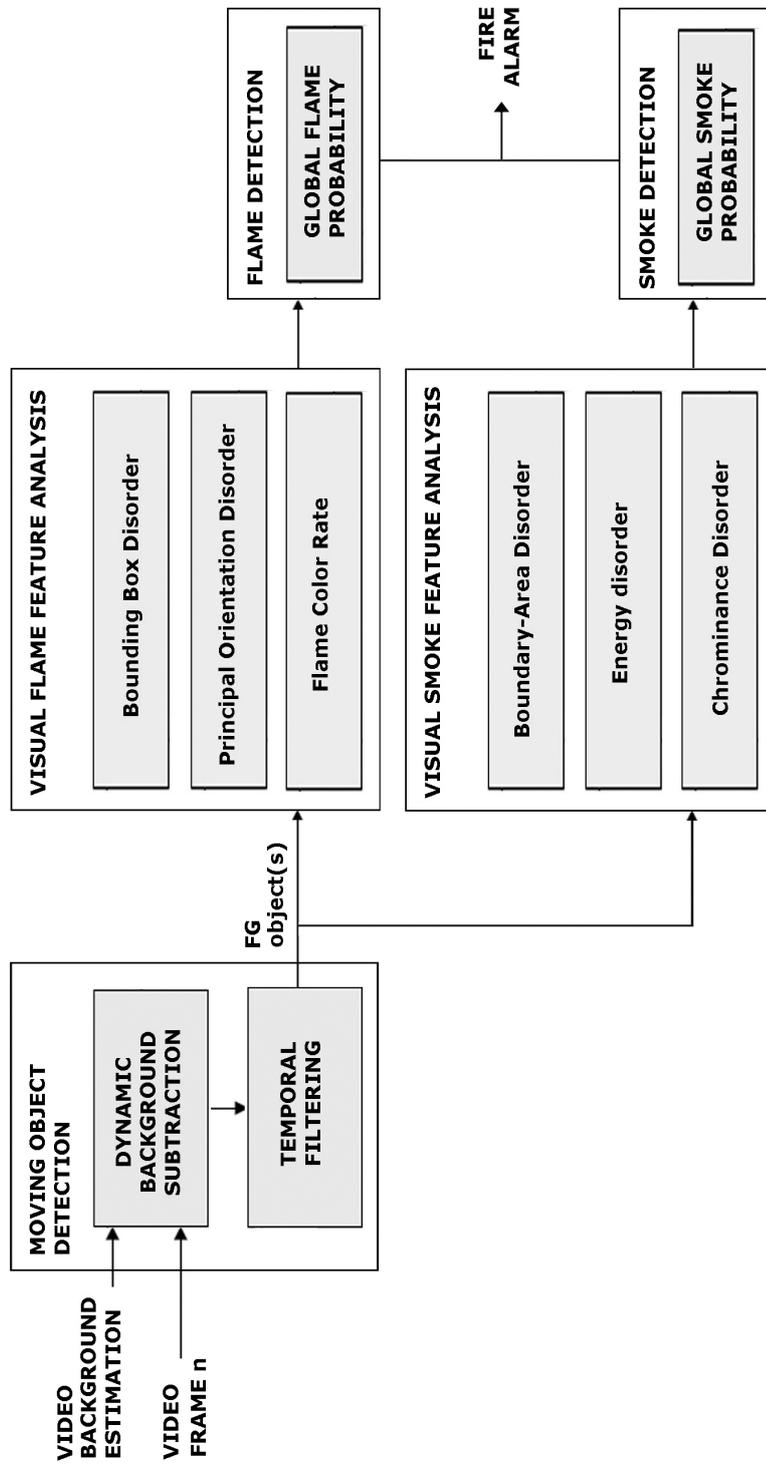


Figure 2.11: Low-cost visible flame/smoke detection.

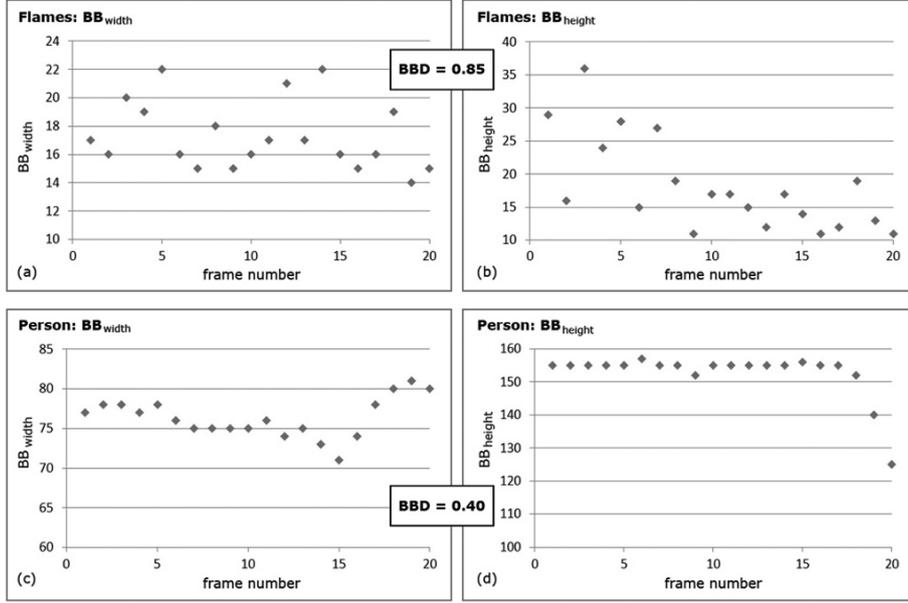


Figure 2.12: Bounding box disorder of flames (a,b) and moving person (c,d).

B. Principal orientation disorder (POD)

During our experiments on a set of fire and non-fire video sequences with varying environmental characteristics, we also found that the disorder in principal orientation is remarkably higher for flames than for more static objects like people (as is shown in Fig. 2.13). This orientation equals the angle α between the x-axis and the major axis of the ellipse that has the same second-moments as the object region. The *POD* (Eq. 2.3) focuses on this orientation disorder and is calculated in a similar way as the *BBD*:

$$POD = \frac{|extrema(\alpha_{1:N})|}{N/2}. \quad (2.3)$$

C. Flame color rate (FCR)

Based on our experiments and the work of others [32,46], it is also reasonable to assume that the color of flames belongs to the red-yellow color range. The flame color rate *FCR* focuses on this color-related aspect of flames in order to eliminate non-flame-colored objects. The *FCR* (Eq. 2.4) is defined as the ratio of the number of pixels $\#_{R-Y}(\Omega)$ within the red-yellow range ($R \geq G \gg B$) and the total number of pixels $\#_{pixels}(\Omega)$ within the object region Ω :

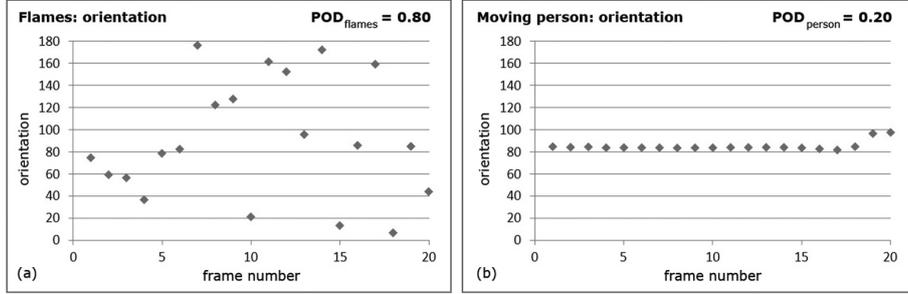


Figure 2.13: Principal orientation disorder of flames (a) and moving person (b).

$$FCR = \frac{\#_{R-Y}(\Omega)}{\#_{pixels}(\Omega)}. \quad (2.4)$$

Furthermore, experiments showed that the flame color does not remain steady, i.e., flames are composed of several varying colors. If necessary, one can also incorporate this color-changing aspect of flames in order to better eliminate ordinary flame-colored objects with a solid flame color. However, since we want to keep the proposed algorithm low-cost, the detector already performs well and preliminary experiments show that the gain of using this color-changing feature is limited, this is not incorporated in our detector.

D. Global flame risk value

Each of the proposed visual flame features possesses a value between 0 and 1, indicating whether the object has the flame characteristic. The global flame risk value P_{flame}^{visual} (Eq. 2.5) combines these different features, i.e., equals the average of the three feature values, and indicates whether the object should be classified as flames. P_{flame}^{visual} is defined as follows:

$$P_{flame}^{visual} = \frac{BBD + POD + FCR}{3}. \quad (2.5)$$

In our experiments it was found that a P_{flame}^{visual} of 0.7 is an appropriate value to generate the fire alarm. As such, the t_{flame} threshold is set to 0.7. If the P_{flame}^{visual} of an object exceeds t_{flame} , the frame in which it appears is labeled as fire. The sensitivity of t_{flame} , however, needs to be further investigated. Most important is that, based on the flame risk value, operators can concentrate their attention on the sequences which most probably contain flames.

2.3.2 Low-cost smoke feature analysis

As can be seen in (Fig. 2.11), the smoke detection starts with the same moving object segmentation and temporal filtering as the flame detection algorithm. Next, a set of low-cost visual smoke features, i.e., the *boundary-area disorder*, the *energy disorder* and the *chrominance disorder*, is used to analyze the remaining *FG* objects. Similar as for the *BBD* and the *POD* in the flame detection algorithm, the selected features vary considerably over time in case of a smoke object. As such, they are appropriate features to distinguish between smoke and other objects. Finally, the values of these features are combined into a *global smoke risk value*.

Due to the dynamic character of smoke, the perimeter and the area of smoke objects show a high degree of disorder [40]. By temporal analysis of the boundary-area roughness *BAR*, which focuses on the area *A* and perimeter *P* of the *FG* object (Eq. 2.6), this disorder can be detected. The boundary-area disorder *BAD* is calculated in a similar way as the flame *BBD* and *POD* features and is related to the number of extrema in the set of *N* consecutive *BAR* data points:

$$BAR = \frac{P}{2\sqrt{\pi A}}, \quad (2.6)$$

$$BAD = \frac{|\text{extrema}(BAR_{1:N})|}{N/2}.$$

The energy disorder feature focuses on the temporal behavior of the energy within smoke regions. In order to measure the energy *E*, the same discrete wavelet transform (DWT) based function as in the work of Calderara and Piccinini [42, 43] is used. In presence of smoke, the energy value of the smoke region varies significantly over time. Using the Energy Disorder *ED* (Eq. 2.7), which is related to the number of extrema in the set of *N* consecutive *E* data points, this energy disorder is measured:

$$ED = \frac{|\text{extrema}(E_{1:N})|}{N/2}. \quad (2.7)$$

Experiments also revealed that the chrominance values of smoke regions change a lot in the beginning of a fire. The chrominance disorder *CD* focuses on this chrominance-related behavior of smoke and is based on the number of extrema in the average chrominance value *C* of the smoke region:

$$CD = \frac{|extrema(\overline{C_{1:N}})|}{N/2}. \quad (2.8)$$

Each of the proposed smoke features also possesses a value between 0 (non-smoke) and 1 (smoke), indicating whether the object should be classified as non-smoke or smoke respectively. As such, the global smoke risk value can be calculated in the same way as the global flame risk value. The P_{smoke} (Eq. 2.9) equals the average of the boundary-area disorder BAD , the energy disorder ED and the chrominance disorder CD . Based on this global risk value, fire alarm is given if the risk value exceeds the smoke threshold t_{smoke} of 0.6, which was found experimentally. The sensitivity of t_{smoke} , however, also needs to be further investigated. P_{smoke}^{visual} is defined as follows:

$$P_{smoke}^{visual} = \frac{BAD + ED + CD}{3}. \quad (2.9)$$

2.3.3 Evaluation of (single-view) smoke and flame detector

The video images in Fig. 2.14 are some exemplary frames of the fire and non-fire realistic video sequences which were captured to test the proposed flame and smoke detection algorithm. As can be seen, different types of fires were investigated. This gives us the opportunity to ensure that the proposed detectors are suited for different environments and will work under different conditions.

In order to objectively evaluate the detection results of the proposed algorithms, and to compare them to other state-of-the-art methods, the ‘detection rate’ metric [94] is used. This metric is comparable to the evaluation methods used by Celik et al. [96] and Toreyin et al. [35]. The detection rate equals the ratio of the number of correctly detected frames as fire, i.e., the detected frames as fire minus the number of falsely detected frames, to the number of frames with fire in the manually created ground truth (GT).

Table 2.4 summarizes the detection results for all the tested sequences. As these results indicate, the combination of both algorithms yields good detection results, which are comparable to, and sometimes better than, the evaluated state-of-the-art (SOTA) results. The SOTA methods, which were chosen for comparison with the proposed smoke and flame detection algorithm (Method 1), are a combination of the flame detection method by Celik et al. [96] and the smoke detection by Toreyin et al. [36] (Method 2) and a combination of the feature-based flame detection method by Borges et al. [45] and the smoke detection method by Xiong et al. [40] (Method 3).

Table 2.4: Comparison of the proposed smoke and flame detection method (Method 1), the combined method based on the flame detector by Celik et al. [96] and the smoke detector described in Toreyin et al. [36] (Method 2), and combination of the feature-based flame detection method by Borges et al. [45] and the smoke detection method by Xiong et al. [40] (Method 3).

video sequence (# frames)	# fire frames <i>ground truth</i>	# detected fire frames			# false positive frames			detection rate *		
		method 1	method 2	method 3	method 1	method 2	method 3	method 1	method 2	method 3
Paper fire (1550)	956	897	922	874	9	17	22	0.93	0.95	0.89
Car fire (2043)	1415	1293	1224	1037	3	8	13	0.91	0.86	0.73
Moving people (886)	0	5	0	28	5	0	28	-	-	-
Wood fire (592)	522	510	489	504	17	9	16	0.94	0.92	0.93
Bunsen burner (115)	98	59	53	32	0	0	0	0.60	0.54	0.34
Moving car (332)	0	0	13	11	0	13	11	-	-	-
Straw fire (938)	721	679	698	673	16	21	12	0.92	0.93	0.92
Smoke/fog machine (1733)	923	834	654	789	9	34	52	0.89	0.67	0.80
Pool fire (2260)	1844	1665	1634	1618	0	0	0	0.90	0.89	0.88

* $detection\ rate = (\#\ detected\ fire\ frames - \#\ false\ alarms) / \#\ fire\ frames$

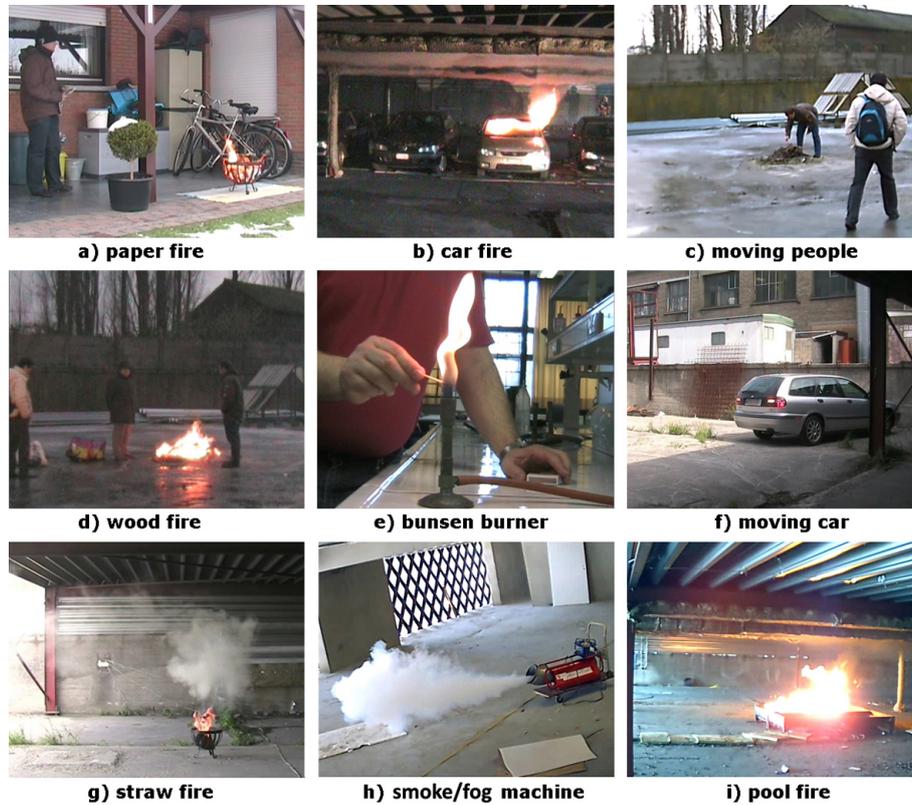


Figure 2.14: Fire and non-fire realistic test sequences.

The reason for making the comparison with the flame detection methods by Celik and Borges and the smoke detection methods by Toreyin and Xiong, is that these four methods are commonly referenced methods which contain similar techniques as those used by the methods proposed throughout this dissertation. However, in order to keep the computational cost low, the complexity of the techniques used within the proposed detectors is kept as low as possible, contrarily to those (sometimes) used in the investigated SOTA algorithms. Also important to mention is that switching the evaluated SOTA combinations has almost no effect on the combined detection results.

To end this section, Fig. 2.15 gives some examples of false positive frames, which were on the basis for investigating other image modalities such as time-of-flight and infrared (thermal) images. In these examples, the BBD and POD of flames are very low, due to the specific circumstances of the fire.



Figure 2.15: Examples of VFD false positive frames.

2.4 Infrared flame detection

The proposed infrared flame detector operates in the long-wave infrared (LWIR) range, as is discussed in Section 2.4.1. Similar to the visual flame detector, the detection algorithm (Fig. 2.16) starts with a *dynamic background subtraction* (Fig. 2.17a-c) and *morphological filtering*. Then, it automatically *extracts hot objects* (Fig. 2.17d) from the foreground thermal images by histogram-based segmentation (Section 2.4.2). After this thermal filtering, only the relevant hot objects in the scene remain in the foreground. These objects are then further analyzed using a set of three LWIR fire features: *bounding box disorder*, *principal orientation disorder*, and *histogram roughness* (Section 2.4.3). 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 values of these fast retrievable features we are able to detect the fire at an early stage.

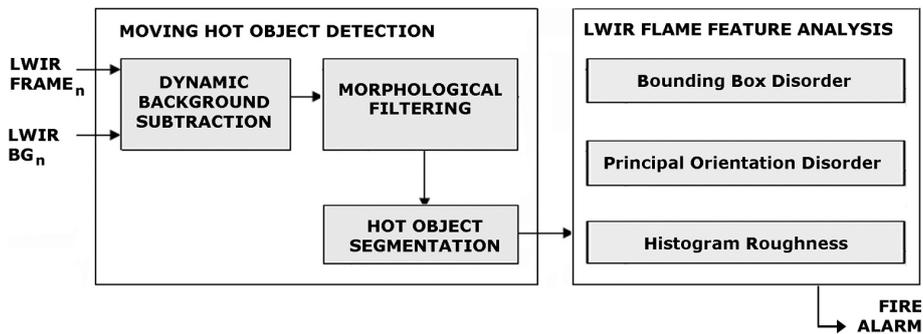


Figure 2.16: Low-cost LWIR flame detector: moving hot object detection and flame feature analysis.

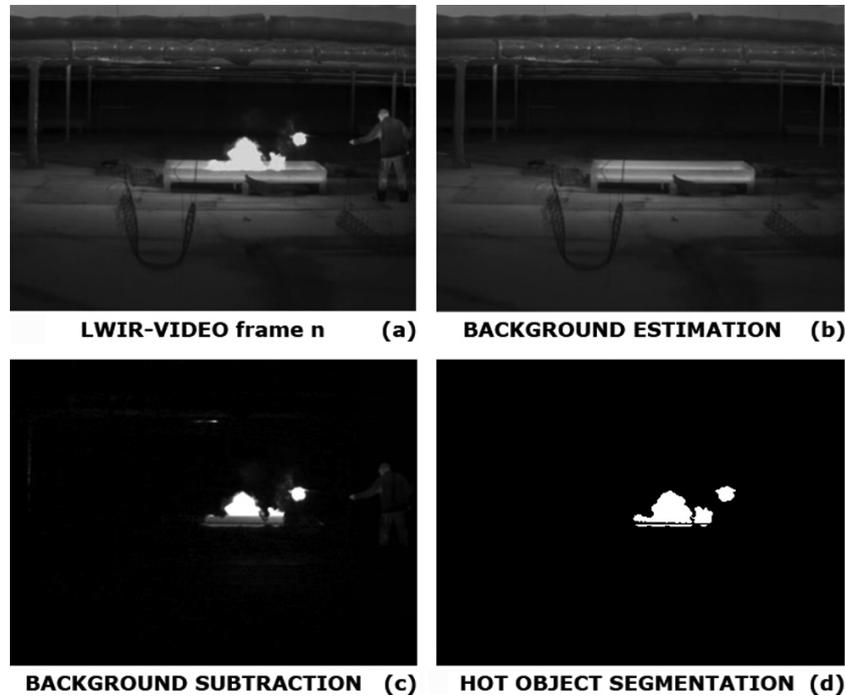


Figure 2.17: Hot object segmentation by BG subtraction and histogram-based dynamic thresholding.

2.4.1 Selection of the spectral range: SWIR, MWIR or LWIR?

Since SWIR is so near the visible bands, its behavior is similar to visible light. Energy in this band must be reflected from the scene in order to produce good imagery, which means there still must be some external illumination. MWIR and LWIR do not have this restriction since they image radiated energy. Also, the further we go in the infrared spectrum the more the visual perceptibility decreases and the thermal perceptibility increases. As such, hot objects like flames will be best visible and less disturbed by other objects in the LWIR spectral range (8-12 μm). Thermal cameras in this range only focus on the temperature: the warmer an object is, the brighter it appears on the images. Moreover these cameras are not sensitive to dust, smoke, and fog, making it possible to even see flames through the smoke.

2.4.2 Histogram-based hot object segmentation

Histogram-based segmentation is used in addition to BG subtraction to extract the hottest objects out of the set of LWIR FG objects. Only these hot FG objects are further analyzed using the set of LWIR flame features. Like in the work of Bosch [75], hot objects representing possible flames are extracted by separating the highly brightened objects from the less brightened objects (Fig. 2.17). This segmentation step uses Otsu's method [97], which automatically performs histogram-based dynamic image thresholding (Fig. 2.18), assuming that the image to be processed contains two classes of objects. Iteratively the optimum threshold t separating those two classes is calculated so that their combined spread, i.e., the intra-class variance, is minimal. The intra-class variance $\sigma_w^2(t)$ is given by:

$$\sigma_w^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t), \quad (2.10)$$

where weights ω_i are the probabilities/cardinality of the two classes separated by a threshold t and σ_i^2 are the variances of these classes.

For thermal images, the rather simple Otsu method is sufficient enough to achieve high accuracy under all circumstances. Of course, more recent thresholding techniques, like local adaptive thresholding [98,99], perform much better than Otsu's method when applied to challenging visual segmentation problems. However, thermal images do not suffer with the illumination/shadow related problems which are the main cause of Otsu's visual thresholding problems. So, because of its advantages of simple implementation and time saving, Otsu is chosen in our approach. Important to mention is that the Otsu algorithm assumes, like many histogram-based algorithms, that its input histogram is bimodal (which is mostly the case in our fire videos). However, this can be seen as one of its limitations as, for example, in the unimodal case segmentations problems can arise. In order to cope with these problems within our 'context', we suggest to quantize the histogram before performing the thresholding and to analyze its local maxima(s). If only one maxima exists, e.g., when only flames or only moving objects occur, no thresholding needs to be performed.

2.4.3 LWIR flame feature analysis

The LWIR flame feature analysis is mainly based on the same features as its visual counterpart, which was discussed earlier in this chapter. The first two features, i.e., the BBD and POD, are the same in LWIR as in visual. As such, they are only briefly discussed. The third feature, i.e., the histogram roughness (HR), is an LWIR specific feature, and is described more in detail.

The LWIR experiments revealed that the bounding box of LWIR flame objects also varies considerably over time in both directions and that this variation shows a high degree of disorder. As such, the BBD is also chosen as a feature to distinguish between flames and other hot objects in LWIR. Flames, with a high number of BB extremes, will have a BBD (Eq. 2.2) close to 1, while for more static hot objects it will be near to 0. During the experiments, we also found that the disorder in principal orientation is remarkably higher for flames than for more static hot objects. For this reason, the LWIR flame feature analysis also uses the POD feature, which focuses on this orientation disorder characteristic. Again, flames, with a high number of orientation extremes, will have a POD (Eq. 2.3) close to 1, while more static hot objects their POD will be near to 0.

By inspection of the histograms H of hot objects, we observed that histograms of flame regions are very rough (Fig. 2.19). Also, we found that the intensities of these regions range almost over the whole histogram, while for non-flame objects these intensities are more centered on some specific intensity bins and have a smaller range. The histogram roughness HR focuses on these two findings. As Eq. 2.11 shows, the HR equals the mean $range$ of the histogram multiplied by the average disorder over all the non-zero $bins$ (which is calculated by extrema analysis and is the indicator of the histogram roughness over time):

$$HR = \frac{\overline{range(H)}}{N} * \frac{|extrema_{bins \neq \emptyset}(H)|}{N/2}. \quad (2.11)$$

In Eq. 2.11, the $range$ is the number of bins over which the histogram is spread and the mean \overline{range} is the average of these ranges within a time window of N frames. The $extrema()$ function counts the local maxima and minima in the set of N consecutive data points.

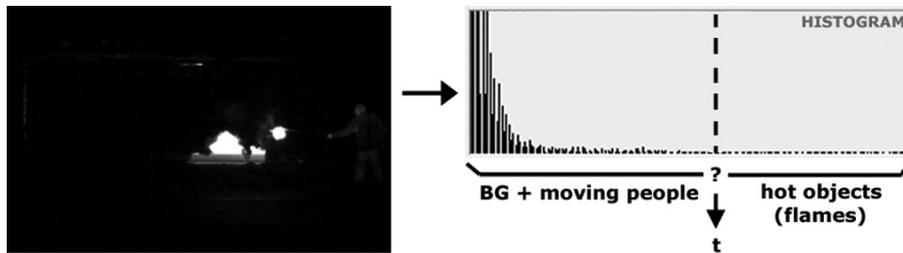


Figure 2.18: Histogram-based dynamic thresholding.

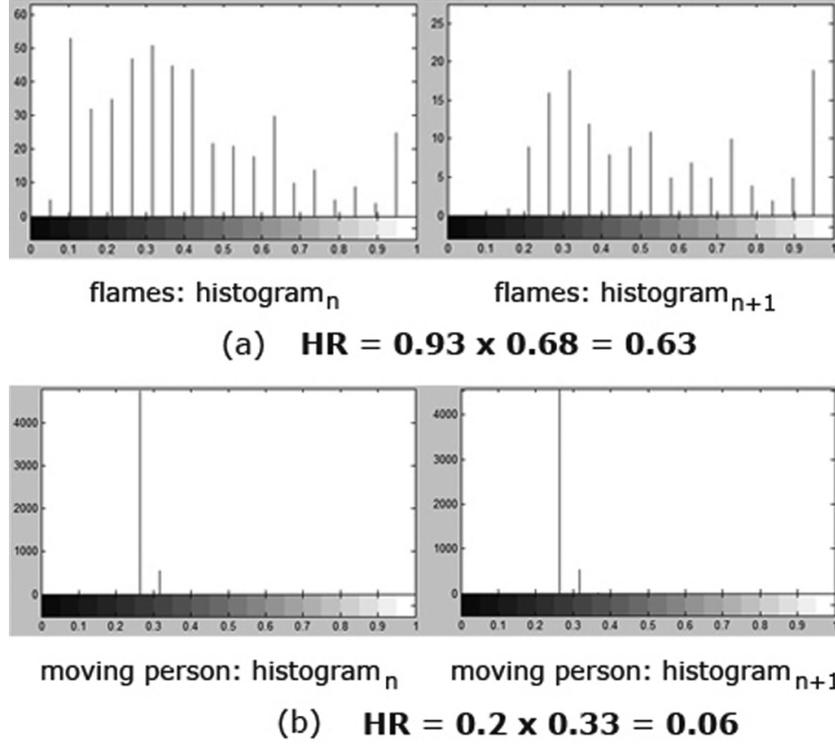


Figure 2.19: Histogram roughness of flames (a) and moving person (b).

Each of the proposed LWIR flame features also possesses a value between 0 and 1, indicating whether the object has the LWIR flame characteristic. The global LWIR flame risk value P_{flame}^{LWIR} (Eq. 2.12) combines these different features, i.e., equals the average of the three values, and indicates whether the hot object should be classified as flames:

$$P_{flame}^{LWIR} = \frac{BBD + POD + HR}{3}. \quad (2.12)$$

In our experiments it was found that a global flame risk value of 0.7 is an appropriate value to generate the fire alarm. As such, the t_{flame}^{LWIR} threshold is set to 0.7. If the P_{flame}^{LWIR} of an object exceeds this threshold, the frame in which it appears is labeled as fire.

2.4.4 Experimental results

The proposed LWIR flame detector was tested with a Xenics Gobi-384 LWIR camera [27], which works in the 8 - 14 μm spectral range. Using the Xenics Xeneth software we were able to extract appropriate grayscale video images out of the thermal imaging camera. These images were then further analyzed by our own LWIR detection algorithm. The LWIR images in Fig. 2.20 are some exemplary frames of the fire and non-fire realistic video sequences, which were captured to test the LWIR flame detection algorithm. For each of these sequences manually annotated ground truth (GT) was created, at which the automatic detection results are evaluated.

Table 2.5 summarizes the detection results for all the tested sequences. As the results indicate, the proposed algorithm already yields good detection results. For uncontrolled fires, e.g., burning paper, the flame detection rate is higher than 90% and for controlled fires, e.g., a Bunsen burner, it is around 75%. Furthermore, the number of false detections is very low. Although no real comparison is made with the discussed SOTA LWIR detection algorithms (Section 2.2.2), it is expected that the proposed method outperforms the luminosity based method by Owruisky et al. [73]. Although Owruisky's fairly simple algorithm seems to produce good results in the reported experiments, its limited constraints do raise questions about its applicability in large open uncontrolled public places. Since it is a global metric, an object heating up, like a radiator, can cause a false alarm. In general, this method does not focus enough on the (real) flame characteristics and just rings the bells (i.e., no further analysis is possible).

Compared to the object-based methods by Toreyin et al. [74] and Bosch et al. [75], which also combine both spatial and temporal clues, the proposed algorithm yields similar results. Contrary to these methods, however, the proposed algorithm is computationally less complex.

The results in Table 2.5 also show that the LWIR detector performs slightly better than our visual flame detector (Section 2.3.1). However, under other circumstances (e.g., when there are a lot of thermal reflections), the visible flame detection can improve the LWIR flame detection. As such, it is our strong belief that only by further analyzing the detection results using multi-modal VFD (\sim Chapter 3), a 'better' detector can be achieved providing high accuracy under all circumstances.

Table 2.5: Performance evaluation of LWIR-based flame detection.

Video sequence (# frames)	# fire frames (GT)	# detected fire frames	# false positive detections	LWIR flame detection rate	visible flame detection rate
Attic - fire (337)	264	255	9	0.93	0.91
Attic - fire and people (2123)	1461	1296	34	0.86	0.80
Attic - moving people (886)	0	14	14	-	-
Lab - Bunsen burner (115)	98	77	0	0.79	0.23
Corridor - hot object (184)	0	8	8	-	-
Car fire (1185)	733	682	5	0.92	0.78
Forklift (694)	0	7	7	-	-

* $\text{detection rate} = (\# \text{ detected fire frames} - \# \text{ false alarms}) / \# \text{ fire frames}$

2.4.5 Performance of LWIR cameras in a real fire environment?

To end this section, we briefly comment on the performance of LWIR cameras in a real fire environment. As most of our recordings were performed using cold smoke generated by a smoking machine or by controlled fires, we did (mainly due to security reasons) not yet investigate the impact of real fires on the camera technology. However, it is definitely something that needs further investigation. The same holds for the other types of cameras that are investigated throughout this dissertation.

If we ask the performance question to thermal imager manufacturers they ensure incomparable vision, even in very hot, smoke-filled environments (as could be expected). However, during the car fire tests of the car park fire project, the fire engineers from the Brandweer Vereniging Vlaanderen [100], highlighted some limitations of their thermal imager when the smoke layer became too hot. Do we see hot smoke/flames or is it the wall which is heating up? It was sometimes difficult to say. This triggered us to further investigate this topic and so we found an article [101] by the National Institute of Standards and Technology (NIST). In this NIST Technical Note an answer is given to this performance-related question of LWIR (thermal) cameras in a real fire environment.

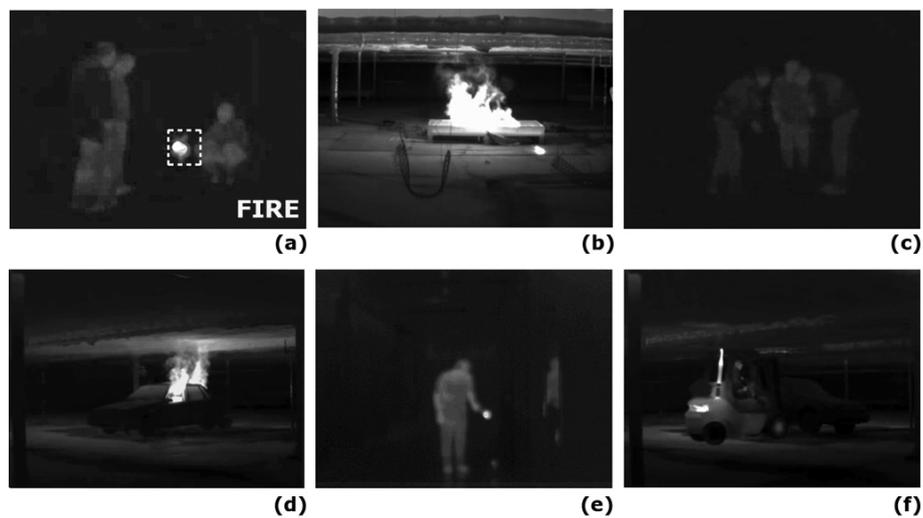


Figure 2.20: LWIR test sequences: (a) Attic - fire and people, (b) Pool fire, (c) Attic - moving people, (d) Car fire, (e) Corridor and (f) Forklift.

The Technical Note of NIST starts by reporting that thermal imaging cameras (TIC) are becoming an important tool for many firefighters and other first responders. However, due to the lack of performance standards for TIC, a wide variety of designs and capabilities are provided to end users with little consistency in reported performance. In order to understand the performance characteristics of TIC during fire fighting applications, it is critical that a set of performance metrics and standard testing protocols be developed to allow the fire service to evaluate TICs. This was the reason for NIST to conduct research to characterize and understand TIC performance. Their research is mainly based on first responder feedback, literature search, and full- and bench-scale testing results and serves as a basis for defining testing conditions that challenge TIC in meaningful ways.

In order to evaluate the TIC's performance, NIST proposes the following metrics:

- Contrast: measures how well the thermal imaging camera can represent temperature differences.
- Effective temperature range: measures the maximum temperature at which the TIC is able to produce an image.
- Spatial resolution: measures how well the thermal imaging camera can discern small details.
- Non-uniformity: measures the quantity of noise present in the image.
- Thermal sensitivity: measures the smallest possible temperature difference within the thermal image.

Based on these metrics, [101] proposes a testing methodology. Following this methodology, the authors have performed some preliminary experiments on which they also report in the technical note. One of these experiments is shown in Fig. 2.21. Three different types of TIC are simultaneously viewing an identical thermal scene: a long corridor with a heated mannequin on the floor, and reflective and heated targets mounted on the wall at the end. A fire room is located adjacent to the corridor on the right. Visual inspection, i.e., subjective evaluation, already shows that camera technology has definitely a big influence on what we see.

Within the proposed testing methodology, NIST also reports on multiple kind of test setups. In Fig. 2.22, for example, TICs are positioned in the hot upper layer viewing a target through heavy toluene smoke, and in the lower layer viewing flames and high heat conditions.

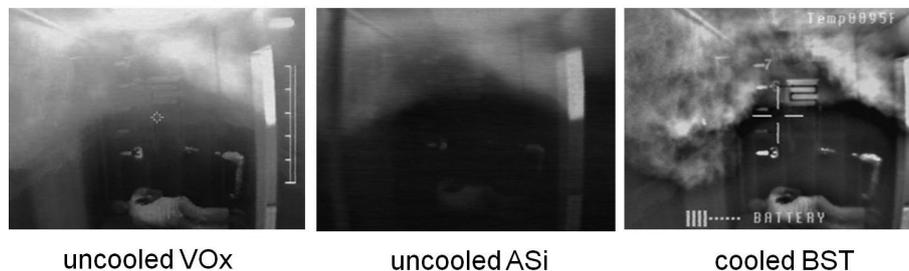


Figure 2.21: NIST experiment on TIC camera technology: camera technology has a big influence on what we see.

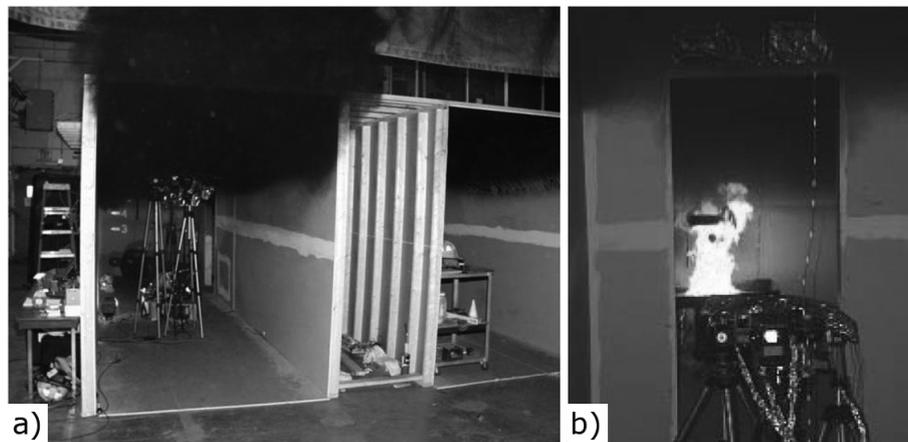


Figure 2.22: NIST test setups: in (a) TIC is positioned in the hot upper layer viewing a target through heavy toluene smoke, and in (b) TIC is positioned in the lower layer viewing flames and high heat conditions.

Based on these preliminary experiments, NIST has formulated the following conclusions (and recommendations):

- Presence of dust and water vapor do not significantly impair the imaging performance of TIC, even in high concentrations.
- Hot smoke and flames in the field of view, however, do have a negative impact on TIC imaging performance. It was found that the TIC detector technologies tested, an amorphous silicon (ASi), a vanadium oxide (VOx), and a barium-strontium-titanate (BST), did not perform consistently across the different test conditions. Although the detector technology alone is not a sufficient indicator of TIC imaging performance, it

appeared that low heat conditions were more amenable to microbolometers (VOx and Asi) than BST detectors. Conversely, conditions in which a wider range of gas and surface temperatures are present appear to be more suitable for BST detector technology.

- Some problems related to the practical application: besides the fire-related limitations, these devices also have a poor depth perception, i.e., it is difficult to judge how far away objects are. An additional limitation of infrared technology is that, since materials at the same temperature are shown as the same color, the display will not depict many details normally viewable in visible light. Finally, firefighters inside the structure, whether they are using the TIC to assist in fire attack or primary search, must remember to not become overconfident because this tool allows them to see in virtual zero visibility.

Based on the above discussed performance metrics, the test methodology and the experimental test setups proposed by NIST, we plan to pay more attention to this real fire performance aspect in our future investigations.

2.5 Influence of the background model

Inspection of the several visual and infrared flame and smoke detection algorithms that have been proposed in literature shows that most of them start from simple background subtraction in spatial domain, e.g., frame differencing and running average. The influence of the background model, as such, is not yet fully explored. This section describes a first attempt in this direction and investigates the added value of a DWT based background subtraction method for segmenting the input scene during video fire detection. The proposed method focuses on both the high-pass, i.e., energy-rich, and low-pass images of the DWT input video frames in spatial domain. Experimental results show that the DWT based method leads to better fire detection results than non-wavelet based background subtraction methods in both visual and infrared spectral range. 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 setup.

First, each of the ‘traditional’ BG subtraction methods from our evaluation are briefly discussed. Subsequently, the novel DWT based background subtraction method is discussed more in detail. Finally, all these methods are evaluated on a set of visual and infrared test videos and results of this evaluation are analyzed.

2.5.1 Running average based background subtraction

Currently, the majority of fire detectors is based on rather simple dynamic background subtraction methods, such as the running average based method which is used in [36, 74, 75]. 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:

$$BG_{n+1}[x, y] = \begin{cases} \alpha BG_n[x, y] + (1 - \alpha)F_n[x, y] & \text{if } F_n[x, y] \rightarrow BG \\ BG_n[x, y] & \text{if } F_n[x, y] \rightarrow FG, \end{cases} \quad (2.13)$$

where the update parameter α is a time constant that specifies how fast new information supplants old observations. Here $\alpha (=0.95)$ was chosen close to 1 as in the work of Toreyin et al. [74]. 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 $BG_{n+1} = BG_n$.

2.5.2 Advanced MGM: simple mixture of models (SMM)

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. [102]. 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 [102] 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 [102].

2.5.3 Discrete Wavelet Transform based FG extraction

Our novel Discrete Wavelet Transform (DWT) based FG extraction algorithm is schematized in Fig. 2.23. First, the input video frame is transformed using a DWT, which convolves the image with several banks of filters. 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 2.5.1. Only its input differs: here also the previous extracted foreground FG_{n-1}^c 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. 2.1. This kind of energy analysis is also used with success in [36,42] for flame feature analysis. However, to the best of our knowledge, it is the first time this is used for BG subtraction.

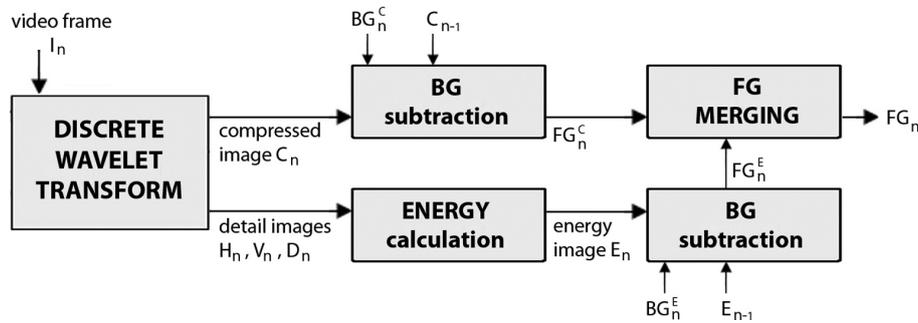


Figure 2.23: DWT based FG extraction.

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 foreground (FG), i.e., moving object. Some exemplary test results are shown in Fig. 2.24. As these examples show, flame reflections or fire based illumination changes do not cause any problem.

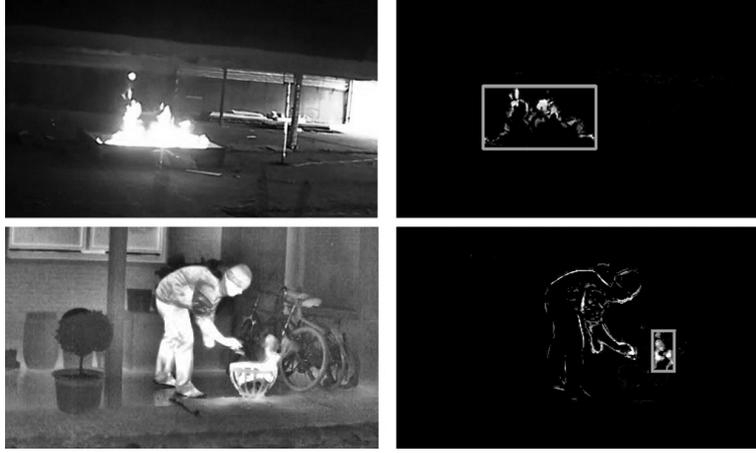


Figure 2.24: Visual and LWIR example of DWT based FG extraction. Orange bounding boxes show detected fire regions.

2.5.4 Evaluation

In order to objectively evaluate the detection results of the proposed wavelet-based BG subtraction method, and to compare it to state-of-the-art non-wavelet based moving object detectors, the detection results are evaluated against manually created ground truth (GT) data. Important to mention is that this evaluation is done on an object level basis, which is more strict than the more commonly used frame-based evaluation techniques [103]. 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 we calculate four detection metrics [104]:

$$precision = \frac{TP}{TP + FP}, \quad (2.14)$$

$$recall = \frac{TP}{TP + FN}, \quad (2.15)$$

$$specificity = \frac{TN}{TN + FP}, \quad (2.16)$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (2.17)$$

TP = True Positive; TN = True Negative
FP = False Positive; FN = False Negative

Table 2.6 summarizes the detection results for all the tested sequences. By comparing the four detection metrics (Eq. 2.14 - Eq. 2.17) for the investigated BG subtraction methods, the added value of wavelet versus non-wavelet based BG subtraction can easily be seen. The higher these metrics score, the better the flame detector, and more specific its BG subtraction, performs. As the results indicate, the DWT yields best detection results. It performs 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. By further inspecting these results one can also see that the overall gain of using wavelet based BG subtraction is bigger in the visual domain than in the thermal domain. This is logic, 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 of the number of objects which are falsely detected as flames. Also in this case, the wavelet-based method outperforms the non-wavelet based methods. In order to further evaluate the quality of the detectors we have also proposed a confidence metric in [105], which is related to the BB area overlap (percentage overlap).

2.6 Time-of-flight based fire detection

The main topic of this section is a novel time-of-flight based fire detection method for indoor fire detection. The indoor detector is based on the depth and amplitude image of a time-of-flight camera. Using the information from both image modalities, flames can be detected very accurately by fast changing depth and amplitude disorder detection. In order to detect the fast changing depth, depth differences between consecutive frames are accumulated over time. Regions which have multiple pixels with a high accumulated depth difference are labeled as candidate flame regions. Simultaneously, the amplitude disorder is also investigated. Regions with high accumulative amplitude differences and high values in all detail images of the amplitude image its discrete wavelet transform, are also labeled as candidate flame regions. Finally, if one of the depth and amplitude candidate flame regions overlap, fire alarm is given.

At the end of this section, first steps towards TOF based smoke detection are also discussed. The proposed histogram based smoke detector mainly focuses on global changes in the depth maps, which have no influence on the amplitude images. Preliminary experiments already show the effectiveness of the proposed approach.

Table 2.6: Performance evaluation of ‘traditional’ and wavelet-based BG subtraction methods for visual and LWIR flame detection.

sequence <i>method / range</i>	precision	recall	specificity	accuracy
hexane pool fire				
<i>SIMPLE / visual</i>	0	0	0	0
<i>/ IR</i>	0	0	0	0
<i>SMM / visual</i>	0.46	0.44	0	0.29
<i>/ IR</i>	0.97	0.97	0.86	0.95
<i>DWT / visual</i>	0.74	0.74	0.52	0.70
<i>/ IR</i>	0.97	0.92	0.81	0.91
outdoor pit fire				
<i>SIMPLE / visual</i>	0.59	0.66	-	0.49
<i>/ IR</i>	0.76	0.82	-	0.68
<i>SMM / visual</i>	0.55	0.89	-	0.52
<i>/ IR</i>	0.84	0.91	-	0.79
<i>DWT / visual</i>	0.81	0.98	-	0.80
<i>/ IR</i>	0.88	0.92	-	0.83
car park fire				
<i>SIMPLE / visual</i>	0.58	0.53	0.47	0.52
<i>/ IR</i>	0.92	0.44	0.97	0.63
<i>SMM / visual</i>	0.67	0.58	0.52	0.55
<i>/ IR</i>	0.99	0.42	1	0.65
<i>DWT / visual</i>	0.78	0.58	0.76	0.65
<i>/ IR</i>	0.98	0.52	0.98	0.70
human actions				
<i>SIMPLE / visual</i>	-	-	0.47	0.47
<i>/ IR</i>	-	-	0.59	0.59
<i>SMM / visual</i>	-	-	0.54	0.54
<i>/ IR</i>	-	-	0.66	0.66
<i>DWT / visual</i>	-	-	0.82	0.82
<i>/ IR</i>	-	-	0.96	0.96

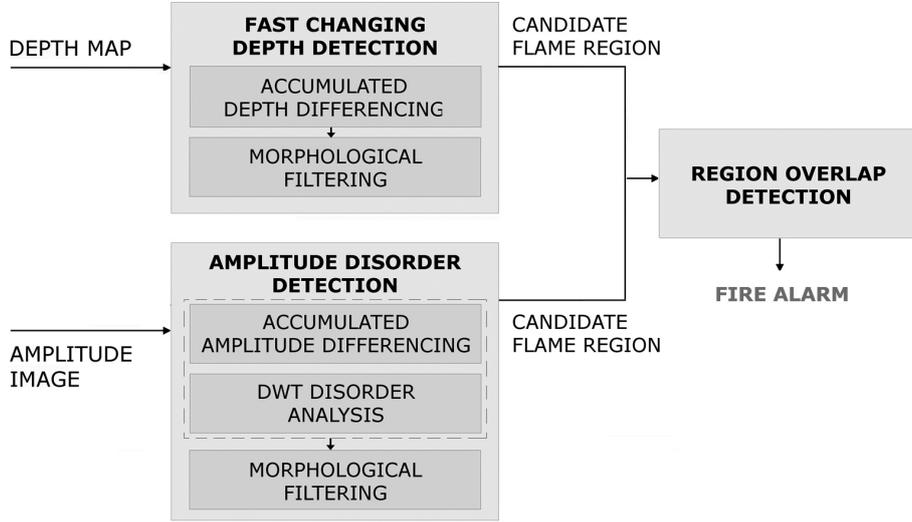


Figure 2.25: General scheme of the TOF based flame detector. By combining the detection results of the fast changing depth detection and the amplitude disorder detection, flames can be detected.

2.6.1 Indoor TOF based flame detector (distance < 10m)

A general scheme of the indoor TOF based flame detector is shown in Fig. 2.25. The proposed algorithm consists of three stages. The first two stages, i.e., the fast changing depth detection and the amplitude disorder detection, are processed simultaneously. The last stage, i.e., the region overlap detection, investigates the overlap between the resulting candidate flame regions of the prior stages. If there is an overlap, fire alarm is given. Because the proposed algorithm requires reliable depth maps, its detection distance is limited to the range of the TOF camera, which is between one and ten meter.

A. Fast changing depth detection

The fast changing depth detection starts with calculating the accumulated frame difference AFD_n (Eq. 2.18) between the current depth frame F_n^{depth} and the previous and the next depth frame, i.e., F_{n-1}^{depth} and F_{n+1}^{depth} respectively. By rounding the absolute frame differences, the AFD_n^{depth} is able to distinguish fast changing flames from more slowly moving ordinary objects. Pixels which AFD_n^{depth} is greater than zero, get a label 1 in the candidate flames image $Flames_n^{depth}$. Other pixels get a label zero:

$$AFD_n^{depth} = \lfloor |F_n^{depth} - F_{n+1}^{depth}| + |F_n^{depth} - F_{n-1}^{depth}| \rfloor, \quad (2.18)$$

$$Flames_n^{depth} = \begin{cases} 1 & \text{where } AFD_n^{depth} > 0 \\ 0 & \text{otherwise} . \end{cases} \quad (2.19)$$

Next, a morphological closing with a 3×3 structuring element connects neighboring candidate flame pixels, i.e., pixels with a label 1 in $Flames_n^{depth}$. Subsequently, a morphological opening filters out isolated candidate flame pixels using the same structuring element. The resulting connected flame pixel group(s) of $Flames_n^{depth}$ form(s) the depth candidate flame region(s). An example of the fast changing depth detection is shown in Fig. 2.26.

B. Amplitude disorder detection

The amplitude disorder detection starts with a similar accumulated frame differencing as the one which was used for the fast changing depth detection:

$$AFD_n^{amp} = \lfloor |F_n^{amp} - F_{n+1}^{amp}| + |F_n^{amp} - F_{n-1}^{amp}| \rfloor. \quad (2.20)$$

However, as high AFD_n^{amp} frame differences also occur at the boundary pixels of ordinary moving objects which are close to the TOF sensor, this feature alone is not enough for accurate flame detection.

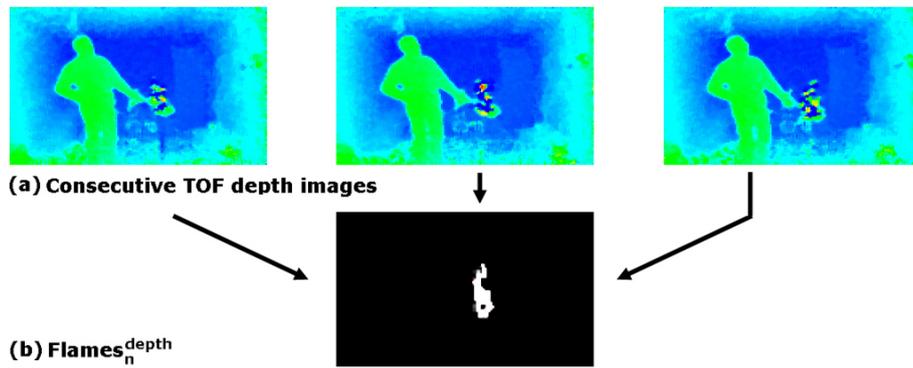


Figure 2.26: Fast changing depth detection: (a) consecutive TOF depth images and their (b) morphologically filtered accumulated depth difference ($Flames_n^{depth}$).

In order to distinguish flame pixels from the boundary pixels of ordinary ‘close’ moving objects, which also have high amplitude values when appearing in front of the sensor, the discrete wavelet transform (DWT) [106] of the amplitude image is also investigated. Experiments (Fig. 2.27) revealed that flame regions are uniquely characterized by high values in the horizontal H , vertical V and diagonal D detail images of the DWT. Ordinary ‘close’ moving objects do not have this characteristic. For this reason, an AFD_n^{amp} region Ω with high accumulated amplitude differences is only labeled as candidate flame region if at least one pixel with a maximal H value (=1), one pixel with a maximal V value (=1) and one pixel with a maximal D value (=1) can be found in the object region DWT_{Ω}^{detail} (which is detected by accumulated frame differencing of consecutive amplitude images). Only then it gets a value of 1 in $Flames_n^{amp}$:

$$DWT_{\Omega}^{detail} = \begin{cases} 1 & \text{if } \max(H_{\Omega}) \times \max(V_{\Omega}) \times \max(D_{\Omega}) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.21)$$

$$Flames_n^{amp} = \begin{cases} 1 & \text{where } AFD_n^{amp} > 0 \text{ AND } DWT_{\Omega}^{detail} = 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.22)$$

Analogously as in the fast changing depth detection, the morphological filtering connects neighboring candidate flame pixels in $Flames_n^{amp}$ and filters out isolated candidate flame pixels. The resulting flame pixel group(s) of $Flames_n^{amp}$ (Fig. 2.28) form(s) the amplitude candidate flame region(s).

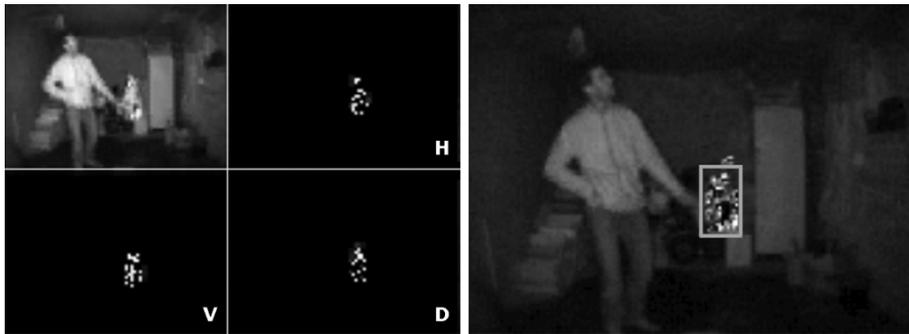


Figure 2.27: Discrete wavelet transform of amplitude image: flames show high values in horizontal (H), vertical (V) and diagonal (D) detail images.

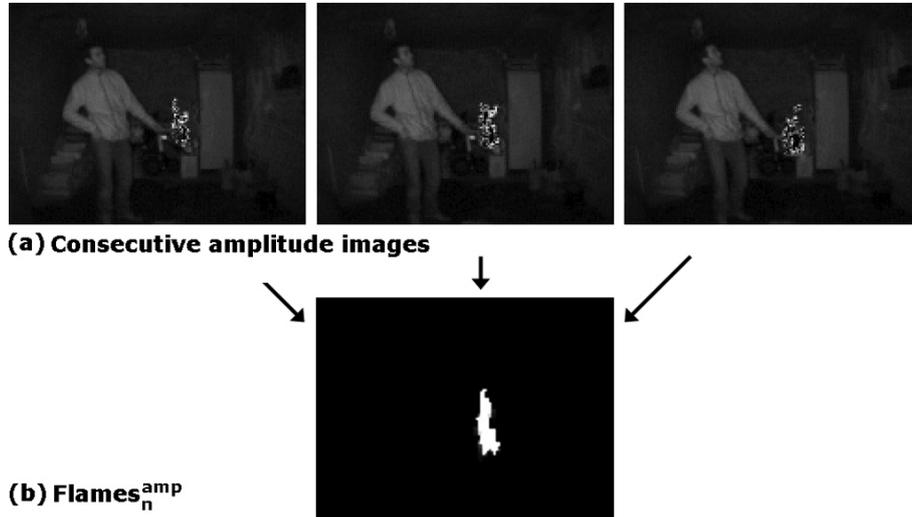


Figure 2.28: Amplitude disorder detection: (a) consecutive TOF amplitude images and their (b) morphologically and DWT filtered accumulated amplitude differences ($Flames_n^{amp}$).

C. Region overlap detection

This last stage investigates the overlap between the depth and the amplitude candidate flame region(s), i.e., $Flames_n^{depth}$ and $Flames_n^{amp}$ respectively. Important to mention is that, in order to do this, the depth map and the amplitude image need to be registered. However, as they are both obtained using the same sensor, both TOF outputs are already aligned on each other. In order to detect the overlap, it is sufficient to perform a logical AND operation between $Flames_n^{amp}$ and $Flames_n^{depth}$. If the resulting binary image contains one or more ‘common’ pixels, i.e., pixels with a value of 1, fire alarm is given. In Fig. 2.29, an example of this region overlap detection is shown.

D. Experimental results

The TOF camera used in this work is the Panasonic D-Imager [93]. 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 Optirima [107]. The technical specifications of the D-Imager are shown in Fig. 2.30.

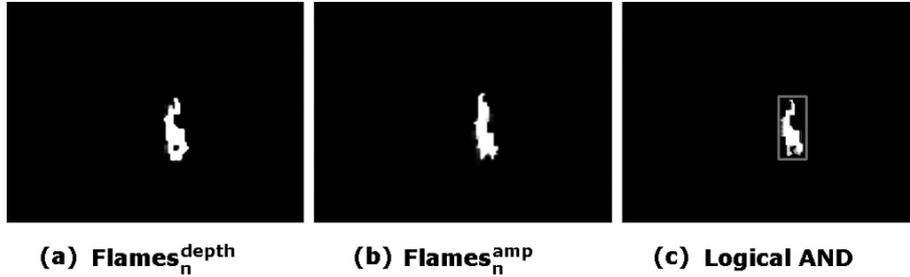


Figure 2.29: Region overlap detection: (c) logical AND of (a) depth and (b) amplitude candidate flame regions.

PANASONIC 3D IMAGE SENSOR "D-IMAGER" (EKL3104)

Pixel Array Size	Horizontal	160 Pixels
	Vertical	120 Pixels
Field of View	Horizontal	60°
	Vertical	40°
Operating Range		1.2m - 9.0m 3.94ft - 29.53ft
Frame Rate		upto 30 FPS (frames/sec)
Absolute Accuracy		± 4cm
Output type	Range data	11bit USB 2.0
	Image data	8bit
Operating temperature		-10°C - +50°C
External dimensions		170mm x 54mm x 49mm
Weight		± 520g

Figure 2.30: D-Imager and its technical specification.

To illustrate the potential use of the proposed indoor TOF based flame detector, several realistic fire and non-fire indoor experiments were performed. An example of these experiments, i.e., the paper fire test, is shown in Fig. 2.31. As can be seen in the depth maps, the measured depth of flames changes very fast. Even between two consecutive frames, very high depth differences are noticeable. In the amplitude images, on the other hand, it can also be seen that the boundaries of the flames have a very high amplitude.

Simultaneously to the TOF recording with the Panasonic D-Imager, we also recorded the experiments with an ordinary video camera. As such, the TOF detection results can be compared to the proposed visible flame detection algorithm (Section 2.3) and state-of-the-art VFD methods.

The results in Table 2.7 show how robust fire detection in indoor environments (distance $< 10\text{m}$) can be obtained with relatively simple TOF image processing. Compared to the VFD detection results of our visible flame detector, i.e., an average detection rate of 93% and an average false positive rate of 2%, the proposed TOF-based flame detector, with its 96% detection rate and no false positive detections, performs better for these primary experiments. The TOF based fire detector, however, is not able to detect the fire in outdoor situations or outside the range of the TOF camera. Main reason of its failing is the fact that its depth maps becomes unreliable under these circumstances. In order to cope with this problem, an outdoor visual-TOF flame detector is introduced in the next chapter (Section 3.6).

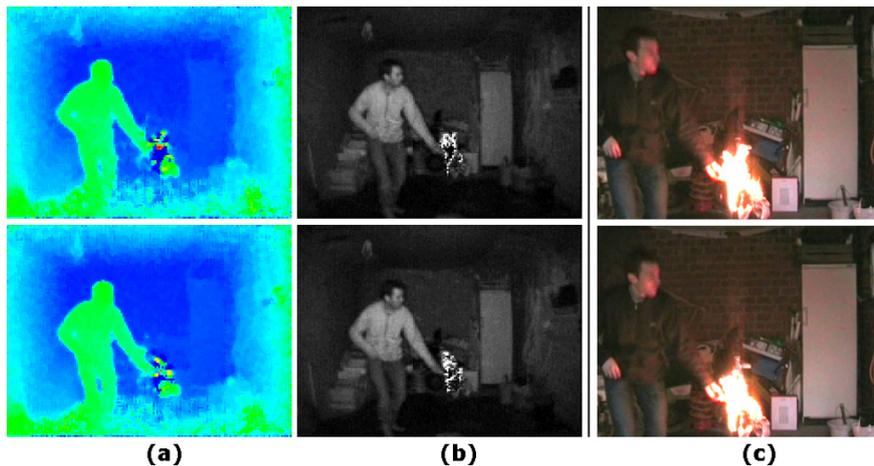


Figure 2.31: Paper fire test: (a) TOF depth map and (b) corresponding amplitude image of two consecutive frames; (c) ordinary video (not registered).

Table 2.7: Performance evaluation of indoor TOF-based fire detection.

Video sequence	# fire frames (GT)	# detected fire frames	# false positive detections	TOF flame detection rate	visible flame detection rate
Paper fire	517	496	0	0.96	0.93
Wood fire	1038	1021	0	0.98	0.94
Straw fire	645	611	0	0.95	0.92
People	0	0	0	-	-

$$* \text{detection rate} = (\# \text{ detected fire frames} - \# \text{ false alarms}) / \# \text{ fire frames}$$

The missing 5% detection rate can mainly be attributed to the fact that the resolution of the TOF camera, for the moment, is too low to detect small objects over long distances. Very small flames (e.g., in the beginning of the fire) are, as such, not detected. Furthermore, fire frames will not always have high detail in H,V,D DWT images. But this is not really a problem, since for detection/alarming purposes, the flame detection rate should not be 100%. For the analysis however, this is of course more important.

Since we did not have both LWIR and TOF sensors available at the same moment, we were not able yet to do a comparison between TOF and LWIR detection. So, since they have only been tested within different experiments, it is difficult to compare their behavior for the moment. Depending on the environment characteristics, however, it is expected that one type of detector will outperform the other and vice versa. Based on this fact, we state that only by using multi-modal VFD, a ‘better’ detector can be achieved providing high accuracy under all circumstances.

2.6.2 TOF based smoke detection

By further analyzing the TOF video sequences of the fire experiments, we noticed that smoke causes a kind of global changing in the depth images. The observed phenomenon (Fig. 2.32) can best be described as if the scene is floating in depth direction. Furthermore, we noted that these smoke related ‘depth changes’ have no significant impact on the amplitude images, i.e., the amplitude images remain nearly the same. Based on this TOF related smoke behavior, a novel TOF based smoke detector was started to be developed. Although this detector is not yet 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 of the TOF based smoke detector is shown in Fig. 2.33. The algorithm consists of three stages. The first two stages, i.e., the amplitude based detection of background blocks and the average depth change detection, are processed simultaneously. The last step, i.e., 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, fire alarm is raised.

The proposed algorithm performs the smoke detection on a block level instead of on pixel level. Each input frame F_t^{depth} and F_t^{amp} at time t is subdivided in ' $n \times n$ ' size blocks, in order to reduce measurement disturbances, i.e., to filter out errors and measurements inaccuracies. Depending the resolution of the camera and the scene characteristics, an appropriate blocksize must be chosen. In our experiments, blocks of 8-by-8 pixels have found to be the most successful. For each block $b_t^{depth}[i, j]$, an average depth value $\overline{F_t^{depth}[i, j]}$ is computed as the average of all the pixel values $F_t^{depth}[x, y]$ in that block:

$$\overline{F_t^{depth}[i, j]} = \frac{\sum_{x=i*n}^{(i+1)*n-1} \sum_{y=j*n}^{(j+1)*n-1} F_t^{depth}[x, y]}{n * n}. \quad (2.23)$$

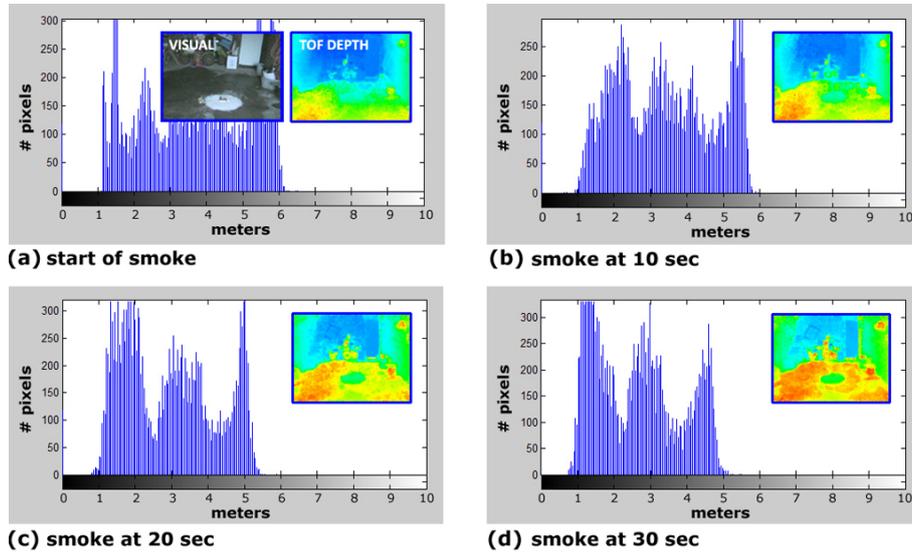


Figure 2.32: TOF smoke behavior: global change of the scene depth. Average depth change between (a) start of smoke and (d) smoke at 30s is almost 1 meter.

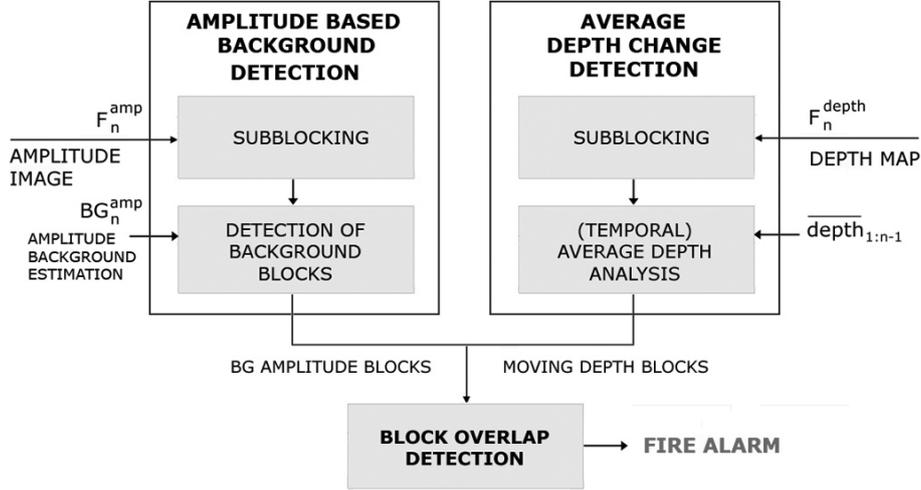


Figure 2.33: General scheme of TOF based smoke detector.

A. Amplitude based background detection

Small and large scale fire experiments revealed that smoke has no significant impact on the amplitude images of a TOF camera. When smoke appears in the field of view of the TOF camera, the amplitude images remains nearly the same. In the depth images, contrarily, smoke causes a global changing in the depth direction (as is discussed further on). 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 algorithm. However, instead of looking for blocks with certain level of motion, we now look for blocks which do not change significantly. A BG amplitude block is determined by comparing the amplitude values of the block $b_n^{amp}[i, j]$ in the current frame F_n^{amp} with the values of the corresponding block in the BG model BG_n^{amp} . If the sum $dif_n^{amp}[i, j]$ of the absolute differences of the block pixels (Eq. 2.24) is lower than the dynamic threshold t_{BG}^{amp} [50, 52], the block is labeled as BG block:

$$dif_n^{amp}[i, j] = \sum_{x=i*n}^{(i+1)*n-1} \sum_{y=j*n}^{(j+1)*n-1} |F_n^{amp}[x, y] - BG_n^{amp}[x, y]|, \quad (2.24)$$

$$b_n^{amp}[i, j] \rightarrow \begin{cases} BG, & \text{if } dif_n^{amp}[i, j] < t_{BG}^{amp} \\ FG, & \text{otherwise.} \end{cases} \quad (2.25)$$

B. Average depth change detection

The average depth change detection performs a temporal analysis of the average depth values (Eq. 2.23) of the current block $b_n^{depth}[i, j]$ and the previous blocks $b_{k:n-1}^{depth}[i, j]$. If the standard deviation of these average depth values exceeds the t_σ^{depth} threshold of 0.1, the block is labeled as moving depth block:

$$b_n^{depth}[i, j] \rightarrow \begin{cases} MOVING, & \text{if } \sigma(\overline{b_{k:n}^{depth}[i, j]}) > t_\sigma^{depth} \\ NON - MOVING, & \text{otherwise.} \end{cases} \quad (2.26)$$

C. Block overlap detection

The block overlap detection, i.e., the last step of our TOF based smoke detector, checks the overlap between BG amplitude blocks $b_n^{amp}[i, j]$ and moving depth blocks $b_n^{depth}[i, j]$. Overlapping blocks, i.e., blocks with an average depth change which does not cause changes in the amplitude values, are labeled as candidate smoke block:

$$b_n[i, j] \rightarrow \begin{cases} SMOKE, & \text{if } b_n^{depth}[i, j] = MOVING \\ & \text{and } b_n^{amp}[i, j] = BG \\ NON - SMOKE, & \text{otherwise.} \end{cases} \quad (2.27)$$

If several, i.e., at least two, candidate smoke blocks occur in three consecutive TOF images, fire/smoke alarm is raised. Depending the ‘monitoring characteristics’, however, the number of candidate smoke possibly needs to be adjusted, but this is currently out of the scope of our work.

D. Experimental results

In order to evaluate the proposed TOF based smoke detector, we performed a wood/paper fire and a Christmas tree fire experiment (Fig. 2.34). The first experiment, i.e., the wood/paper fire, was performed in a garage box of 3-by-7 meter. The Christmas tree fire experiment, on the other hand, was performed in a car park of 30-by-30 meter. Besides the fire, ordinary moving objects were also present in each of the sequences. As a preliminary evaluation, we test how fast the proposed algorithm detects the smoke, and compare it to the GT smoke start. Results (Table 2.8) show that smoke is detected within less than 10 seconds (if it appears in the field of view of the camera). Although this may seem long, this is (often) much faster than traditional sensors.

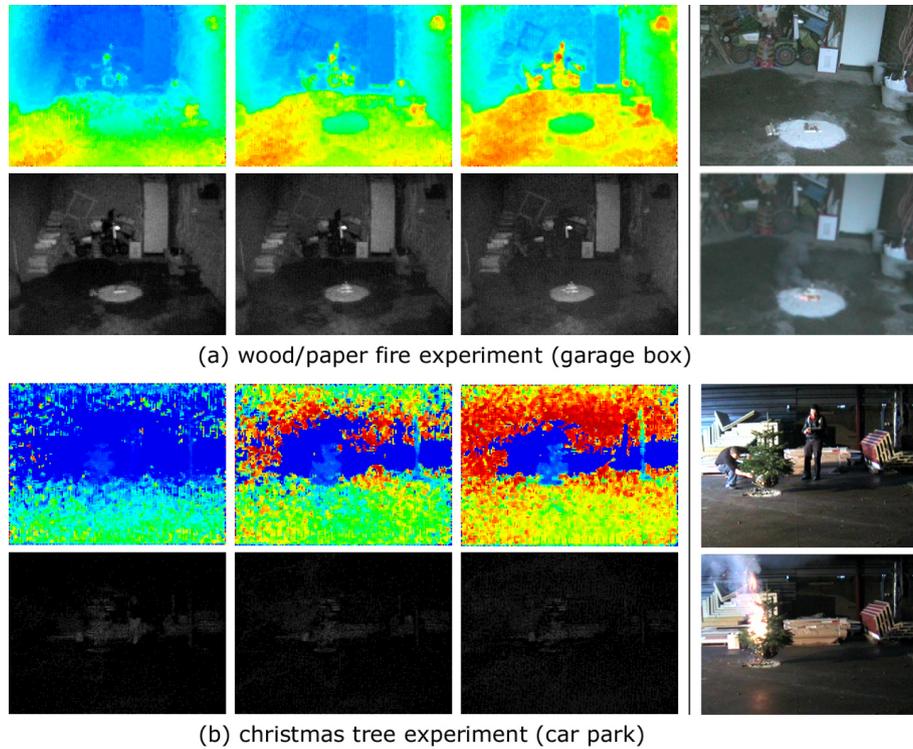


Figure 2.34: Exemplary depth and amplitude frames of TOF smoke experiments: (a) wood/paper fire and (b) Christmas tree fire. Smoke causes a kind of global change of the depth images, which does not influence the amplitude images. Visual images (right column) are given as a reference.

The reason why it takes longer to detect the wood/paper fire compared to the Christmas tree fire can most probably be found in the fact that Christmas trees generate much more smoke in the beginning of the fire and will, as such, faster disturb the sensor. The timing results in Table 2.8 also show that the TOF based smoke detector achieves quasi similar detection times as our visible smoke detector (Section 2.3.2). Furthermore, more recent experiments revealed that depending the environment characteristics, the TOF smoke detector can outperform the visible detector and vice versa. Based on this fact, we state again that it is our strong belief that only by using multi-modal VFD (Chapter 3), a ‘better’ detector can be achieved providing high accuracy under all circumstances. Combining the best of both worlds will always be a win-win. However, in order to keep the complexity low, the multi-modal detectors should be low-cost, i.e., the requirement to which all the proposed single-sensor detectors adhere.

Table 2.8: Performance evaluation of TOF-based smoke detection.

Video sequence	smoke start (GT)	TOF detected smoke start	TOF delay	visible detected delay
Wood/paper fire	17s	26s	9s	13s
Christmas tree fire	36s	41s	5s	4s

2.7 General remarks and future improvements

In order to further improve the proposed detection algorithms, future work will mainly focus on two topics: the investigation/evaluation of feature alternatives and more advanced strategies to combine the feature values. Related to this, some general remarks about our feature-based approach are also discussed in the following two subsections.

2.7.1 Feature combinations and alternatives

As was already indicated in Section 2.3, we use the flame color rate (FCR) to identify pixels that are within the red-yellow range. For the moment, we do not incorporate the color-changing aspect of flames, mainly to keep the algorithm low-cost. However, as the color of flames does not remain as steady as the flame-colored background objects, one could wonder if it might not be more interesting to use this color-changing aspect instead of the flame color rate. However, experiments on our set of fire and non-fire video sequences revealed that FCR is more discriminative than the color-changing aspect. Of course, using this extra feature will improve the detection, however, the added value will be limited. Furthermore, since textured moving objects will have similar color-changing behavior, color filtering will (always) be necessary to distinguish them from flames.

Since the BBD and POD flame features seem the most discriminative, one could also argue to only focus on these two features. However, based on our experimental experience we expect that this will increase the number of missed detections and false alarms. Fig. 2.35a shows some examples of more ‘controlled’ fires which will most possibly not be detected when focusing only on BBD and POD.



Figure 2.35: BBD/POD experiments: examples of (possible) (a) missed detections and (b) false alarms.

Furthermore, objects like flags, moving crowds/groups or a person who is dancing (shown in Fig. 2.35b) can have the same BBD and POD behavior as flames. In these cases, FCR can help in ‘deciding if the object is a fire or non-fire object. Finally, for some types of fires, like pool fires, FCR will be the most discriminative and also in the early stages of a fire, BBD and POD will not always be high. A similar remark holds for fires at large distance. Within this context, we have performed some additional tests (shown in Fig. 2.36), in which our algorithms are used to detect the fire at a distance of more than 1km. If we only focus on BBD and POD, the combined fire risk value is sometimes too low to detect the fire.

Current research [108] at Bilkent University (under supervision of Prof. Enis Cetin), focuses on two new feature alternatives: the axes of bounding ellipse disorder (ABED) and the center of mass disorder (CMD). ABED, on the one hand, can be seen as an alternative for the Bounding Box Disorder (BBD) and Principal Orientation Disorder (POD) which focuses on changes in the axes lengths of the bounding ellipse around the object (Fig. 2.37). It combines the two ‘flame characteristics’ of BBD and POD. As such, it is expected that ABED will be high when POD and/or BBD are high. We do not think that using ABED in combination with POD and BBD will improve the detection performance a lot, since it does not really focus on new flame characteristics.

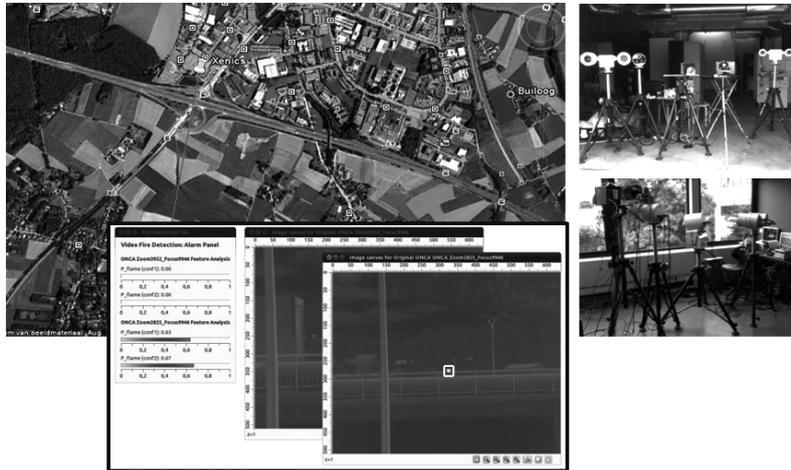


Figure 2.36: BBD/POD long range distance tests: fire at a distance of more than 1km.

However, since it combines both features into one ‘feature’, it can be used as an alternative to POD/BBD, which (possibly) reduces the computational cost. The added value of ABED will further be evaluated within the context of the EU-FP7 FIRESENSE project [109], in which both Bilkent University and UGent-IBBT participate. CMD, on the other hand, will in our opinion be less discriminative than the other features. Moving objects (e.g. dancing people), can have similar CMD behavior, which (possibly) increases the risk on false alarms. However, this will depend on how the features are combined, as is discussed next.

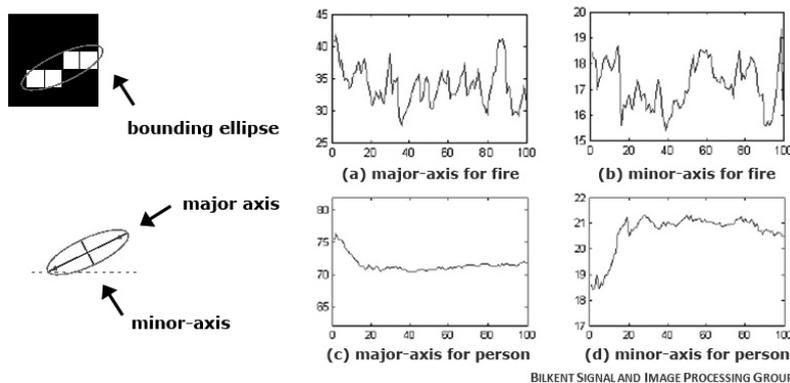


Figure 2.37: Axes of bounding ellipse for (a-b) fire and (c-d) moving person.

2.7.2 Advanced strategies to combine feature weights

Depending the type of the fire, one feature can be more discriminative than the other and vice versa. For example, when analyzing pool fires (i.e., a kind of controlled fire), BBD and POD will not always be high, while the FCR will (more likely) be high. As such, we admit that further research is needed regarding the feature weights, i.e., how the features are combined. Instead of taking the average of the feature values, one could also think of ‘learning/training’, for example using support vector machines (SVM), the feature combinations of flames and non-flames objects. However, in order to have a good fire/non-fire classifier, there is need for an exemplary/training fire dataset. Currently, such a dataset is not yet available. So, although first SVM results [108] of Bilkent University show good results, further research on this topic is needed.

Another way of intelligently combining the features is by using context-dependent feature weights, which is (slightly) suggested in Section 3.4. For example, if there is a lot of wind/air circulation it could be logical to increase the impact of POD. This is also related to sensor-feedback, in which detection criteria (such as thresholds and weights) are linked on other sensors knowledge about the environment. On both topics, i.e., automatic detection of context-dependent feature weights and sensor feedback, we will start working after this PhD.

2.8 Conclusions

Early detection of smoke and flames is a prerequisite to limit both human and material losses in case of fire. However, due to several limitations of traditional sensors, which are still in use today, it is not always possible to timely detect the fire. Especially in large open spaces, crucial time is often lost. In order to deal with this problem, video based fire detection is gaining increasing importance in the last years and can be seen as a viable alternative or complement to the existing fire detection techniques. VFD can be applied in conditions in which conventional methods fail, e.g., to detect the fire from a distance in large open spaces. As soon as smoke or flames occur in one of the camera views, the fire can be detected.

Although VFD has proven useful to solve several problems related to the traditional sensors, real-world fire experiments revealed that the video based

detection of fire also suffers from some vision related problems. For example, missed detections and false alarms often occur when the light conditions are bad. Hence, developing a reliable VFD system in the visible spectrum is shown a tough challenge. Our main contribution to this problem is the introduction of video sensors operating in other spectral bands. Contrary to many other research approaches, the proposed optimizations for the detection of flames and smoke are more in the breadth than in the depth direction. Instead of dealing with ever more complex visual fire detection algorithms, the focus of our research is on investigating the benefit of infrared and time-of-flight image processing for fire detection. The latter one, i.e., TOF fire detection, has not been covered by related work until now.

Firstly, state-of-the-art video fire detection algorithms are investigated. Based on this SOTA and our experiments, a low-cost visual flame and smoke detection algorithm is proposed. In order to ensure real-time detection, the computational cost of both algorithms is kept as low as possible. They both consist of only two building blocks: a moving object detection and a set of ‘low-cost’ fire features, which uniquely describe smoke and flames. The discriminative features for flame detection are the Bounding Box Disorder (BBD), the Principal Orientation Disorder (POD) and the Flame Color Rate (FCR). The smoke detection, on the other hand, uses the Boundary-Area Disorder (BAD), Energy Disorder (ED) and chrominance disorder (CD) features. By analyzing the values of these features, a fire alarm is raised. The concept of using ‘low-cost’ features for fire detection was originally presented by Bosch et al. [75] in their work on object discrimination by infrared image processing. Contrary to most of our features, they use moment-based features (similar to those proposed by Bilkent University [108]). Bosch et al. their features are: intensity, signature and orientation. In their work, no real analysis is done on these features. As such, our extrema-analysis can be seen as a contribution to their work. The extraction of the ‘signature’ also has a higher computational cost, compared to our features. Furthermore, before doing the feature analysis, we perform a dynamic BG subtraction, which is more suitable to cope with the time-varying characteristics of dynamic scenes.

Instead of using a simple (dynamic) background model, we also investigated the benefit of using a DWT based background subtraction method for segmenting the input scene during video fire detection. Experiments revealed that the DWT based method leads 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 setup.

Subsequently, our thermal LWIR flame detector is proposed. When light conditions are bad or when smoke occurs in the field of view of the camera, thermal LWIR vision is a fundamental aid for flame detection. The detector mainly reuses most of the building blocks of our ‘low-cost’ visual flame detector. Additionally, the moving object detection is extended with a hot object segmentation step to extract the hottest objects out of the set of LWIR foreground (FG) objects. Only these hot FG objects are further analyzed using the set of LWIR flame features, i.e., BBD, POD and Histogram Roughness (HR). Again, a fire alarm is raised on the basis of the values of these features. Important to mention is that the cost of a thermal IR camera is still too high to widely use them for video surveillance purposes. Depending the application it can as such, at the moment, be better to choose one of the other detectors.

Finally, we investigated the possibilities of time-of-flight (TOF) based flame and smoke detection. The TOF based fire detection methods presented in this chapter are the first attempts in this direction. Preliminary experiments already show that the combination of amplitude and depth information is a win-win. However, problems arise for flame detection in outdoor situations and outside the range of the TOF camera. Under these circumstances the TOF depth map becomes unreliable and cannot be used anymore for accurate flame detection. In order to cope with this problem, a multi-modal detector using visible and TOF amplitude information is suggested in the next chapter.

Experiments revealed that each of the proposed detectors is able to accurately detect smoke or flames. Depending on the environment characteristics, however, one type of detector outperforms the other and vice versa. Based on this fact, we state that only by using multi-modal VFD, which is discussed in the next chapter, a ‘better’ detector can be achieved providing high accuracy under all circumstances. Each sensor type has its own specific limitations, which can only be compensated by other types of sensors. Nevertheless that, due to cost reasons, it was one of our objectives at the start of this research to develop a fire detection system which could operate on the existing CCTV equipment, the cost of using multiple video sensors does not outweigh the benefit of multi-modal fire analysis. The fact that manufacturers also ensure a decrease in the sensor cost, fully opens the door to multi-modal video analysis.

To finalize this chapter, we shortly discuss one of the main directions for future work, i.e., pan-tilt-zoom (PTZ) based VFD. As should be clear from the examples in this chapter, the proposed algorithms are developed to work on static cameras, i.e., cameras which do not change position during acquisition of the images. Hence, to monitor a large area, many cameras can be needed. For several video surveillance applications, e.g., wildfire detection and monitoring

of large car parks, this is unaffordable. Instead of using multiple cameras, these applications mostly use a dynamic PTZ camera. One of the drawbacks of PTZ, however, is that the position of everything in the scene can change between consecutive image acquisitions, which complicates for example the background modeling and object tracking/analysis. This makes it difficult to predict how the proposed algorithms will behave in a PTZ setting and how costly (required) adaptations for PTZ fire detection will be. A recent literature survey did not reveal many research in this direction [110].

Besides the research on PTZ based VFD, a more thorough investigation of the fire detection capabilities of the different IR spectral bands (NIR/SWIR, MWIR, LWIR) seems also interesting as a future work.

The author's work on video fire detection using visible, infrared and time-of-flight imaging led to the following publications.

- Steven Verstockt, Sofie Van Hoecke, Pieterjan De Potter, Peter Lambert, Charles-Frederik J. Hollemeersch, Bart Merci, Bart Sette, and Rik Van de Walle. Multi-modal time-of-flight based fire detection. *Special issue in Multimedia Tools and Applications: Analysis and Retrieval of Events/Actions and Workflows in Video Streams*, submitted March 2011.
- Steven Verstockt, Sofie Van Hoecke, Nele Tilley, Bart Merci, Bart Sette, Peter Lambert, Charles-Frederik J. Hollemeersch, and Rik Van De Walle. Hot Topics in Video Fire Surveillance. In *Video Surveillance*, pages 443–458. InTech, 2011.
- Steven Verstockt, Ioannis Kypraios, Pieterjan De Potter, Chris Poppe, and Rik Van de Walle. Wavelet-based multi-modal fire detection. In *Proc. of 19th European Signal Processing Conference (EUSIPCO)*, accepted for publication, September 2011.
- Steven Verstockt, Nele Tilley, Bart Merci, Charles-Frederik J. Hollemeersch, Bart Sette, Sofie Van Hoecke, Peter Lambert, and Rik Van de Walle. Future directions for video fire detection. In *Proc. of 10th international IAFSS symposium*, accepted for publication, June 2011.
- Steven Verstockt, Alexander Vanoosthuysse, Bart Merci, Nele Tilley, Bart Sette, Charles-Frederik J. Hollemeersch, Peter Lambert, and Rik Van de Walle. Performance evaluation framework for vision-based fire detection. In *Proc. of 12th International conference on Fire Science and Engineering (Interflam 2010)*, pages 257–268, July 2010.

- Steven Verstockt, Rudy Dekeerschieter, Alexander Vanoosthuysse, Bart Merci, Bart Sette, Peter Lambert, and Rik Van de Walle. Video fire detection using non-visible light. In *Proc. of 6th International seminar on Fire and Explosion Hazards (FEH-6)*, pages 49–50, April 2010.
- Steven Verstockt, Peter Lambert, Rik Van de Walle, Bart Merci, and Bart Sette. State of the art in vision-based fire and smoke detection. In *Proc. of 14th International Conference on Automatic Fire Detection (AUBE)*, pages 285–292, September 2009.
- Steven Verstockt, Bart Merci, Peter Lambert, and Rik Van de Walle. Video processing techniques for early fire detection. In *Proc. of International WildFire Management conference*, June 2009.
- Steven Verstockt, Peter Lambert, Bart Merci, Rik Van de Walle, and Bart Sette. Video fire detection going faster than fire. In *Annual report and newsletter of European Group of Organisations for Fire Testing, Inspection and Certification*, 8:12–13, 2008.

Chapter 3

Multi-modal fire detection

Single-sensor video fire detectors are plagued by a number of difficulties in real-world scenes. Many of these difficulties are mainly caused by limitations due to the type of sensors used. In most cases, each of these sensor specific limitations can be compensated by other types of sensors. As such, instead of dealing with ever more complex single-sensor fire detection algorithms, the focus of this chapter is on investigating and combining multi-modal information from the different types of video sensors that are discussed in the previous chapter. The main contribution of this chapter treats the registration of multi-modal images and proposes a novel silhouette based registration method, which is able to (semi-)automatically align visual, time-of-flight (TOF) and/or infrared (IR) images. The geometric parameters found using this registration method are further used by each of the multi-modal fire detectors that are presented at the end of this chapter. Based on several fire and non-fire experiments, these multi-modal smoke and flame detectors are identified as the best solution to achieve high accuracy under all circumstances. Combining the appropriate 'fire' features of visual, TOF and/or IR imagery, i.e., using the strengths of each medium, is shown to be a win-win situation.

3.1 Introduction

Developing an accurate video fire detector relying on only one type of video sensor is a huge challenge. A visual fire detector, for example, can fail due to noise, shadows, illumination changes, and other visual artifacts. Reflections and IR-blocking, on the other hand, can mislead an infrared detector. Although several 'single-sensor' solutions have already been proposed to cope with these problems, most of these depth-related research techniques cannot be guaranteed to work under all conditions.

In order to better compensate for the specific limitations, i.e., artifacts, of each type of sensors, a study on the combination of multiple sensors is a much more interesting track to follow. Due to the fact that, depending on the environment circumstances, one detector outperforms the other and vice versa, it is our strong belief that only by using multi-modal video fire detection (VFD) a ‘better’ fire detector can be achieved providing high accuracy under all circumstances.

The combined detection in the IR and visual spectral range is not new. The fusion of visual and IR images has already been proposed as a way to improve the detection performance in many application domains. Also in the domain of fire detection some steps are already taken in this direction. Contrarily, on the combination of a visual or IR camera with a time-of-flight (TOF) camera, only recently studies have begun. Concerning multi-modal fire detection, the visual-TOF flame detector presented in this chapter, is the first of its kind. For more details on the previous work in multi-modal fire detection, surveillance and image registration, section 3.2 lists the state-of-the-art in these domains.

An important problem when (f)using/combining detection results of different types of sensors is how to align the corresponding objects in the scene. This problem is also known as the registration problem. The goal of registration is to establish geometric correspondence between the multi-sensor images so that they may be transformed, compared, and analyzed in a common reference frame. Because corresponding objects in visual, thermal and/or depth images may have different sizes, shapes, features, positions and intensities, the fundamental question to address during registration is: what is a good image representation to work with? Section 3.3 treats this multi-modal registration question and proposes a novel silhouette based registration method, which is able to (semi-)automatically align visual, TOF and infrared images. As we assume parallel sensors whose lines of sight are close to each other, the proposed multi-modal registration strategy consists of a rigid transformation, which can be decomposed into a 2-D rotation, scaling and translation.

The proposed multi-modal registration method has successfully been used by each of the multi-modal fire detectors that has been investigated in our work. The first of these detectors is a long-wave infrared (LWIR)-visual flame detector (Section 3.4), which improves the visual flame detector proposed in Section 2.3. Similarly to all of the existing multi-modal fire detectors it focuses on the combined analysis of thermal and visual flame features. The second detector is an LWIR-visual smoke detector. To the best of our knowledge, no literature exists on the multi-modal detection of smoke using visual and infrared imagery. The reason for this can probably be found in the fact that

smoke becomes more and more transparent further in infrared spectrum, and as such, is hard to detect in this spectral range. However, the proposed LWIR-visual smoke detector (Section 3.5) makes use of this transparency feature. By temporal analysis of the silhouette coverage of moving objects in registered long-wave infrared (LWIR) and visual images, this multi-modal smoke detector is able to accurately ‘see’ the smoke. Our third and last multi-modal detector (Section 3.6) is a visual-TOF flame detector which can be used in outdoor situations, outside the range of the TOF camera and in case that smoke appears in the field of view of the TOF camera. Due to the fact that the TOF depth maps become unreliable under these circumstances, the proposed ‘outdoor’ detector only uses the TOF amplitude images in combination with our visual flame detector. Experimentally we found that by combining the different types of video data, the number of missed detections and false alarms can be reduced drastically, which results in a significant improvement of VFD.

Due to cost reasons it was one of our objectives at the start of our research to develop a fire detection system which could operate on the existing closed-circuit television (CCTV) equipment. However, the cost of using multiple video sensors does not outweigh the benefit of multi-modal fire analysis. The fact that manufacturers also ensure a decrease in the sensor cost in the next years, fully opens the door to multi-modal video analysis. Furthermore, increasing the number of sensors does not much affect the software/processing costs, since the detections itself can run in parallel. To summarize, one can say that low cost algorithms running on multiple sensors will start to take over the ever more complex single-sensor algorithms that are proposed in most publications today. It is our strong belief that the fusion of multi-modal data will become the keyword in video surveillance. In the conclusions (Section 3.7), at the end of this chapter, we further elaborate on this statement.

3.2 State-of-the art in multi-modal fire detection

Research on multi-modal fire detection has only started in the last decade and is still limited. The most interesting works are the studies of Arrue et al. [25] and Martinez-de Dios et al. [111] which use visual information to improve infrared detection of wildfires. The former work is an IR-visual false alarm reduction system which discriminates false alarms by analyzing the ratio between the alarm areas in visual and infrared images. In order to reduce the high false alarm rate due to infrared emissions from other sources, the system takes advantage of the information redundancy from visual and infrared cameras.

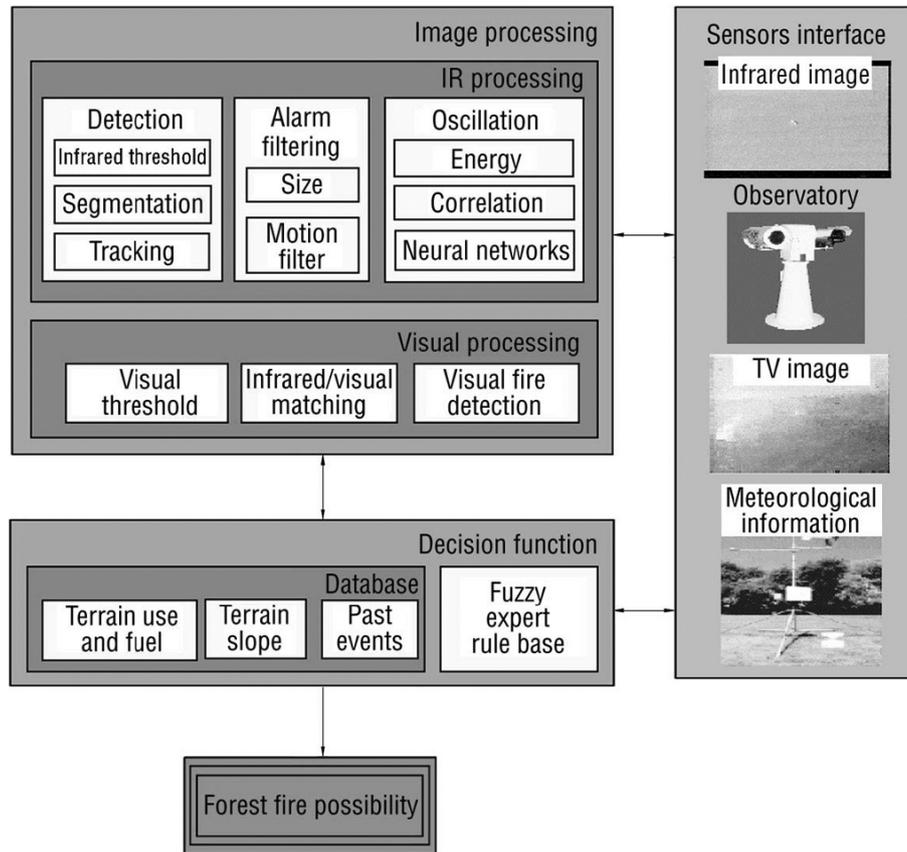


Figure 3.1: False Alarm Reduction system by Arrue et al. [25]

This is also the idea behind the multi-modal flame detectors proposed in this chapter. However, compared to the system of Arrue et al., which is shown in Fig. 3.1, our novel methods are computationally less complex and do not use additional information from meteorological sensors and from a geographical information database.

Similar to the method of Arrue et al., the work by Martinez-de Dios et al. presents how potential fire alarms from both thermal and visual images can be fused to obtain more reliable fire detection characteristics. In order to have direct vision of the (wild)fire this system makes use of unmanned aerial vehicles (UAVs), which also facilitate the computation of the geographical position of the detected alarms. For the specific use case of our research, however, the use of UAVs is too expensive and is practically not possible. For the registration of the multi-modal images Martinez-de Dios et al. make use of a homography-

based estimation of the transformation parameters (using a calibration board). The novel video fire analysis framework proposed in Chapter 4 uses a similar registration technique for the analysis of multi-view detection results. However, for the novel multi-modal detectors presented in this Section, which fuse the detection results of nearly co-located (adjacent and parallel) cameras, a low-complex registration method based on silhouette mapping is computationally more efficient.

Although the trend towards multi-modal VFD is not yet widely ‘visible’, the combination of detection in different spectral ranges has already proven successful in many other types of vision applications, as is further discussed.

3.2.1 Multi-modal video surveillance

The main benefit of (f)using multi-modal image data is that unreliably extracted parts from one sensor might be reliably extracted from the other sensor. This provides an opportunity for improving the detection performance. The combination of several types of imagery yields information about the scene that is rich in color, motion, depth and/or thermal detail. Once registered, such information can be used to successfully detect and analyze activity in the scene with fewer misdetections [24]. As a logical consequence of these benefits, multi-modal imaging is considered a win-win by many authors [19, 25] and has started to be actively used to improve the performance of object detection and recognition [26].

The majority of the research in multi-modal video analysis focuses on the fusion of infrared and visual images, especially in the field of surveillance [112], automatic target recognition [113], tracking [24] and medical image analysis [114]. The most interesting related work is the multi-sensor image fusion for the detection of weapons by Chen et al. [112, 115]. An object which can be seen in infrared, but not in visual, is detected as a candidate weapon. Similarly, we detect candidate smoke regions as moving objects which can be seen in visual, but not in infrared. Without the fusion of multi-modal detection information, i.e., when only relying on one sensor, it would be very difficult to detect these type of objects or to distinguish them from other types of objects which have similar thermal or visual features. By focusing on both objects’ ‘visible-invisible’ feature, which can only be detected with multi-modal video analysis, Chen’s and our method achieve better detection results, be it in different application domains. Compared to Chen’s registration method, which is based on the maximization of the mutual information criterion [114], our silhouette-based multi-modal registration method only performs low-complex geometric operations in order to estimate the transformation parameters.

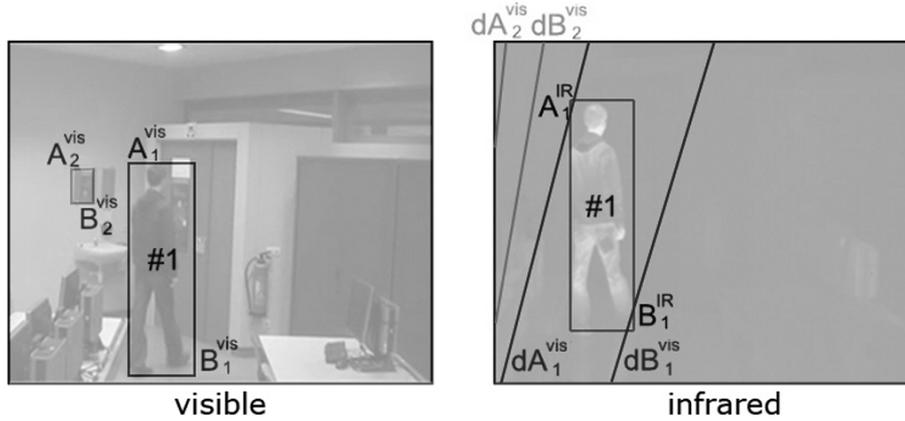


Figure 3.2: Example of decision fusion based on epipolar geometry (Benzeth et al. [116]).

Another interesting work is the optimized people detection and tracking proposed by Benzeth et al. [116]. In this work, it is shown how LWIR and daylight cameras can collaborate inside a stereo vision setup to reduce the false positive rate inherent to their individual use. Based on the epipolar geometry [117] of the stereo-vision system, Benzeth’s multi-modal detector checks if the IR detected object corners (A_i^{IR} , B_i^{IR}) are close enough to the epipolar lines (dA_i^{vis} , dB_i^{vis}) of the projected corners (A_i^{vis} , B_i^{vis}) of a visual object. If so, the object is considered a ‘human’; otherwise, it is discarded. An example of this decision fusion based on epipolar geometry is shown in Fig. 3.2. Although this multi-modal approach optimizes the detection of moving objects, its lack in an exact mapping, for example, makes it infeasible for our multi-modal smoke detector (Section 3.5) which mainly focuses on the percentage overlap of multi-modal moving objects.

The fusion of visual and TOF images, on the other hand, only recently started to be used as a way to improve everyday video analysis tasks. The main reason for this can be found in the fact that the TOF camera is a much ‘younger’ technology which only recently became commercially available. The results of first approaches (Section 2.2.3) already seem very promising and ensure the feasibility of TOF imaging in other domains, such as fire detection. For example, Bleiweiss and Werman [83] fuse time-of-flight depth and RGB color images to solve common problems in tracking and segmentation of RGB images, such as occlusions, fast motion, and objects of similar color. Since the camera they have used provides them directly a registered TOF and visual image, they do not report on the registration problem and mapping but on the fusion of multi-modal data itself. As such, their focus is different than ours.

In [118], Gould et al. augment a 2D object detector with 3D information from a depth sensor to produce a multi-modal object detector for robots. Their experimental results show that the multi-modal detector provides an average gain of 15% over a 2D detector for the detection of common household/office objects, i.e., an example of the multi-modal win-win. Sabeti et al. [84] describe two separate particle filter trackers, one using color and the other using TOF data, and compare them on a variety of video sequences. They conclude that each performs better in different environments and that combining the two would be beneficial, i.e., a similar statement as the one made in our introduction.

3.2.2 Multi-modal image registration

In order to combine the information in a multi-modal setup, e.g., for medical imaging and computer vision [119, 120], the corresponding objects in the scene need to be registered. The goal of registration is to establish geometric correspondence between the multi-modal images so that they may be transformed, compared, and analyzed in a common reference frame [121]. Since corresponding multi-modal objects may have different sizes, shapes, features, positions and intensities, as is shown in Fig. 3.3, a good image representation needs to be found that brings out the common information between the two multi-sensor images, while suppressing the non common information [122].

When choosing an appropriate registration method, a first distinction can be made between automatic and manual registration. In applications with manual registration, e.g., using a calibration checkerboard [24], a set of corresponding points are manually selected from the two images to compute the parameters of the transformation. The registration performance is evaluated by subjectively comparing the registered images based on these parameters. This is repeated several times until the registration performance is satisfied. If the background changes, e.g., due to camera movement, the entire procedure needs to be repeated. Because this manual process is labor intensive, automatic registration is more desirable and therefore preferred in our system.

Within the range of (automatic) registration methods, a distinction can be made between region, line and point feature-based methods [123]. It is necessary to use features that are stable with respect to the sensors, i.e., the same physical artifact produces features in both images. Compared to the correspondence of individual points and lines, region-based methods, such as silhouette mapping, provide more reliable correspondence between color, IR and TOF image pairs [115]. For example, comparing the visual, IR and TOF images in Fig. 3.3, one can see that some information varies a lot, but what is most similar are the silhouettes. Therefore, the proposed image registration method

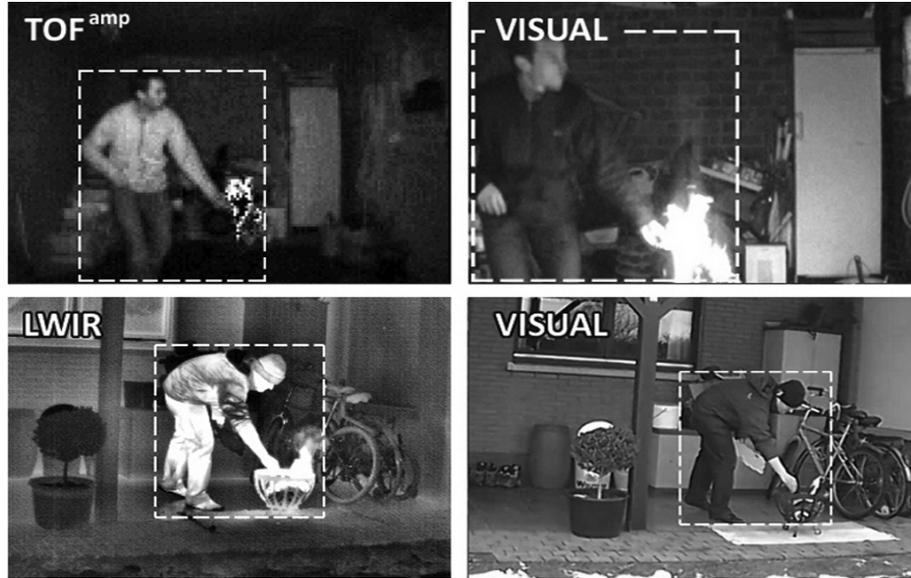


Figure 3.3: Comparison of corresponding objects in LWIR, visual and TOF images.

performs a match of the transformed color silhouette of the calibration object, i.e., a moving person, to its TOF or IR silhouette. The mutual information, i.e., the silhouette coverage, is assumed to reach its maximal value when both images are registered. However, knowing that the same silhouettes extracted from visual, TOF and/or IR images can have different details (as can be seen in Fig. 3.3), a complete exact match is (quasi) impossible. It is also important to mention that instead of using a person as the calibration object, also other objects in the scene can be used, as long as they are moving.

Although the proposed (semi-)automatic registration method performs good results, full automatic methods, such as the one proposed in [124], are the most desirable. This method automatically registers IR with visual image data using geometric structures that are matched with a partial graph matching algorithm. Another benefit of this method is that re-calibration can be done without the need of moving objects, as the method can calibrate on background geometry. However, due to its higher computational cost, we decided to use our own low-cost silhouette based registration method which, contrarily to the cited SOTA works, fulfills all the requirements stated in Section 1.6. Basically we can also align the multi-modal data using the calibration knowledge of the sensors, and always move the sensors together. However, in a real situation, it is more useful to align the images without imposing anything on the starting condition of the cameras and their relative position.

Based on this state-of-the-art basis on multi-modal video surveillance and image registration, the following section presents our novel silhouette based multi-modal image registration. The proposed registration method has successfully been used by each of the multi-modal fire detectors that are discussed at the end of this chapter.

3.3 Silhouette based multi-modal image registration

The proposed silhouette contour based image registration algorithm, shown in Fig. 3.4, coarsely registers the images taken simultaneously from the visual, LWIR and/or TOF parallel sensors whose lines of sight are close to each other. The registration starts with a silhouette extraction [115] in the visual, thermal and/or TOF amplitude image to separate the calibration objects, i.e., the moving foreground, from the background, which is assumed to be static. The novelty of the proposed silhouette extraction method is its combination of several existing concepts, which improves the performance of each of the ‘stand-alone’ concepts. So, none of its building blocks are really new, but it is their combination that makes the difference.

Key components of the moving object silhouette extraction are the dynamic background subtraction, automatic thresholding and morphological filtering with growing structuring elements, which grow iteratively until a resulting silhouette is suitable for multi-modal silhouette matching. After silhouette extraction, the multi-modal silhouettes are used to estimate the transformation parameters for the image registration. First, 1D contour vectors are generated from the resulting visual, thermal and/or TOF silhouettes using silhouette boundary extraction. This is followed by a Cartesian to polar transformation and a radial vector analysis. Next, in order to retrieve the rotation angle and the scale factor between the visual, LWIR and/or TOF image, these contours are mapped onto each other using circular cross correlation [125] and contour scaling. Finally, the silhouette mapping calculates the translation between the multi-modal images using maximization of binary correlation. The retrieved transformation parameters are used to align the multi-modal image(s).

In what follows, a more detailed description is given of each step in the silhouette-based registration. First, the extraction of visual (Section 3.3.1), thermal (Section 3.3.2) and/or TOF silhouettes (Section 3.3.3) of the calibration object is discussed. Next, the image registration is presented for the anal-

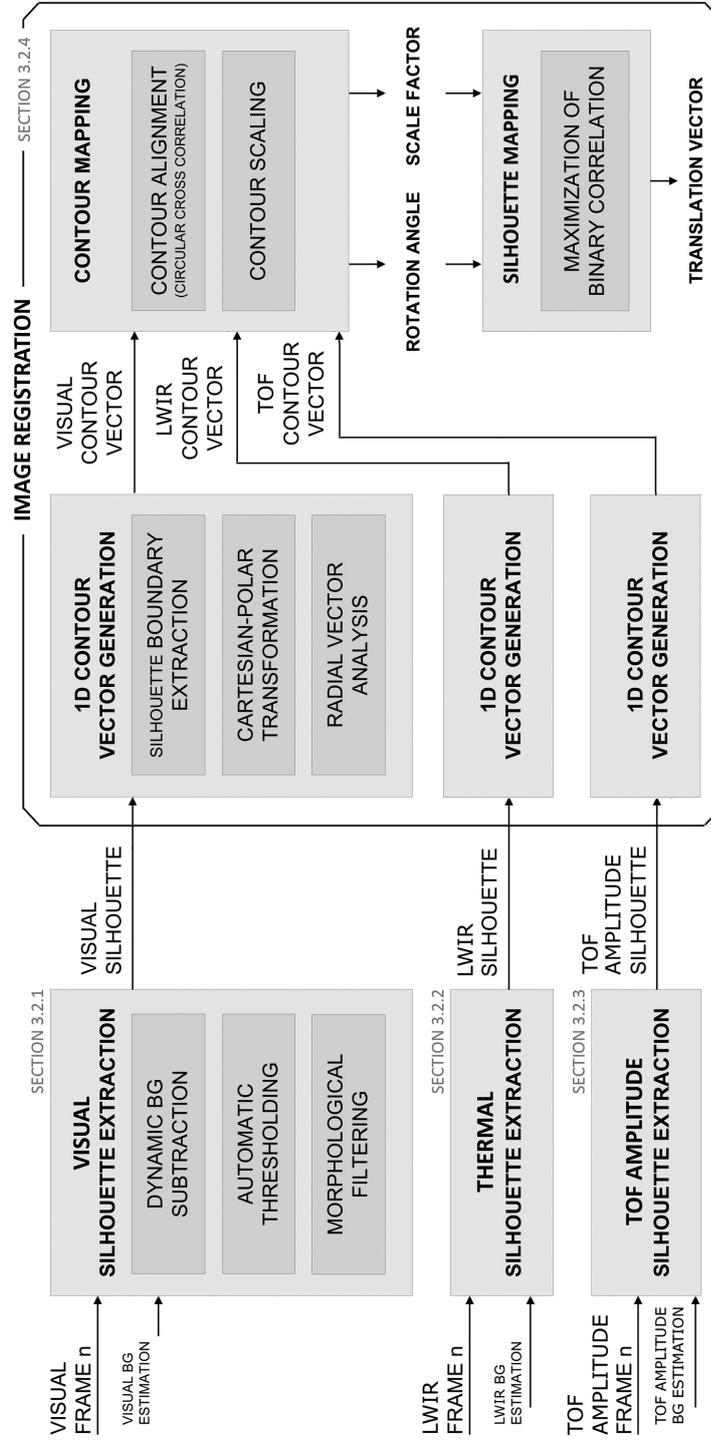


Figure 3.4: Silhouette-based image registration of visual and TOF amplitude images.

ysis of the visual and thermal silhouettes (Section 3.3.4). In order to detect the rotation, scale and transformation between the visual-LWIR multi-modal images, we perform a 1D contour vector generation, a contour mapping and a silhouette mapping. In the same way, image registration between other types of multi-modal images, e.g., visual-TOF and TOF-LWIR, can be performed. Due to the high similarity in the registration process, these other types of multi-modal image registrations, however, will not be discussed.

At the end of the section, examples of the registration process in different real-case scenarios are given to illustrate the accuracy of the proposed technique and to show its superiority over SOTA alternatives (Section 3.3.5).

3.3.1 Visual silhouette extraction

In order to extract the visual silhouette of the calibration person from the background, we propose the algorithm shown in Fig. 3.5, in which intensity, color and edge information of the moving part of the visual images are merged. Merging these three types of information is the only way we can guarantee the entire moving object silhouette is found under all circumstances (in our experiments). The algorithm uses the visual frame F_n , i.e., the input RGB video frame at time n , in which the calibration person is in the scene (Fig. 3.6b), and the visual background estimation BG_n , in which we assume that no moving objects occur (Fig. 3.6a). The algorithm starts with two **image transformations** to convert F_n into the intensity image I_n and the color image C_n . The color image C_n equals the ratio of input image F_n by the intensity image I_n :

$$C_n = F_n / I_n . \quad (3.1)$$

In short, the pixel values of each of the RGB color bands of F_n are divided by the intensity values of the corresponding pixels in the intensity/grayscale image I_n . This gives us the color values of the color image C_n (Eq. 3.1). More detailed information on the creation of C_n can be found in [126].

Next, a **dynamic background subtraction** [60] extracts the moving foreground (FG) out of I_n and C_n using the intensity and color image of the visual background estimation. By computing the absolute difference between I_n and C_n with everything in the scene that remains constant over time, i.e., BG_n , only the moving part of those images remains. The intensity BG estimation is updated dynamically after each segmentation using:

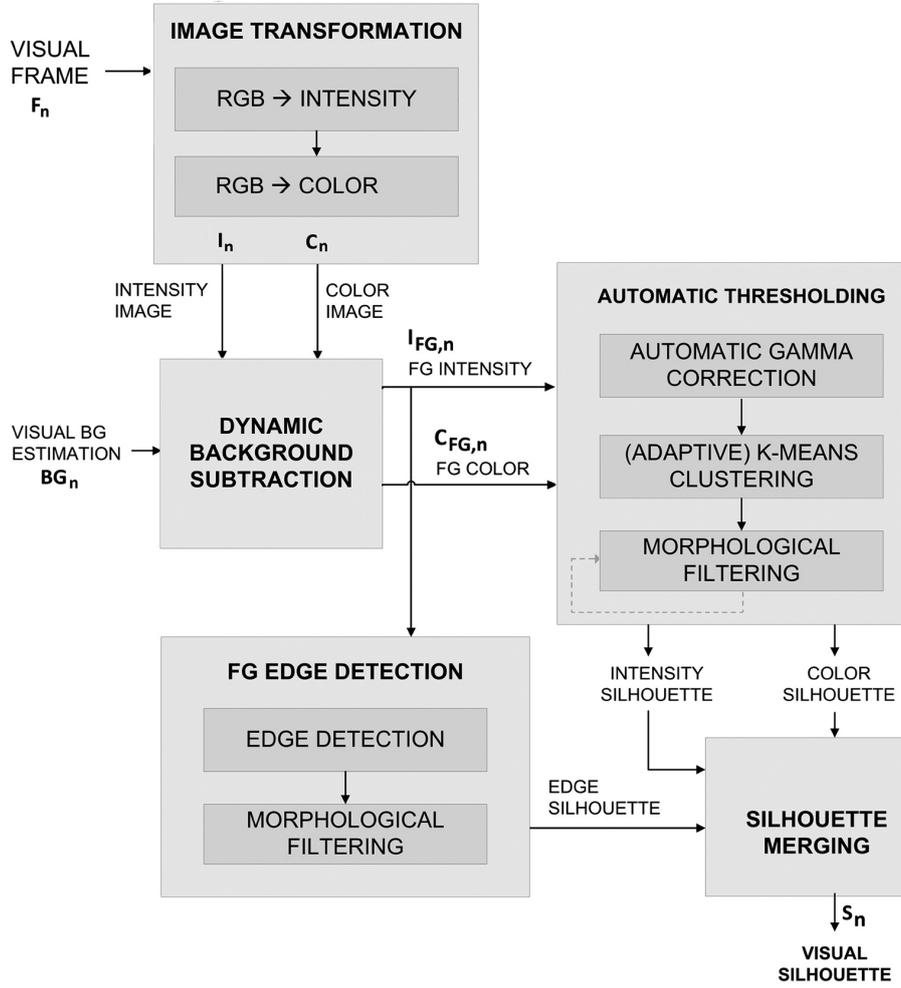


Figure 3.5: Visual silhouette extraction.

$$BG_{n+1}[x, y] = \begin{cases} \alpha BG_n[x, y] + (1 - \alpha)I_n[x, y] & \text{if } S_n[x, y] \rightarrow BG \\ BG_n[x, y] & \text{if } S_n[x, y] \rightarrow FG, \end{cases} \quad (3.2)$$

in which $[x, y]$ are the pixel coordinates, S_n is the final silhouette image and α is the update parameter. The closer α is to 1, the faster new information replaces old observations. Here $\alpha (= 0.95)$ was chosen close to 1, as suggested by Toreyin et al. [35]. The color BG estimation is updated analogously.

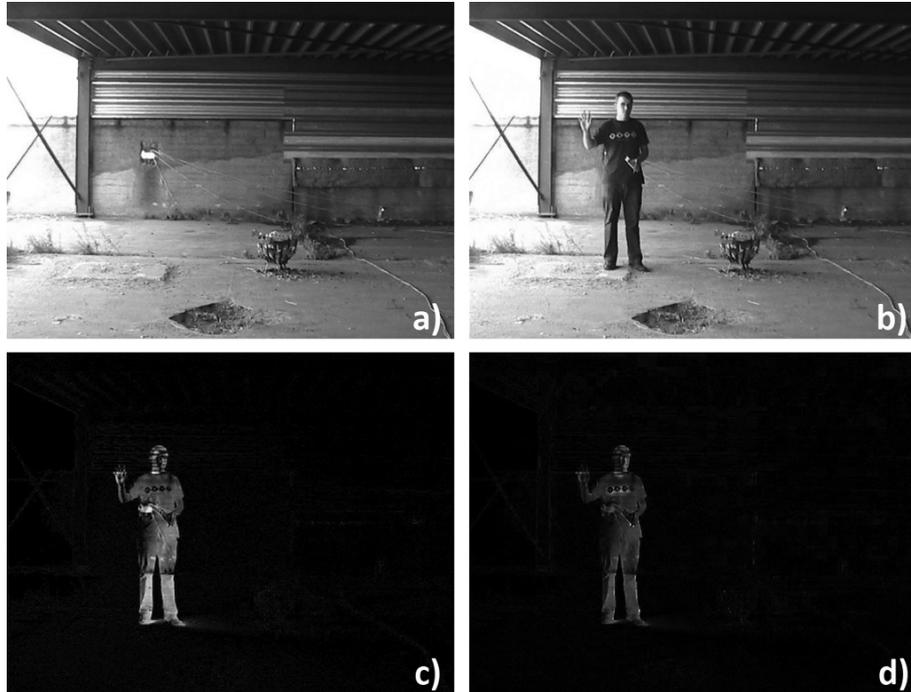


Figure 3.6: a) Background (BG) estimation and b) calibration input frame; c) BG subtracted intensity and d) BG subtracted color image.

After background subtraction, the resulting intensity and color foreground $I_{FG,n}$ and $C_{FG,n}$ are **thresholded automatically** using automatic *gamma correction*, (*adaptive*) *k-means clustering* and *morphological filtering* with growing structuring elements, which grow iteratively until the resulting silhouette is suitable for thermal-visual silhouette matching. It was found in our experiments, for example in Section 3.5.3, that the combination of these three steps gave the best results, compared to other frequently used segmentation techniques such as histogram equalization and contrast stretching [127, 128].

Gamma correction changes the brightness distribution of images. Using an appropriate gamma, this correction results in a more useful input image for k-means clustering, making details in both light and dark portions of the image more visible. To automatically generate an appropriate value an automatic gamma correction was used (details are presented in [129]). Based on the mean and standard deviation of the input image, for example $I_{FG,n}$, the gamma value γ for the automatic correction is calculated using:

$$\begin{aligned}
& \text{IF } \overline{I_{FG,n}} > 0.5 \\
& \quad \gamma = 1 + \frac{|0.5 - \overline{I_{FG,n}}|}{\sigma} \\
& \text{ELSE} \\
& \quad \gamma = \frac{1}{1 + \frac{|0.5 - \overline{I_{FG,n}}|}{\sigma}}.
\end{aligned} \tag{3.3}$$

In the same way, the gamma value for $C_{FG,n}$ can be calculated.

As the images in Fig. 3.7 and our results in Section 3.5.3 show, the gamma correction improves the segmentation results a lot when light conditions are bad or the color difference between the human calibration object and the background is minimal. Similar results can be achieved with homomorphic filtering [130]. However, the proposed technique is computationally less complex, which is essential to achieve the system requirements (Section 1.6).

For the extraction of the human silhouette in the gamma-corrected color and intensity foreground images, different thresholding techniques can be used. Among all of these techniques, automatic thresholding, like the Otsu method [97] and k-means clustering [131], are widely used because of their simple implementation and low computational cost. These methods automatically select an optimal gray-level threshold value for separating objects of interest from the background, based on their gray level distribution. However, since these standard methods only focus on Euclidean intensity distance, they are sometimes insufficient in forming the desired clusters in real-world image segmentation. The Otsu method, for example, fails if the histogram is unimodal or close to unimodal [132], as is the case in our experiments. Instead, a weighted distance measure, such as the spatial constrained k-means [133], the two-dimensional Otsu [134], the histogram valley emphasis [132], or the k-means adaptive clustering [135], performs much better by utilizing both local pixel/histogram information and pixel intensity. In our work the latter *k-means adaptive clustering*, with two clusters, is used. It was the only method which gave satisfactory results for each of the experimental setups, in with varying environmental conditions were tested. As the color silhouette extraction in Fig. 3.7 and the results in Section 3.5.3 show, this clustering achieves favorable results, even in low-light images. Similar results are retrieved for the intensity silhouette extraction.

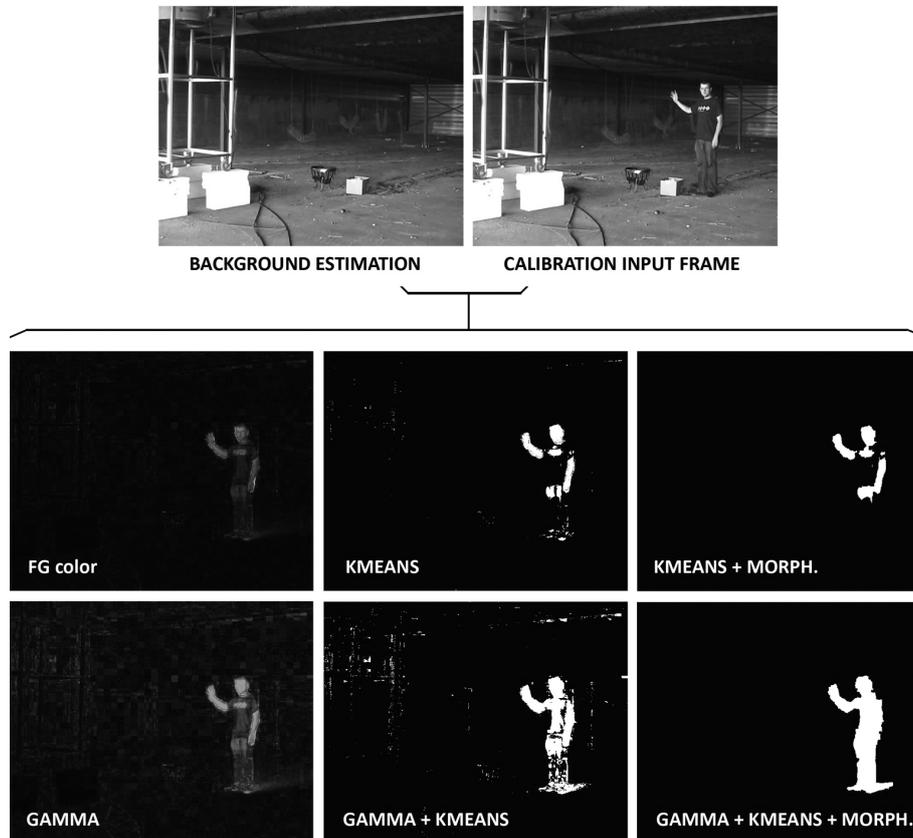


Figure 3.7: Results of color silhouette extraction with and without gamma correction.

In order to discard noisy objects and to improve the quality of the color and intensity silhouette, *morphological filtering* [70] is performed on the binary images after k-means clustering. First, small noisy FG objects are removed using a blob filter. Next, a morphological closing connects neighboring silhouette parts. Finally, a filling operator fills the remaining holes in the silhouette. The results of this morphological filtering are shown in Fig. 3.7. Combining gamma correction, k-means and morphological filters clearly results in appropriate silhouette extraction. In order to determine an optimal size for the structuring elements of the morphological filters, the structuring elements grow iteratively until the resulting silhouette is suitable for thermal-visual silhouette matching, i.e., until one FG silhouette object remains with adequate thermal-visual correspondence.

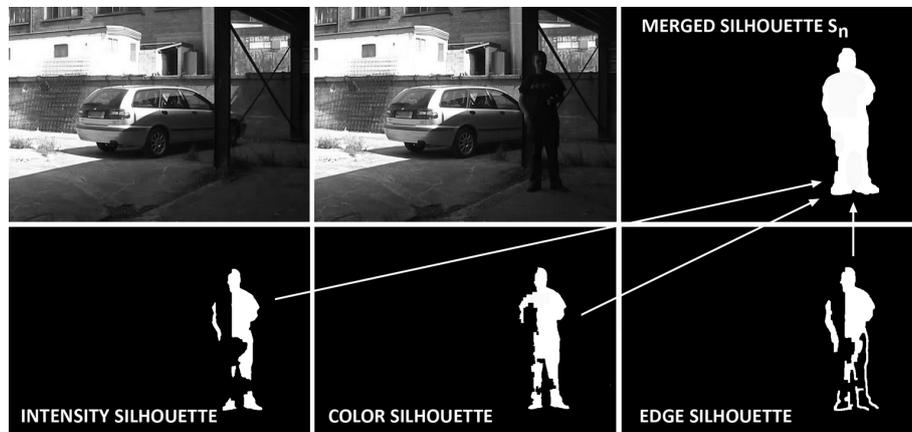


Figure 3.8: Silhouette merging.

As it is not always possible to extract the full silhouette out of the color or intensity images, as is shown in Fig. 3.8, we finally merge the color and intensity silhouette. In addition, the resulting silhouette is also merged with an edge silhouette, which is created using a standard Canny *edge detection* [70] and morphological filtering on the FG intensity images. The main reason why the Canny edge detector is used on the intensity images, and not on the color images, is that the intensity images are much richer in edge information than the color images. This is logical, as in the construction of the color images a lot of edge information, i.e., intensity variations, is discarded by dividing the input frame by the intensity image.

By *merging* the three different types of information into the final silhouette image S_n , accurate visual silhouette extraction is achieved. The merging itself starts by adding the binary values of the intensity, color and edge image together. Next, the resulting image is thresholded. Non-zero regions which contain one or more values that are bigger than 1, i.e., pixels that are foreground in more than one of the silhouettes, are mapped to foreground. All other regions are mapped to background. In this way, objects which have only been detected in one of the three silhouettes are discarded. So, the proposed algorithm is also able to cope with specific visual artifacts, such as (disconnected) shadows. As such, the combination of color, intensity and edges is a winning combination.

3.3.2 Thermal silhouette extraction

As already stated in Section 2.4.1, the body of a person can seldom be imaged as a whole warm object in thermal images. Due to the insulating properties of some clothes, for example, it is difficult to segment the whole body from the background. As such, problems can arise during silhouette extraction. The here proposed silhouette extraction copes with those problems by focusing on the absolute intensity differences between the current frame and the thermal BG estimation, instead of focusing on the pure intensity values.

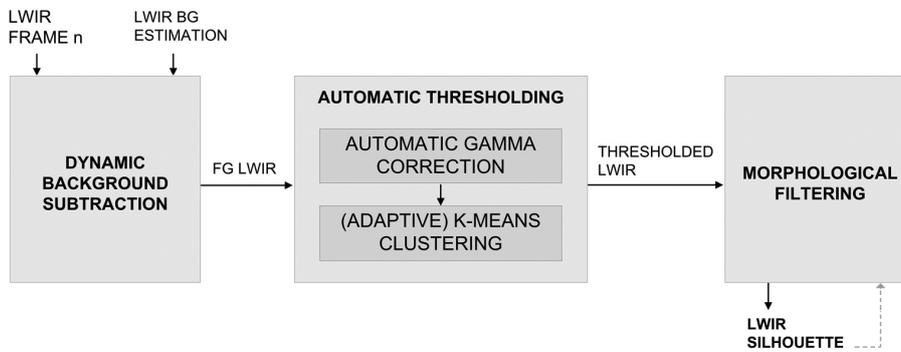


Figure 3.9: LWIR silhouette extraction.

The main steps of the thermal LWIR silhouette extraction algorithm are shown in Fig. 3.9. The algorithm uses the thermal frame F_n^{LWIR} , in which the calibration person is in the scene (Fig. 3.10a), and the thermal BG estimation BG_n^{LWIR} , in which no moving objects occur (Fig. 3.10b). The algorithm starts with the same **dynamic background subtraction** as the one used for visual extraction. The BG subtraction extracts the thermal foreground FG_n^{LWIR} (Fig. 3.10c) out of F_n^{LWIR} by calculating the absolute difference of F_n^{LWIR} and the thermal background estimation BG_n^{LWIR} , which is also updated dynamically using the final thermal silhouette S_n^{LWIR} .

Subsequently, **automatic thresholding** extracts the candidate thermal silhouette out of FG_n^{LWIR} using the same *automatic gamma correction* and *k-means clustering* as for the visual silhouette extraction. Finally, the thermal extraction also uses *morphological filtering* with iterative growing structuring elements to discard remaining noisy objects and to improve the silhouette quality of the thermal silhouette S_n^{LWIR} . As shown in Fig. 3.10d, the combination of these steps produces satisfactory results not only for visual, but also when applied to thermal images.

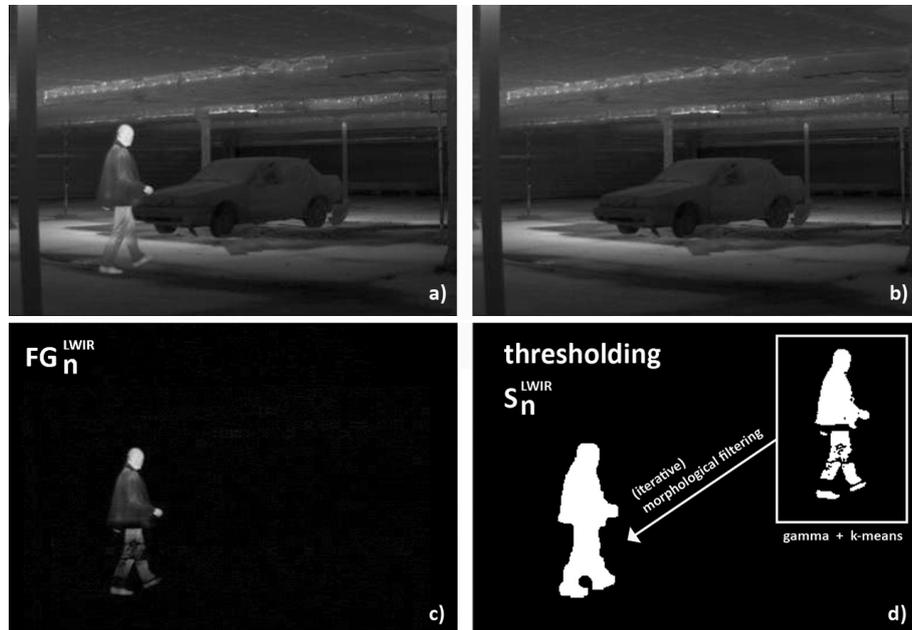


Figure 3.10: Example of LWIR silhouette extraction: calibration person (a); thermal BG estimation (b); BG subtraction (c); thresholding and morphological filtering (d).

3.3.3 TOF silhouette extraction

As shown in Fig. 3.11, the TOF silhouette extraction follows the same three steps as the thermal silhouette extraction. First, the dynamic background subtraction extracts the TOF amplitude foreground FG_n^{TOF} from the TOF amplitude image F_n^{TOF} (Fig. 3.12a) using the TOF BG estimation BG_n^{LWIR} (Fig. 3.12b). Next, the TOF amplitude FG image FG_n^{TOF} (Fig. 3.10c) is thresholded automatically and morphologically filtered in the same way as the thermal and visual FG images.

The silhouette extraction result in Fig. 3.12d illustrates that the combination of dynamic background subtraction, automatic thresholding and morphological filtering produces satisfactory results not only for visual and thermal, but also when applied to TOF amplitude images. The silhouette of the calibration person is successfully segmented from the background. As such, the proposed technique can be seen as a ‘generic’ silhouette extraction technique for video-based sensors which produces satisfactory results on visual, thermal and TOF images.

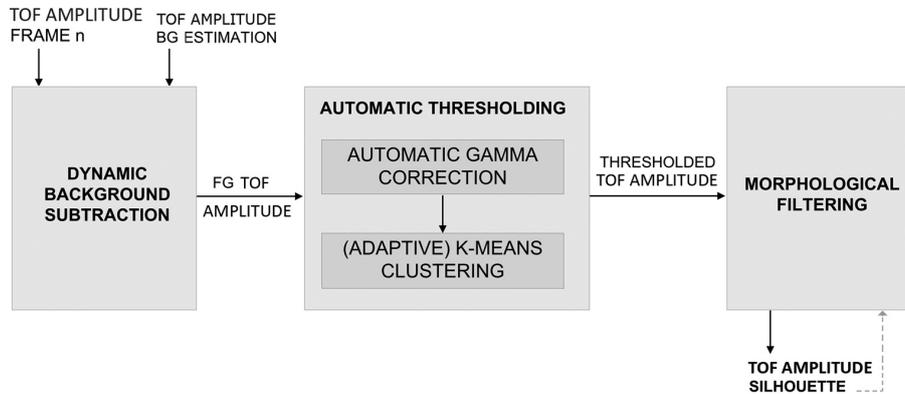


Figure 3.11: LWIR silhouette extraction.

Further on, we will discuss the visual and LWIR image registration using the visual and thermal body silhouettes. In the same way, image registration between other types of multi-modal images, e.g., visual-TOF and TOF-LWIR, can be performed. Due to the high similarity in the registration process, these other types of multi-modal image registrations will not further be discussed in detail. In our experiments (Section 3.3.5), we report on their performance.

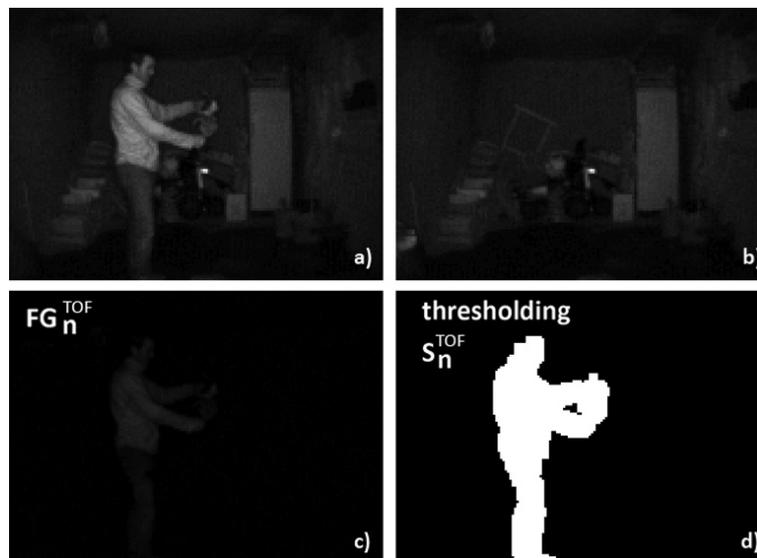


Figure 3.12: Example of TOF amplitude silhouette extraction: calibration person (a); BG estimation (b); BG subtraction (c); thresholding and morphological filtering (d).

3.3.4 Visual and LWIR image registration

After the extraction of the visual and thermal body silhouettes from the color image and its synchronous thermal image respectively, registration of both images is performed using a three-step registration algorithm. The goal is to determine the transformation parameters in order to align the LWIR with the visual image. Assuming that the distance between the cameras and the calibration person is large, the human surface from the camera view can be approximated as planar and the geometric transformation can be strictly represented by a projective transformation. Furthermore, assuming that the image planes of both visual and LWIR cameras are approximately parallel, the geometric transformation can be further simplified to a rigid transformation, which can be decomposed into a 2-D rotation, scaling and translation [71]. As such, a point (X, Y) in the visual image plane is transformed into the point (X', Y') in the thermal image plane as follows:

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = s \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} \Delta X \\ \Delta Y \end{pmatrix}, \quad (3.4)$$

where θ is the rotation angle, s is the scaling factor and $(\Delta X, \Delta Y)$ is the translation vector. A similar geometric transformation for image registration is also proposed by Liu et al. [136].

In order to estimate each of the three geometric parameters, i.e., rotation angle θ , scaling factor s and translation vector $(\Delta X, \Delta Y)$, the contours and the correlation of the visual and thermal silhouettes are analyzed, as is discussed in detail in the following subsections. First, the rotation is computed using silhouette contour extraction and circular cross correlation. Next, contour scaling is used to estimate the thermal-visual scale factor. Finally, the translation vector is estimated by maximization of binary correlation.

A. Contour vector generation

In order to estimate the rotation angle between the two silhouettes (\sim rotation angle between the two camera views), we propose to analyze the translation of the 1-D contour centroid distance (CCD) of both silhouettes. The contour centroid distance $CCD(u)$ (in which u is the index of the extracted boundary point) represents the distance between the boundary points $(x(u), y(u))$ and the centroid (x_c, y_c) of the silhouette. As such, the 2-D silhouette matching problem is converted to a one-dimensional signal matching problem, i.e., the matching of silhouette contours.

The 1D contour vector is generated from both visual and thermal LWIR silhouette using a *boundary extraction* algorithm [137], and by measuring the one-dimensional signal from the center of mass, i.e., the centroid, to the boundary for each silhouette [125]. The centroid (x_c, y_c) is computed as follows:

$$x_c = \frac{1}{N} \sum_{u=0}^{N-1} x(u) \quad y_c = \frac{1}{N} \sum_{u=0}^{N-1} y(u). \quad (3.5)$$

The $CCD(u)$ is computed as follows:

$$CCD(u) = \sqrt{(x(u) - x_c)^2 + (y(u) - y_c)^2}. \quad (3.6)$$

In Fig. 3.13, both the boundary extraction (Fig. 3.13a,b) and the one-dimensional visual and thermal CCD (Fig. 3.13c,d) of the calibration silhouettes are shown. Although visual inspection of the CCDs can already reveal a rough estimation of the rotation, automatic analysis on this 1-D signal is not straightforward due to the different number of boundary points of both CCDs. For direct comparison of both CCDs and in order to estimate the rotation and scale, they must have the same size. Therefore, we propose to convert each contour point $(x(u), y(u))$ from *Cartesian to polar* coordinates [138] using Eq. 3.7 and compute the one-dimensional $CCD^{polar}(u)$. This $CCD^{polar}(u)$ is obtained by computing the distance $r(u)$ ($= CCD(u)$) from the centroid (x_c, y_c) of the silhouette to the silhouette boundary as a function of the turning angle $\theta(u)$ ($-\pi \leq \theta(u) < \pi$):

$$\begin{aligned} r(u) &= \sqrt{(x(u) - x_c)^2 + (y(u) - y_c)^2} = CCD(u) \\ \theta(u) &= \tan^{-1}\left(\frac{y(u) - y_c}{x(u) - x_c}\right) \end{aligned} \quad (3.7)$$

$$CCD^{polar}(\theta(u)) = r(u).$$

The CCD^{polar} of the thermal and visual calibration silhouette are shown in Fig. 3.13e,f. Although the range of the CCDs is already equal ($[-\pi, \pi]$), the number of points in both signals is still different due to the fact that multiple boundary points can be detected under the same angle. To cope with this problem, we propose to perform a *radial vector analysis*, i.e., a novel CCD mapping technique which discretizes the CCD^{polar} signal over 64 equally spaced intervals.

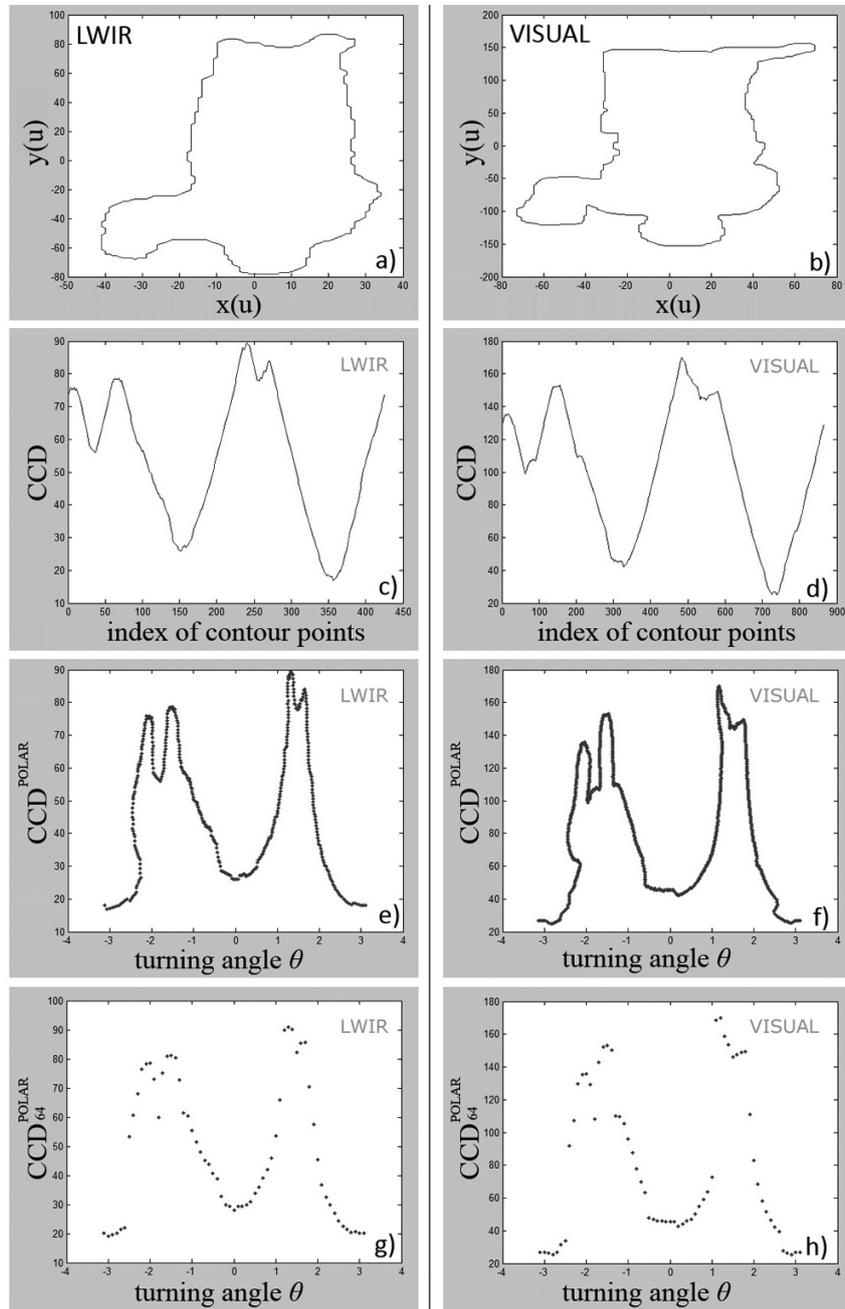


Figure 3.13: One-dimensional visual and thermal CCD of the calibration silhouettes: boundary extraction (a,b); CCD (c,d); polar CCD (e,f); discretized polar CCD (g,h).

The reason for choosing 64 intervals is that the turning angle $\theta(u)$ is quantized over $[-\pi, \pi[$ with a step-size of 0.1. Within each interval, the maximum $\max(CCD^{polar})$ in that interval is chosen as the representative boundary value for the interval, since those points best match the outer part of the silhouette, and as such, only a limited amount of information is lost. The resulting CCD_{64}^{polar} values are shown in Fig. 3.13g,h. Alternatively, it is also possible to super-sample the smallest signal, as in [125]. However, by converting to polar coordinates and quantize the signal, we reduce the 2D silhouette boundary to a 64-element vector and keep the computational cost low.

B. Contour mapping

Contour alignment (for rotation estimation)

Using the CCD_{64}^{polar} of the thermal and visual silhouette, the rotation of both camera views can easily be calculated by finding the translation which maximizes the thermal-visual CCD_{64}^{polar} correlation. This is based on the fact that translating the 1-D signals in centroid contour distance space over k locations corresponds to rotating the associated silhouette image in 2D pixel space over $k/64 * 360^\circ$. The thermal-visual CCD_{64}^{polar} translation is found by calculating the location k at which the circular cross-correlation $CXC(k)$ reaches its maximum. The circular cross-correlation [125, 139] of $CCD_{64}^{polar}(S_n^{LWIR})$ and $CCD_{64}^{polar}(S_n^{VISUAL})$ is defined by:

$$CXC(k) = \sum_{i=1}^{64} CCD_{64,i}^{polar}(S_n^{LWIR}) * CCD_{64,i \oplus k}^{polar}(S_n^{VIS}), \quad (3.8)$$

with $k = 0 \dots 63$ and $\oplus =$ addition modulo 64.

For the CCDs shown in Fig. 3.13, we found that CXC reaches its maximum for $k=0$. As such, the rotation angle between the thermal and visual silhouette equals $0/64 * 360 = 0$, as could be expected based on the rough visual estimation. In Fig. 3.14 we also show an example with rotated views. In this example, the thermal camera is placed upside down (i.e., the rotation angle = 180°). The ‘rotation estimation’ result of this experiment, which is 174° ($k = 31$), shows that the proposed technique is a good rotation estimator.

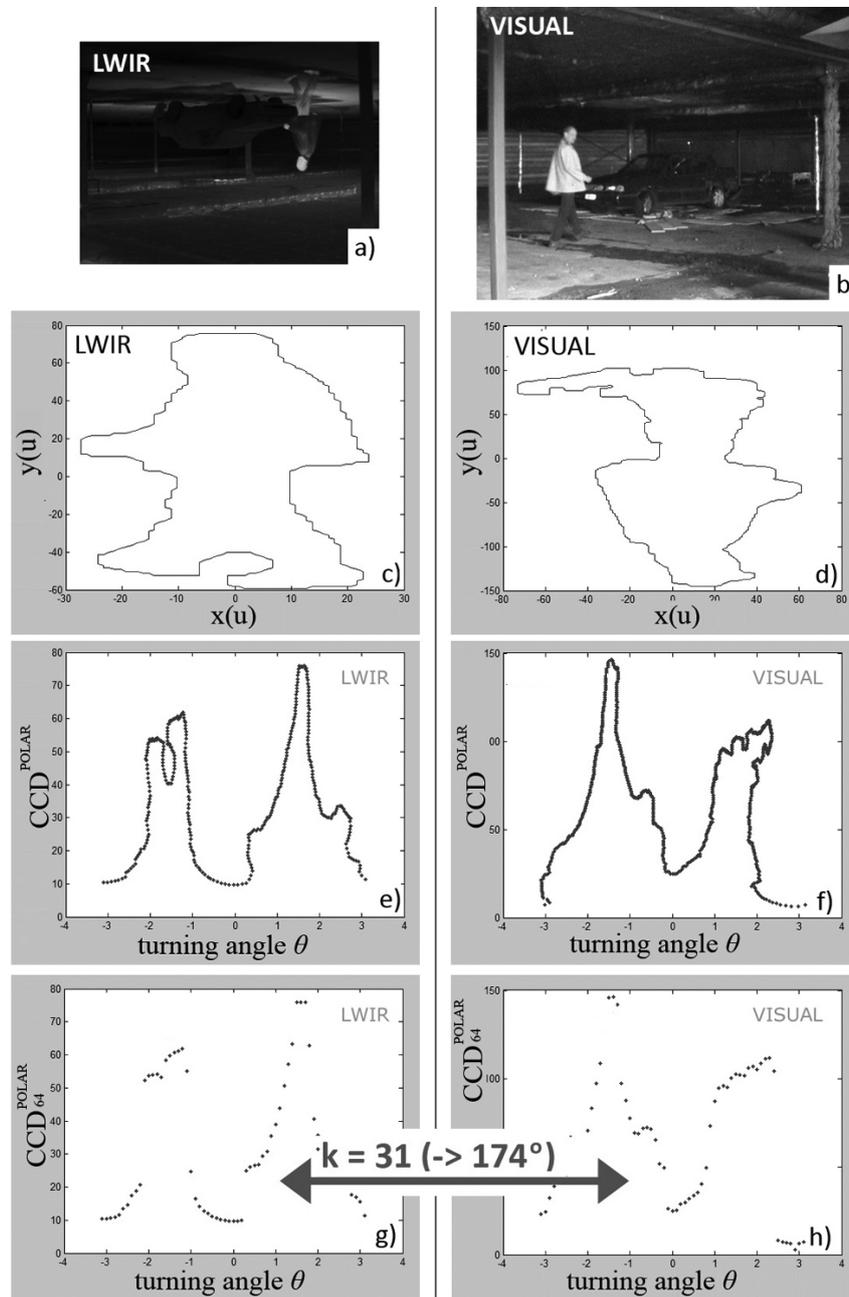


Figure 3.14: Rotation estimation for rotated thermal camera(a) and visual camera(b); boundary extraction (c,d); polar CCD (e,f); discretized polar CCD (g,h).

Contour scaling (for scale factor estimation)

After rotating, i.e., aligning the thermal and the visual CCD, the scale factor between both views is estimated by analyzing the ratio (CCD^{ratio}) of the thermal and visual aligned CCDs. The ratios for the calibration example are shown in Fig. 3.15. As can be seen in the image, the ratios are not constant and show some disorder. The reason for this behavior is twofold. First of all, the horizontal and vertical dimensions of both sensor images do not relate equally, which implies some deformation and influences the vertical-horizontal scale ratios. Second, the edge transitions in visual and thermal images are not always identical and, as such, the thermal and visual boundaries can differ. Furthermore, visual and thermal artifacts, such as (connected) shadows and reflections, can also increase the ratio disorder.

In order to cope with the CCD^{ratio} -related problems, we propose to use the median ratio as the scale factor. The main reason for choosing the median ratio instead of, for example, the mean ratio is that the median ratio is not influenced by outliers, while for the mean ratio this cannot be guaranteed. The calculation of s is shown in Eq. 3.9.

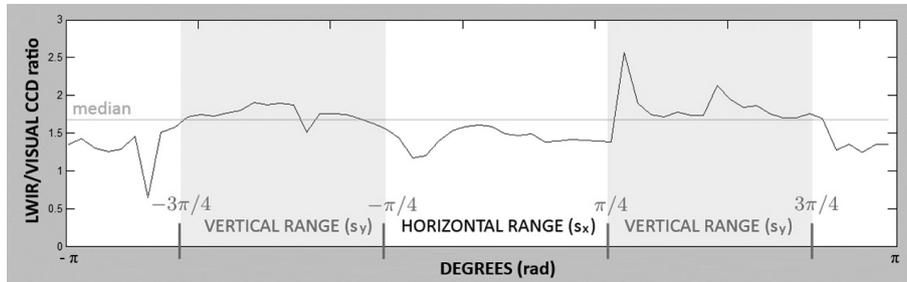


Figure 3.15: Scale factor estimation based on CCD ratio analysis.

Instead of using one scale factor s for both horizontal and vertical direction, it is also possible to use different scale factors s_x and s_y for each direction. For example, in case of reasonable vertical-horizontal deformation, due to non-parallel sensor placement or highly non-related sensor dimensions, different scale factors can be necessary to coarsely map the silhouettes. To estimate both the vertical and horizontal scale factor, we propose to use the median ratio in $[-3\pi/4, -\pi/4]$ and $[\pi/4, 3\pi/4]$ for s_y , and the median ratio in the remaining ranges for s_x (as illustrated in Fig. 3.15):

$$\begin{aligned}
 CCD^{ratio} &= \frac{CCD_{64}^{polar}(S_n^{LWIR})}{CCD_{64,k}^{polar}(S_n^{VISUAL})} \\
 s &= \text{median}(CCD^{ratio}) \\
 s_x &= \text{median}(CCD^{ratio}[\frac{-\pi}{4} : \frac{\pi}{4}, \frac{3\pi}{4} : \frac{-3\pi}{4}]) \\
 s_y &= \text{median}(CCD^{ratio}[\frac{-3\pi}{4} : \frac{-\pi}{4}, \frac{\pi}{4} : \frac{3\pi}{4}]).
 \end{aligned} \tag{3.9}$$

C. Silhouette mapping (for translation estimation)

The last transformation parameter, estimated by the registration algorithm, is the translation vector $(\Delta X, \Delta Y)$. Translation can occur due to the placement of the cameras, but also due to the different sensor resolutions, i.e., the image of one sensor can be a cropped version of the other. To correct this translation and to be able to perfectly align the thermal and visual image, the binary correlation technique (Fig. 3.16) proposed by Chen et al. [112] is used to determine the x- and y-displacements.

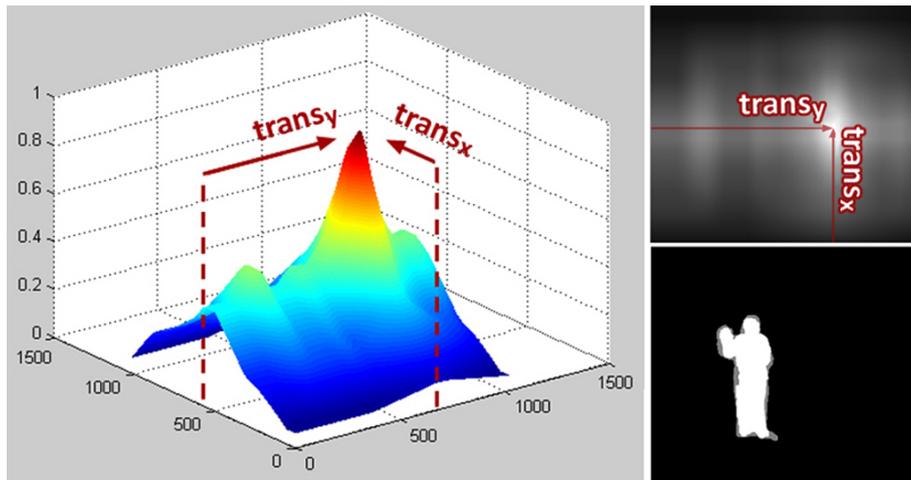


Figure 3.16: 2D/3D correlation-based translation estimation and registration result.

After rotating and scaling up the LWIR image using the estimated rotation angle θ and the scaling factor s , the translation vector $(\Delta X, \Delta Y)$ is computed by binary correlation, i.e., template matching, in the frequency domain. The correlation between the thermal image and the visual image is computed by rotating the thermal image 180° and then using the Fast Fourier Transform (FFT)-based convolution technique. This can be done since convolution is equivalent to correlation when rotating the kernel by 180° . Similar to [112], we represented the two levels of the silhouette images by -1 (BG) and 1 (FG), so that by maximizing the correlation function both parts are matched as much as possible. The 2D/3D result of correlating the thermal silhouette with the visual silhouette is shown in Fig. 3.16.

The point $(trans_x, trans_y)$, at which the correlation reaches its maximum, is used to calculate $(\Delta X, \Delta Y)$ as follows:

$$\begin{pmatrix} \Delta X \\ \Delta Y \end{pmatrix} = \begin{pmatrix} trans_x \\ trans_y \end{pmatrix} - \begin{pmatrix} size_x(S_n^{LWIR}) \\ size_y(S_n^{LWIR}) \end{pmatrix}, \quad (3.10)$$

with $size_x(S_n^{LWIR})$ and $size_y(S_n^{LWIR})$ the x and y dimension of S_n^{LWIR} .

The estimation of the translation vector finishes the proposed three-step registration algorithm, and using the retrieved transformation parameters θ , s and $(\Delta X, \Delta Y)$, registration between LWIR and visual images can be performed. As the registration result in Fig. 3.16 shows, the visual and thermal silhouette of the calibration object map coarsely. The overlapping part of the silhouettes is shown in white and the non-overlapping part is shown in gray.

3.3.5 Experimental results

A. Silhouette based LWIR-visual image registration

A visual, i.e., subjective, evaluation of the silhouette based LWIR-visual registration experiments in Fig. 3.17 already indicates that the proposed registration algorithm is able to coarsely align the thermal and visual images. However, in order to evaluate the registration more objectively, we propose to use the coverage metric COV which equals the percentage overlap between the thermal S_n^{LWIR} and visual S_n^{Visual} registered silhouettes:

$$COV(S_n^{LWIR}, S_n^{Visual}) = \frac{S_n^{LWIR} \cap S_n^{Visual}}{S_n^{LWIR} \cup S_n^{Visual}}. \quad (3.11)$$

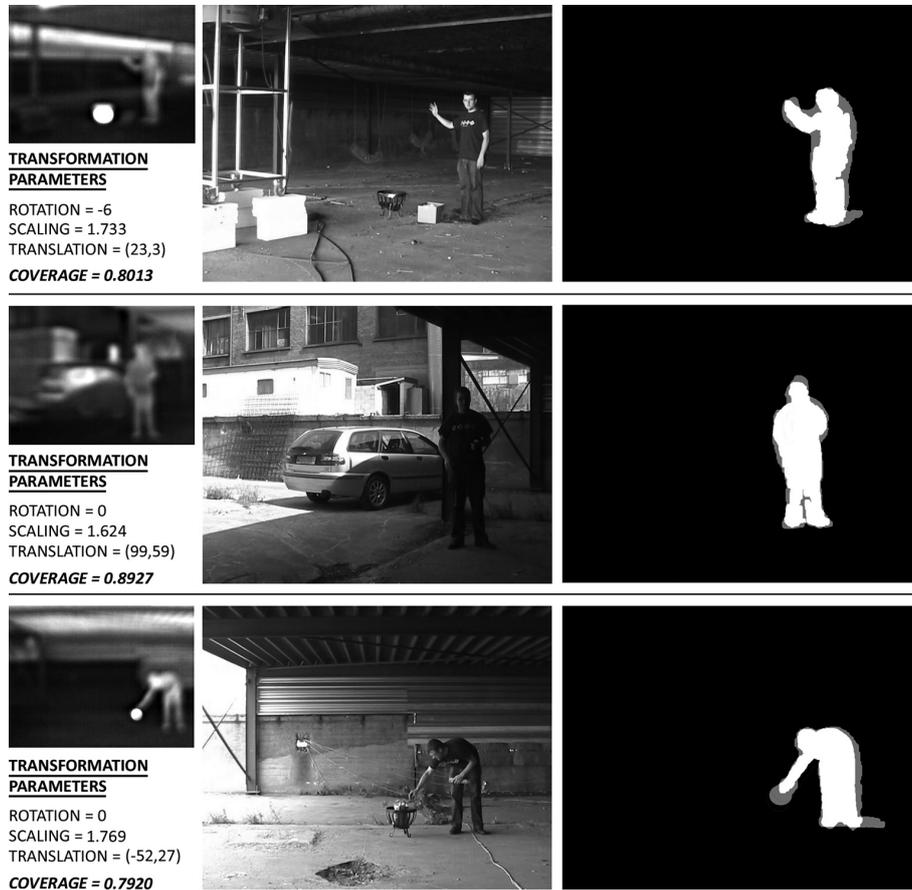


Figure 3.17: Experimental results of LWIR-visual registration.

Since COV depends on the performance of the silhouette extraction methods, one can also use the registration precision proposed in [71]. Similarly to Eq. 3.11, the registration precision is defined as $P(A, B) = (A \cap B) / (A \cup B)$, where A and B are manually labeled human silhouette pixel sets from the original visual image and the transformed thermal image, respectively. However, different to the approach of Han and Bhanu [71] where the registration is done manually, our method automatically calculates the COV precision.

The silhouette maps and the coverage results in Fig. 3.17 show that the proposed approach achieves good performance for image registration between color and thermal image sequences. The visual and IR silhouette of the person are coarsely mapped onto each other with an average coverage above 80%.

However, due to the individual sensor limitations, such as shadows in visual images, thermal reflections and soft thermal boundaries in LWIR, exact match is quasi impossible. For example, in Fig. 3.17, small artifacts at the boundary of the merged silhouettes can still be noticed. Also, if the cameras are not perfectly aligned, i.e., the assumption of parallel image planes is not satisfied, or if the vertical and horizontal dimensions of both sensors do not relate proportionally, deformation can arise between the detected objects and coverage can be low.

To cope with the above mentioned problems, the proposed approach can be extended using more complex moving object detectors and transformation models, such as, for example, is done in the work of Benezeth et al. [116], which is based on epipolar geometry. Further improvement can (possibly) also be achieved by averaging the results over multiple frames instead of using only one frame or by refining the registration results, for example by maximization of mutual information using the techniques described by Maes et al. [114] and Liu [136]. However, for our application, the average calibration coverage above 80% is sufficient. Also, compared to the results of related work, e.g., the registration method in [71], our proposed method achieves similar results. For more details, the reader is referred to [127, 128].

B. TOF-visual image registration

Similarly to the LWIR-visual registration experiments, tests were also performed on the automatic registration of TOF-visual images. Also here, the proposed registration method was used to automatically find the correspondence between the ‘moving’ silhouettes extracted from synchronous TOF and visual images. Again, we found both subjectively and objectively that the proposed registration algorithm is able to coarsely map the multi-sensor TOF and visual images. Some exemplary TOF-visual ‘mappings’ in Fig. 3.18 show the effectiveness of the TOF-visual registration.

Due to the absence of appropriate TOF-thermal video sequences, i.e., sequences captured with parallel TOF and thermal sensors whose lines of sight are close to each other, we have not yet evaluated the automatic registration of TOF-thermal images. However, similar results as the above are to be expected. This also holds for the combined registration of visual, thermal and TOF images.

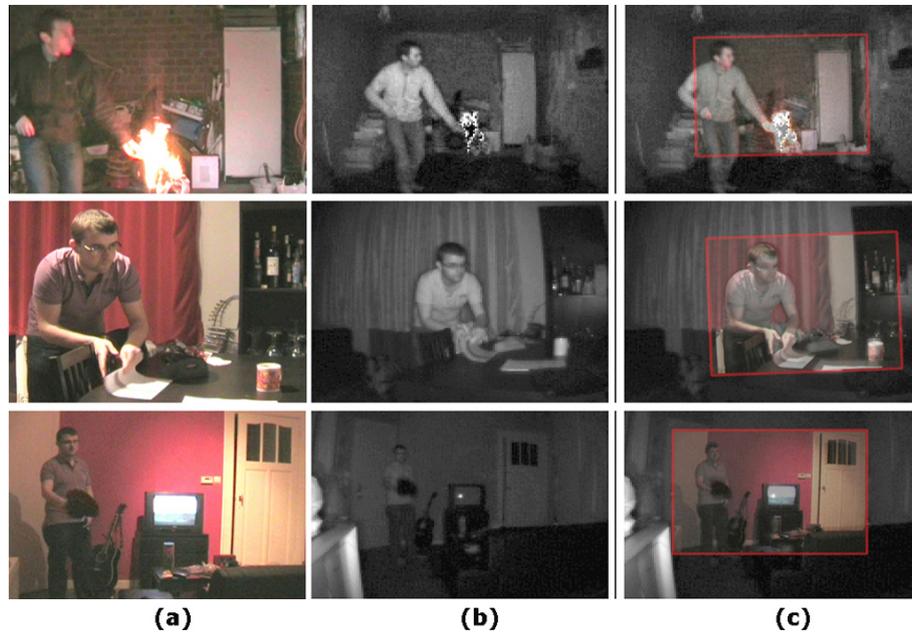


Figure 3.18: Examples of visual and TOF amplitude image registration: (a) visual and (b) TOF amplitude images; (c) registration check.

In the following sections, the proposed silhouette based registration algorithms are used to map the detection results of single-sensor LWIR, visual and/or TOF fire detectors on each other. As such, multi-modal fire detectors are created which outperform the single-sensor detectors discussed in Chapter 2. The first of these multi-modal detectors is a LWIR-visual flame detector, which combines the detection results of a thermal LWIR detector and a visual detector in order to improve the accuracy of each of the individual detectors.

3.4 LWIR-visual flame detection

The multi-modal LWIR-visual flame detector, shown in Fig. 3.19, first searches for candidate flame objects in both LWIR and visual images by using moving object detection and flame feature analysis. These steps have already been discussed in Section 2.4 and Section 2.3.1 of the previous chapter. Next, it uses the registration information, i.e., rotation angle, scale factor and translation vector, to map the LWIR and visual candidate flame objects on each other (Section 3.3). Finally, the global flame risk value is calculated using the risk values of the mapped objects. In case objects are detected with a high combined multi-sensor risk value, a fire alarm is given.

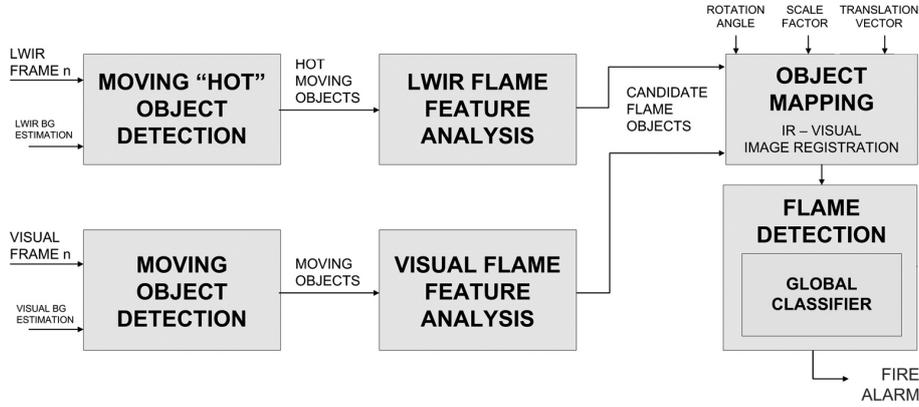


Figure 3.19: Multi-modal LWIR-visual flame detection.

3.4.1 Global flame risk value

As discussed in the previous chapter, each of the visual and LWIR flame features possesses a value between 0 and 1, indicating whether the object has the flame characteristic. By averaging these flame values, the risk values P_{flame}^{LWIR} and P_{flame}^{visual} are retrieved, which indicate whether the object should be classified as flames in the respective spectral range. The global flame risk value combines these two risk values, using:

$$P_{flame} = \beta * P_{flame}^{LWIR} + (1 - \beta) * P_{flame}^{visual}, \quad (3.12)$$

into an overall flame risk value P_{flame} . The parameter β in this equation is a weight factor that specifies how much of P_{flame}^{LWIR} and P_{flame}^{visual} must be taken into account in the overall flame risk calculation. Depending the circumstances, e.g., night or day, an appropriate β value can be chosen. The (automatic) selection of such an appropriate β value is related to the general remark in Section 2.7 on context-dependent feature weights. At the end, the overall risk value P_{flame} is compared to an alarm threshold t_{flame} . If the flame risk value exceeds this threshold, a fire alarm is raised. In our experiments it was found that a good value for t_{flame} is 0.7. The sensitivity of this threshold, however, needs to be further investigated in future work. Most important is that, based on the risk value, operators can concentrate their attention on the sequences which most probably contain flames. Furthermore, it is important to mention is that changing the β value in Eq. 3.12 to 0 or 1, transforms the multi-sensor detector into a standalone visual or LWIR detector respectively.

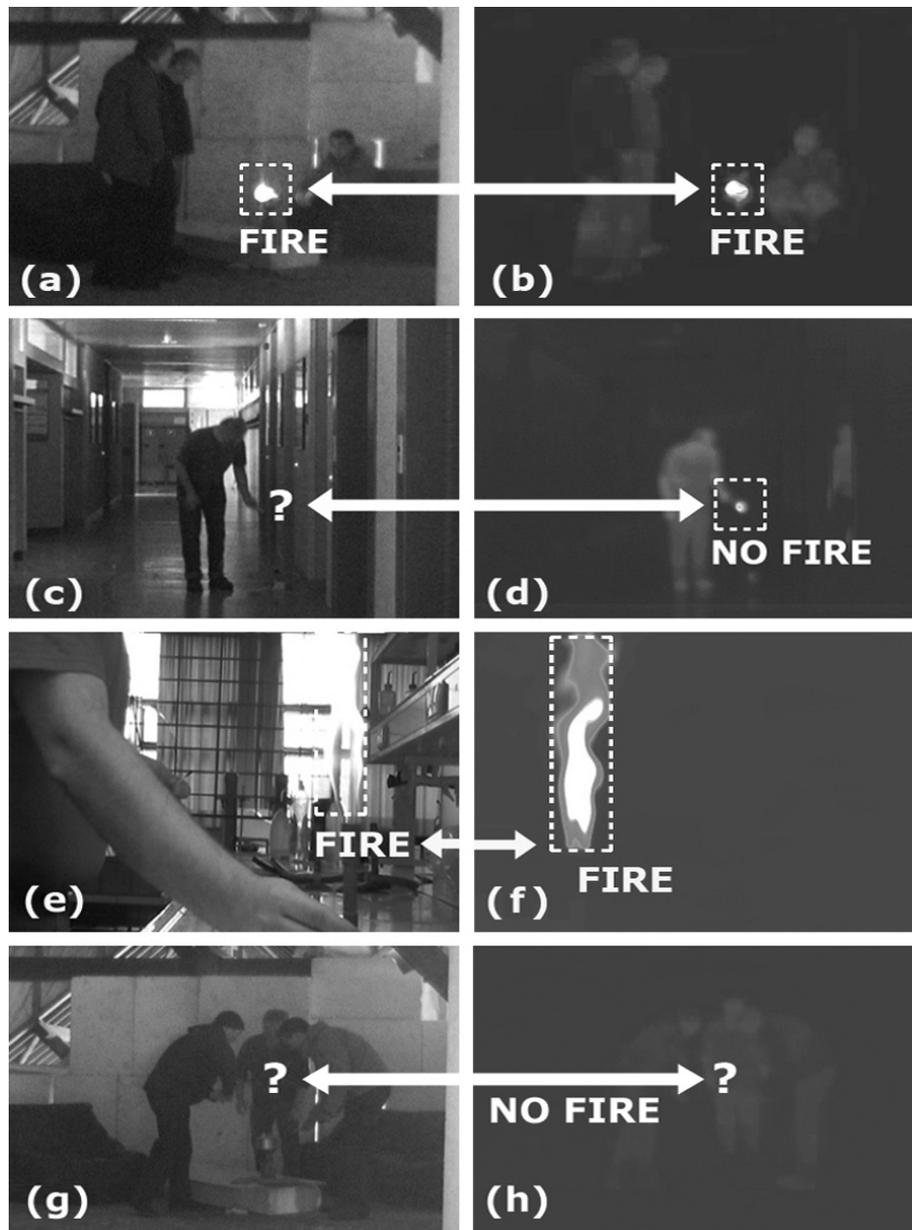


Figure 3.20: Multi-modal LWIR-visual flame detection results of visual (a,c,e,g) and LWIR (b,d,f,h) Attic, Corridor and Lab sequence.

3.4.2 Experimental results

In order to verify the proposed LWIR-visual flame detector we performed several fire and non-fire experiments. Exemplary shots of these experiments are shown in Fig. 3.20. The multi-modal sequences were acquired by a Xenics Gobi-384 LWIR camera and a Canon MD110 camera, which work in the $8 - 14\mu\text{m}$ spectral range and the visual spectrum respectively. The Gobi thermal imager has a resolution of 384×288 pixels and a frame rate of 28-30fps. The Canon's resolution is 576×720 and its frame-rate is 25fps. In order to cope with the different frame rates and resolutions, and also with the differences in the field of view of the cameras, the LWIR-visual frames are spatio-temporal registered using temporal frame alignment and the proposed silhouette-based registration.

As can be seen in Table 3.1, the multi-sensor flame detector yields better results than the LWIR detector alone (Table 2.5). In particular for uncontrolled fires, a higher flame detection rate with fewer false alarms is achieved. Compared to the rather limited results of standalone visual flame detectors [1], the multi-sensor detection results are also more positive. As such, the combined detector is a win-win. As the images of the experiments (Fig. 3.20) show, only objects which are detected as fire by both sensors do raise the fire alarm.

3.5 LWIR-visual smoke detection

In the previous section we described the working principle of a novel multi-modal LWIR-visual flame detector and evaluated its performance over individual video flame detectors. Similarly, this section presents a novel LWIR-visual smoke detector, which also takes advantage of the different kinds of information represented by visual and thermal LWIR imaging sensors.

The LWIR-visual smoke detector analyzes the silhouette coverage of moving objects in visual and long-wave infrared registered (\sim aligned) images. The registration is also performed using the proposed silhouette-based image registration method (Section 3.3) which detects the rotation, scale and translation between moving objects in the multi-spectral images. The geometric parameters found at this stage are then further used to coarsely map the silhouette images and coverage between them is calculated. Since smoke is invisible in long-wave infrared its silhouette will, contrarily to ordinary moving objects, only be detected in visual images. As such, the coverage of thermal and visual silhouettes will start to decrease in case of smoke.

Table 3.1: Performance evaluation of multi-sensor LWIR-visual frame detection.

Video sequence (# frames)	# fire frames (GT)	# detected fire frames	# false positive detections	frame detection rate *
Attic - fire (337)	264	259	6	0.96
Attic - fire and people (2123)	1461	1352	19	0.91
Attic - moving people (886)	0	5	5	-
Lab - Bunsen burner (115)	98	74	0	0.75
Corridor - hot object (184)	0	3	3	-

* $detection rate = (\# detected fire frames - \# false alarms) / \# fire frames$

Due to the dynamic character of the smoke, the visual silhouette will also show a high degree of disorder. By focusing on both the visible-invisible character of smoke in visual-LWIR images and visual smoke silhouette disorder behavior, the system is able to accurately detect the smoke and to distinguish between smoke and non-smoke moving objects. Experiments on smoke and non-smoke multi-sensor sequences indicate that using the low-cost silhouette analysis, a fast warning, with a low number of false alarms, can be given. It is important to mention that, as smoke becomes more and more transparent further in the infrared spectrum, IR cameras in the long wave IR range (LWIR, $8 - 12\mu\text{m}$) have the highest added value for detecting smoke. As is illustrated in Fig. 3.21, a LWIR camera can even look through the smoke. By focusing on the visible-invisible character of smoke in visual-LWIR images, our multi-sensor detector can detect the smoke very accurately.

The detection algorithm, shown in Fig. 3.22, starts with the similar moving object silhouette extraction (Section 3.3) as the one used for image registration in LWIR-visual flame detection (Section 3.4). Then, it uses the registration information, i.e., the rotation angle, the scale factor and the translation vector, to map the thermal and visual silhouette images onto each other. Next, the coverage of the resulting thermal-visual silhouette map is computed and is analyzed over time using the silhouette coverage analysis (SCA). This SCA is the first phase of our novel two-phase decision algorithm. The SCA focuses on the silhouette coverage of the thermal-visual registered images and gives a kind of first smoke warning when a decrease in silhouette coverage occurs. In the second phase, the smoke warning is further investigated by analyzing the disorder characteristics of the visual silhouette S_n^{Visual} . If this silhouette shows a high degree of disorder, the smoke hypotheses is confirmed and a fire alarm is raised. In the next few sections we will discuss each of these phases more in detail and evaluate their performance in real-world experiments.

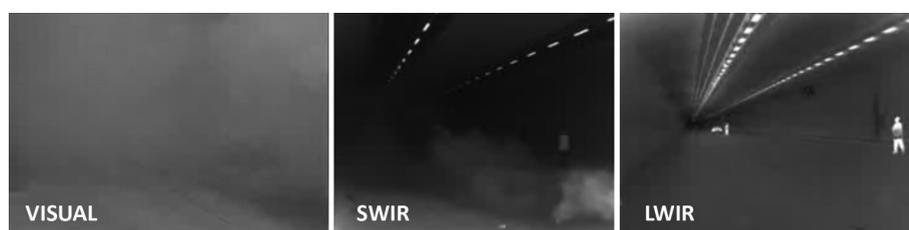


Figure 3.21: Smoke transparency in visual, short wave infrared (SWIR) and long wave infrared (LWIR) range. (source: www.xenics.com)

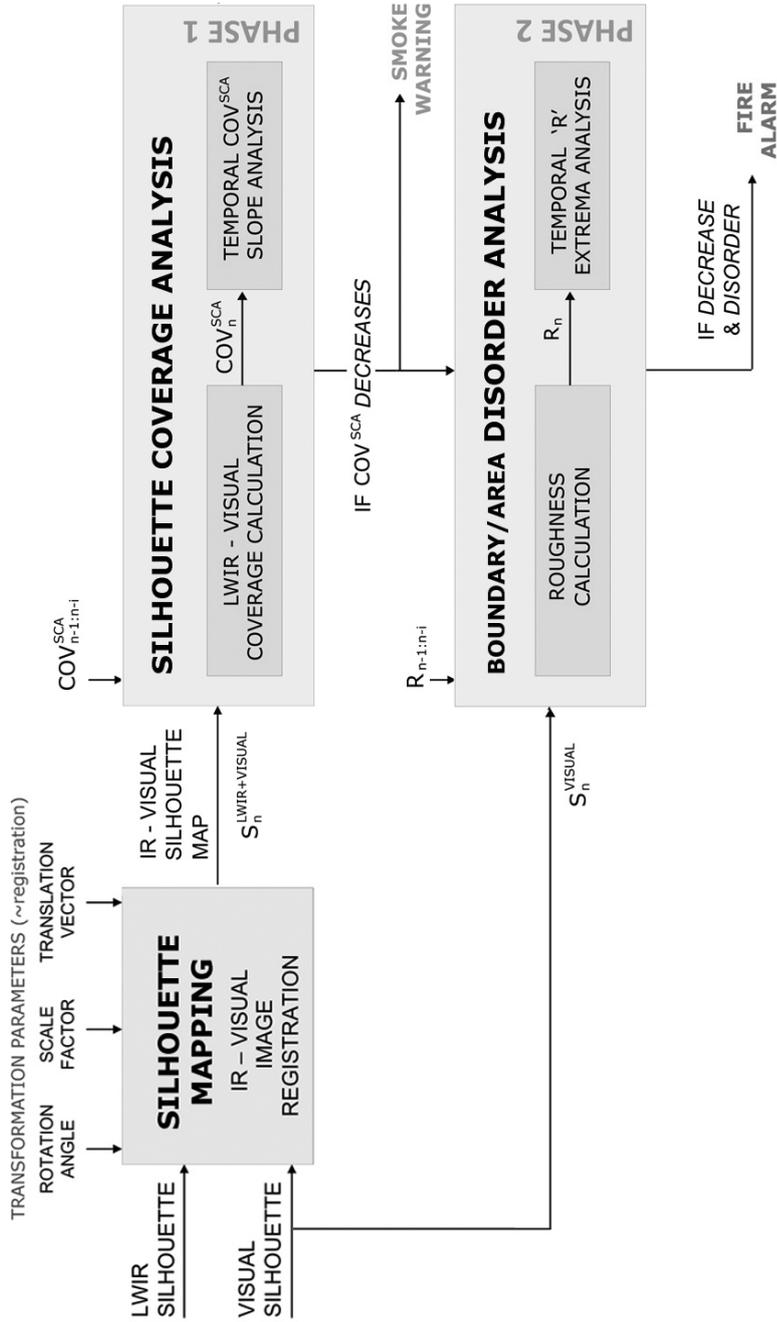


Figure 3.22: Multi-sensor LWIR-visual smoke detection using 2-phase smoke detection algorithm: (PHASE 1) temporal slope based silhouette coverage analysis; (PHASE 2) boundary/area disorder analysis.

3.5.1 Phase 1: silhouette coverage analysis

The SCA starts with the calculation of the LWIR-visual coverage of the registered visual S_n^{VISUAL} and thermal silhouette S_n^{LWIR} . Contrary to the COV coverage metric introduced for registration in Eq. 3.11, the SCA uses a slightly different metric, since we are only interested in the percentage of the visual silhouette that is also detected by the thermal silhouette. The SCA coverage metric COV^{SCA} is defined as:

$$COV^{SCA}(S_n^{LWIR}, S_n^{Visual}) = \frac{S_n^{LWIR} \cap S_n^{Visual}}{S_n^{Visual}}. \quad (3.13)$$

Under normal conditions, if there is no smoke, the COV^{SCA} does not change much over time. This is also shown by the silhouette coverage graph of the moving person sequence in Fig. 3.23a, where the COV^{SCA} stays within the $[0.8; 1]$ coverage range. Contrarily, in the case of smoke (Fig. 3.23b), the COV^{SCA} strongly decreases below 0.8. Even when no moving objects are present in the scene ($COV^{SCA} = 1$), a similar decrease is noticeable when smoke occurs. For detection of this decrease, we propose a sequence/scene independent technique based on slope analysis of the linear fit, i.e., trend line, over the ten most recent silhouette coverage values. If the slope of this trend line is negative and decreases continuously, smoke warning is given.

Since it is the global trend of a sequence of adjacent points which is analyzed, low noise coverage results do not cause any problems. Furthermore, the delay caused by analyzing the set of adjacent points is negligible. Since the algorithm is able to run at 25 fps, a delay of 10 frames, for example, does not much affect the ‘real-time’ character of the proposed method.

The trend line, i.e., the linear fit of consecutive coverage results, is found by linear regression [140]. Suppose there are n data points $[COV_i^{SCA}, x_i]$ where $i = 1, 2, \dots, n$ and $x_i = i$. The goal is to find the equation of the straight line $COV^{SCA} = \alpha + \beta x$ which would provide a best fit for the data points, i.e., the line which minimizes the sum of squared residuals of the linear regression model. Using the least squares method the problem can be formulated as follows:

$$\begin{aligned} &\text{Find } \min_{\alpha, \beta} Q(\alpha, \beta) \\ &\text{where } Q(\alpha, \beta) = \sum_{i=1}^n (COV_i^{SCA} - \alpha - \beta x_i)^2. \end{aligned} \quad (3.14)$$

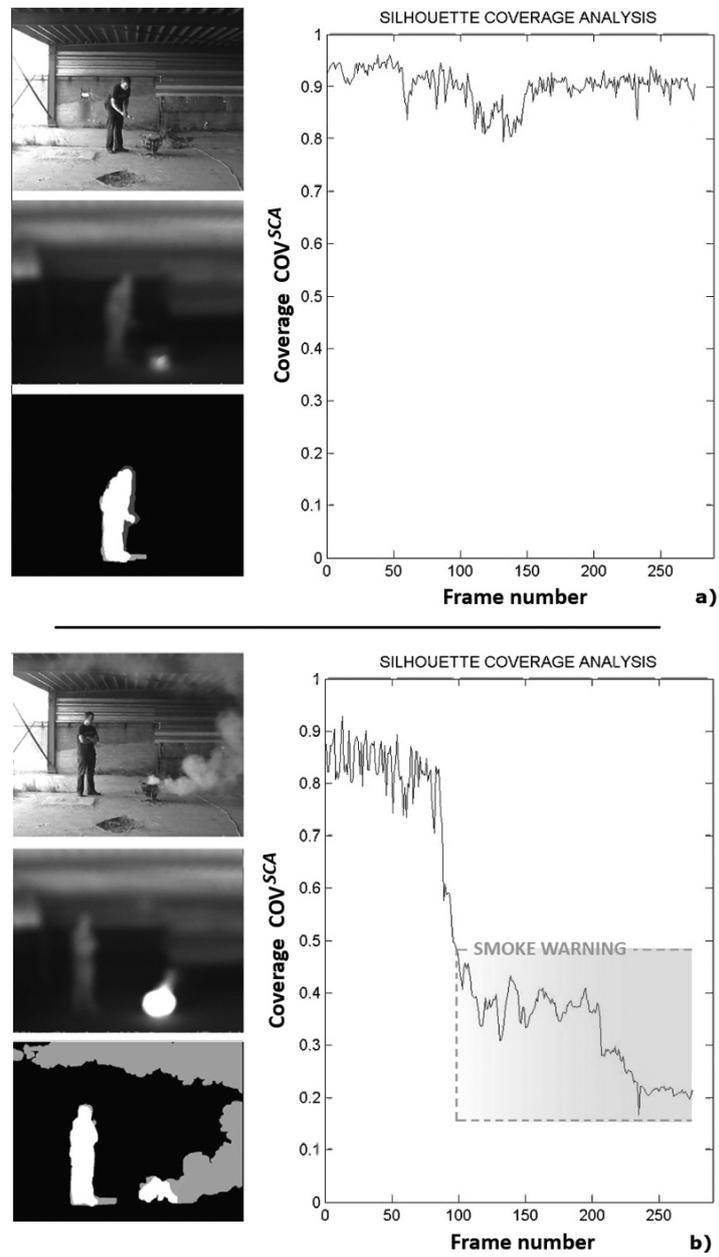


Figure 3.23: Silhouette Coverage Analysis (SCA). a) SCA of moving person sequence; b) SCA of smoke (straw fire) sequence.

It can be shown [140] that the values of α and β that minimize Q are:

$$\begin{aligned}\beta &= \text{corr}(x, COV^{SCA}) \frac{\sigma(COV^{SCA})}{\sigma(x)} \\ \alpha &= \overline{COV^{SCA}} - \beta \bar{x}.\end{aligned}\quad (3.15)$$

where $\text{corr}()$ is the correlation coefficient, $\sigma()$ is the standard deviation and \bar{x} and $\overline{COV^{SCA}}$ are the means of x and COV^{SCA} . Substituting the values of α and β in $COV^{SCA} = \alpha + \beta x$ provides the equation of the trend line.

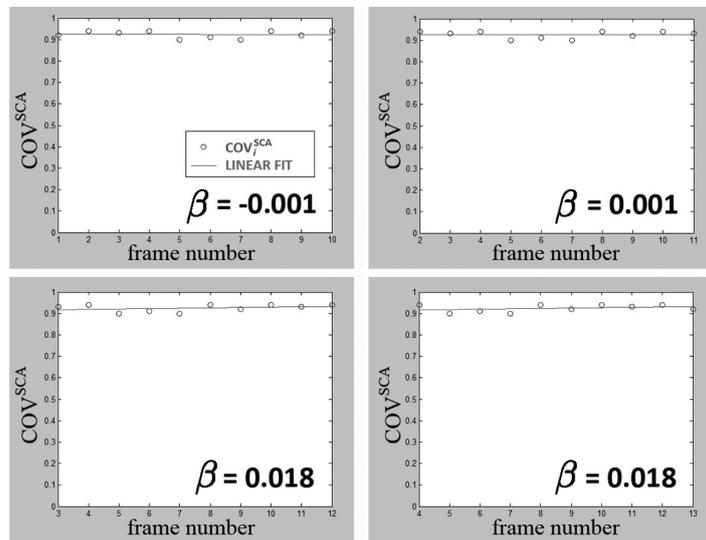
A positive slope of the trend line indicates that the line increases, whereas a negative slope indicates a decrease. As such, in order to detect a continuous decrease in silhouette coverage, it is sufficient to analyze the slope over time. If more than two consecutive slope values are negative and grow in the negative direction, smoke warning is given. Since the silhouette coverage of ordinary objects can also have a small negative slope over time, due to the thermal-visual differences, small negative slopes ($\beta > -0.1$) are not taken into account in the slope analysis. The analysis of the slope also causes a small delay. However, this delay of three frames is also negligible.

An example of the slope analysis for 4 consecutive frames from the moving person sequence and the smoke sequence is shown in Fig. 3.24. The slope for the moving person is very small and does not change much over time. Contrarily, in the smoke sequence, the slope becomes negative and grows in the negative direction as soon as smoke occurs. After more than two consecutive negative slope decreases, the smoke warning is given.

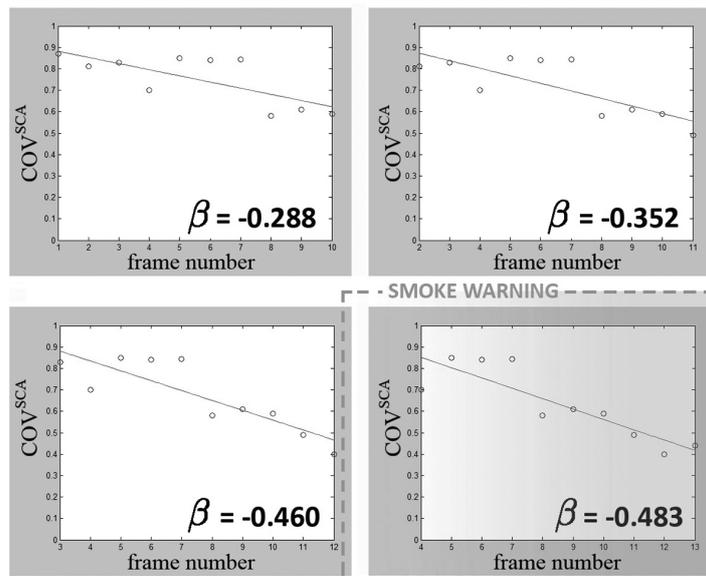
3.5.2 Phase 2: disorder analysis of visual silhouette

The second phase of the multi-sensor smoke detection is only executed if a smoke warning is given in the first phase. If a warning is given, foreground (FG) objects in the visual silhouette are further investigated by temporal disorder analysis in order to distinguish true detections from false alarms, such as shadows. Due to the dynamic character of smoke, the perimeter and the area of FG smoke objects in the visual silhouette S_n^{Visual} show a high degree of disorder. By temporal analysis of the boundary-area roughness R [40], which focuses on both the area A and perimeter P of the FG object, this disorder can be detected. The R of a FG object in S_n^{Visual} is given by:

$$R = \frac{P}{2\sqrt{\pi A}}. \quad (3.16)$$



a) moving person sequence - COV^{SCA} slope (β) analysis



b) smoke sequence - COV^{SCA} slope (β) analysis

Figure 3.24: Slope analysis for a) moving person and b) smoke sequence (\sim Fig. 3.23). Graphs show frame number versus visual/thermal coverage COV^{SCA} for four consecutive frames. If more than two consecutive negative slope decreases occur, smoke warning is given.

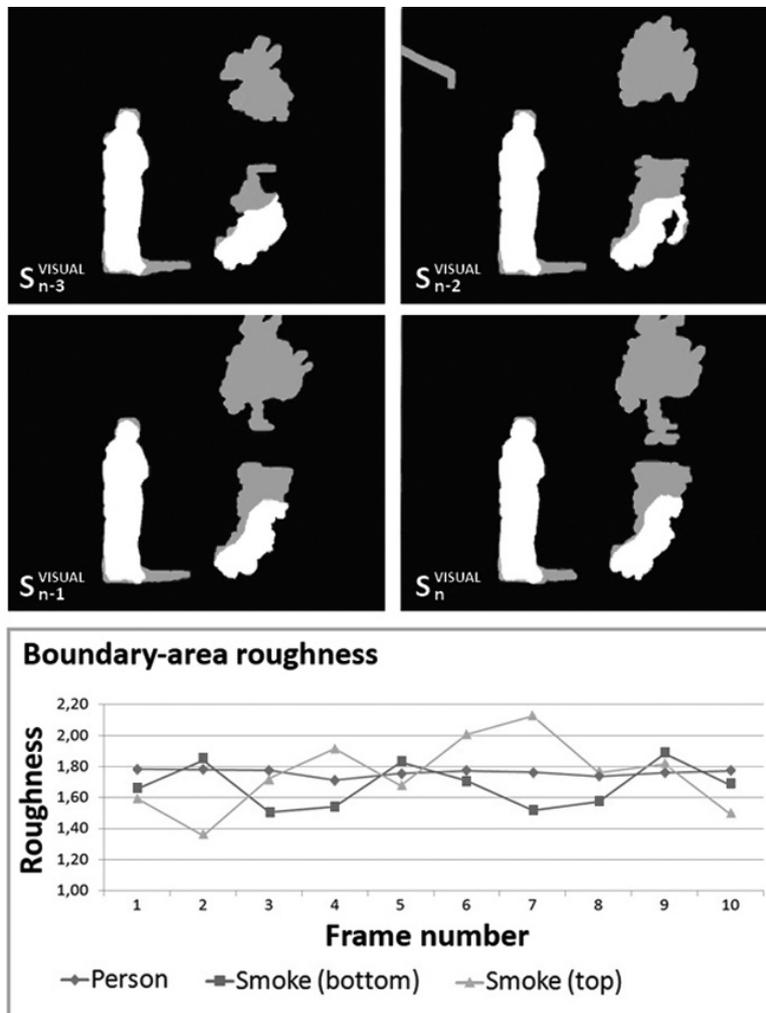


Figure 3.25: Boundary-area roughness for moving person and smoke objects.

As Fig. 3.25 shows, the boundary-area roughness of both smoke objects shows a high temporal disorder, while the disorder for the person remains quasi constant. For each object, its degree of disorder can automatically be detected by low-cost extrema analysis [94] using the roughness variance metric:

$$R_{var} = \frac{|extrema(R)|}{N/2}, \quad (3.17)$$

which is related to the number of extrema $|extrema(R)|$, i.e., local maxima and minima, in the set of N consecutive R data points.

By smoothing these data points using a moving average filter, small differences between consecutive points are filtered out and are not taken into account in the extrema calculation, which increases the strength of the disorder feature. Smoke, with a high roughness disorder, will have a R_{var} close to 1, while for more static objects it will be close to 0. Important to mention is that, in order to keep the delay low, N is chosen equal to the number of coverage points for trend line analysis. Since both can be calculated simultaneously, the buffering and (optional) analysis of the boundary-area roughness values causes no extra delay.

If for one (or more) FG object(s) R_{var} is high, i.e., close to 1, the fire alarm is raised. If necessary, further analysis of the visual silhouette using other low cost smoke-features can be performed. However, by only focusing on both proposed silhouette behaviors, the multi-sensor smoke detector is already able to accurately detect the smoke, as shown by the experimental results below.

3.5.3 Experimental set-up and results

In order to verify the proposed LWIR-visual smoke detector we performed several real-life fire and non-fire experiments in a closed car park at Warrington-FireGent [28]. An example of these real case scenarios is shown in Fig. 3.26a-b, where the left-most images are the visual and LWIR camera views of the 'moving people' and 'car fire' test sequence.

As the graphs in Fig. 3.26 show, the moving people sequence has a quasi constant silhouette coverage, and as such, no smoke warning is given so phase 2, i.e., the visual disorder analysis, is not performed. Contrarily, the silhouette coverage of the smoke sequence shows a high decrease after 45 frames, which activates the smoke warning. As a reaction to this warning, phase 2 is activated and analyzes the boundary-area roughness variance R_{var} of the visual silhouette objects. Since the R_{var} for the largest object is high, fire alarm is given.

In order to objectively evaluate the proposed method, we performed five different test setups: car fire, straw fire, moving people, moving car, and paper fire. For each of these fire and non-fire test setups we generated several video sequences. In total the test set contains 18 multi-modal fire videos and 13 non-fire video sequences with varying environment characteristics. For each of these sequences, we also generated a manual ground truth (GT). The performance results in Table 3.2 summarize the experimental results of all these tests for different algorithm configurations. Each of these configurations will be analyzed in more detail below.

Table 3.2: Fire and non-fire test results of different configurations of proposed algorithm.

configuration	FIRE TESTS			NON-FIRE TESTS		
	smoke warning	fire alarm	false alarm	smoke warning	smoke warning	false alarm
<i>proposed approach</i>	98%	98%	2%	5%	5%	2%
without gamma correction	74%	71%	5%	7%	7%	5%
without visual merge	67%	63%	4%	8%	8%	4%
without adaptive k-means	93%	91%	2%	7%	7%	2%
silhouette-based FG extraction alternatives						
frame differencing and dilation	69%	66%	7%	11%	11%	7%
dynamic background subtraction and dilation	76%	75%	4%	6%	6%	4%
silhouette disorder analysis alternatives						
randomness of area size disorder	98%	95%	3%	5%	5%	3%
similarity disorder of distance transformations	98%	97%	2%	5%	5%	2%

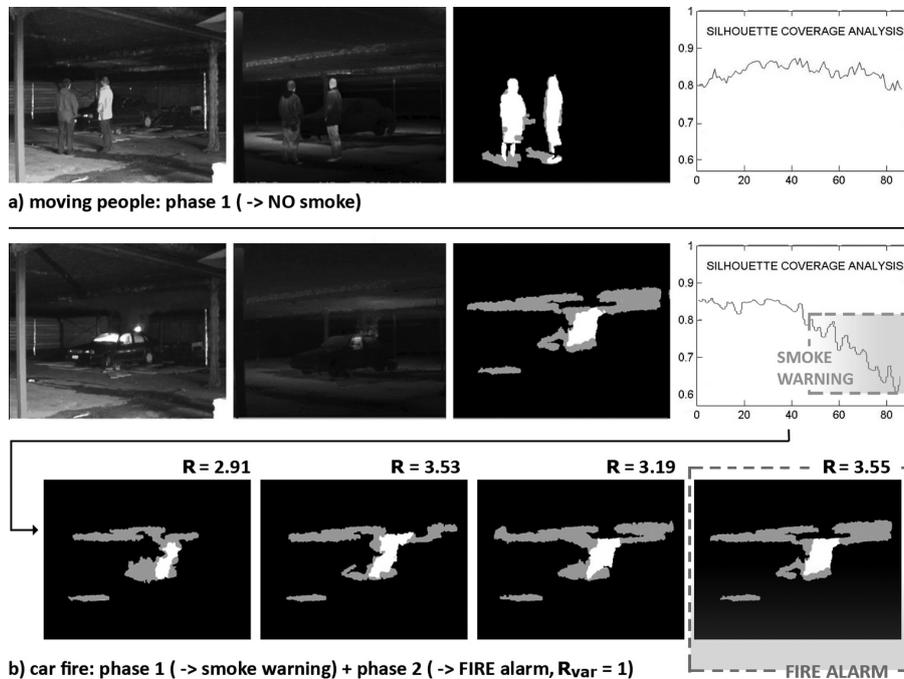


Figure 3.26: Multi-sensor smoke detection experiments: a) (non-smoke) moving person and b) smoke (car fire) sequence.

A. Evaluation of different algorithm configurations

During the tests, four different configurations of the algorithm were tested to evaluate and justify the steps taken in the proposed approach: proposed setup without gamma correction; without merge of visual color, edge and intensity information (only intensity was used); without adaptive k-means clustering (as an alternative, we used the Otsu method); and the proposed approach.

As the results indicate, the proposed configuration yields the best smoke warning / fire alarm rate. For the fire tests, the overall fire alarm rate is 98%. This means that almost each of the tested fires, which were manually annotated during ground truth (GT) creation, is detected. Without gamma correction or visual merge, these results are significantly lower. It can also be seen that the influence of the adaptive k-means is not so big, but since this kind of automatic thresholding results in an extra seven percent gain, its use in the proposed approach is justified.

By comparing the smoke warning and the fire alarm rate of the non-fire tests, the influence of the visual disorder analysis, i.e., the second phase in the smoke detector, becomes also visible. As most of the non-fire test sequences which falsely generate a smoke warning are corrected by the visual disorder analysis, the fire alarm percentage for the proposed approach is close to zero. As such, the number of false alarms is very low, one of the main requirements mentioned in the introduction. Only if the moving object has similar temperature profile as the background AND its Boundary Area Roughness is high, a false alarm will occur. Besides smoke, it is expected that not much objects will have both characteristics simultaneously.

B. Evaluation of silhouette FG extraction alternatives

In order to evaluate the effectiveness of the proposed foreground extraction, the results of the silhouette-based approach are compared to a simple frame differencing and dilation algorithm. Furthermore, comparison is also made with a popular running average based dynamic background subtraction algorithm [60].

The results in Table 3.2 show that the detection results of the proposed method outperform the results of the dynamic background subtraction, which on his turn achieves better results than the simpler frame difference approach. Especially when light conditions are bad, like in the car park experiments, the frame differencing and the dynamic BG subtraction have a lot of FG detection problems, which do not occur when using the proposed method. As such, the use of the silhouette-based FG extraction is objectively found more efficient than the investigated FG extraction alternatives.

C. Evaluation of disorder analysis alternatives

Over the last decade, many solutions for object shape changing detection have been proposed in literature, e.g., randomness of area size, boundary (area) roughness and the similarity disorder of distance transformations. In previous work [1], we already discussed some of these state-of-the-art disorder detection metrics and stated that their performance is quasi identical. Recent experiments, of which the results are shown in Table 3.2, also confirm this hypothesis. It is found that, although each of these shape changing detection techniques differ in definition, the outcome of each of them is almost identical. Furthermore, the experiments revealed that the disorder analysis of the boundary (area) roughness and the randomness of area size are computationally more efficient than the distance transformation technique. As such, one of these prior techniques is chosen.

D. Comparison with SOTA alternatives

As can be seen in Table 3.3, the proposed 2-phase multi-sensor detector yields good detection results, which outperforms the investigated state-of-the-art smoke detectors of:

- Xiong et al. [40]: based on BG subtraction and flicker/disorder analysis;
- Calderara et al. [42]: based on mixture of Gaussians (MoG) of DWT energy variation and color blending;
- Toreyin et al. [36]: based on block-based spatial wavelet analysis and Hidden Markov Model (HMM).

Especially when light conditions are bad, like in the car park experiments, the proposed algorithm detects the smoke more accurately. The main reason for this can be found in the fact that, contrarily to the SOTA (visual) clues that can be used for smoke detection, the change rate of FG area between visual and IR video images focuses on the visible-invisible character of visual-LWIR smoke regions, which is much less prone to misdetections. From these alternatives, the method of Toreyin et al. performs best. In the FireSense project [109], our method will further be evaluated against this method.

Since other aerosols such as fog and dust can possess similar visual-LWIR silhouette behavior as smoke, further visual object investigation, for example by energy (\sim visual obscuration) analysis [35, 42, 61] and dynamic texture analysis [68, 141, 142], can be necessary to eliminate those phenomena. However, this is out of the scope of this dissertation and will be part of future work.

Table 3.3: Comparison of proposed algorithm to state-of-the-art smoke detectors of Xiong et al. [40], Calderara et al. [42] and Toreyin et al. [36].

configuration	FIRE TESTS		NON-FIRE TESTS	
	smoke warning	fire alarm	smoke warning	false alarm
<i>proposed approach</i>	98%	98%	5%	2%
Xiong et al.	71%	68%	9%	7%
Calderara et al.	85%	83%	6%	4%
Toreyin et al.	88%	86%	7%	4%

3.6 Visual - TOF flame detection

In Section 3.4 we proposed a novel multi-modal LWIR-visual flame detector and evaluated its performance over individual video flame detectors. Similarly, this section presents a novel visual-TOF flame detector, which takes advantage of the different kinds of information represented by registered visual and TOF imaging sensors. The registration between the sensors is also performed using the silhouette-based image registration method proposed in Section 3.3.

The visual-TOF flame detector builds on the proven concepts of the single-sensor TOF flame detector proposed in Section 2.6.1. This TOF flame detector is based on the combined analysis of the depth and amplitude images of a TOF camera. Experiments (Section 2.6.1) revealed that, using this multi-modal TOF information, flames can be detected very accurately. One of the drawbacks of this detector, however, is that it is limited to indoor detection within the range of the TOF camera. In outdoor situations or outside the range of the TOF camera (distance $\geq 10\text{m}$), the detection fails. Main reason of this failing is the fact that the depth maps become unreliable under these circumstances. In order to cope with this problem one could think of only using the TOF amplitude information. However, relying on this feature alone causes a lot of mis-detections. A better approach is to combine the TOF amplitude detection (Section 2.6.1) with a visual flame detector (Section 2.3), what is done in the visual-TOF flame detector presented in this section. Since the amplitude image and the visual image focus on different ‘characteristics’ of the fire, we believe that combining both image modalities is definitely a win-win.

3.6.1 General scheme of visual-TOF flame detector

A general scheme of the ‘outdoor’ visual-TOF based flame detector is shown in Fig. 3.27. The proposed algorithm consists of three stages and is similar to the ‘indoor’ TOF based flame detection algorithm described in Section 2.6.1. The first two stages, i.e., the low-cost visual flame detection and the amplitude disorder detection are processed simultaneously. The last stage, i.e., the region overlap detection, investigates the overlap between the resulting candidate flame regions of the prior stages. If they overlap, fire alarm is given.

The low-cost visual flame detector is the one which was discussed in Section 2.3, and the amplitude disorder and region overlap detection are also the same as in Section 2.6.1. The novelty of the ‘outdoor’ visual-TOF flame detector, compared to the indoor TOF flame detector, is that we now combine visual and TOF detection results and analyze them together using the region overlap detection. Important to mention is that, in order to perform the region overlap detection, the visual and amplitude images need to be registered.

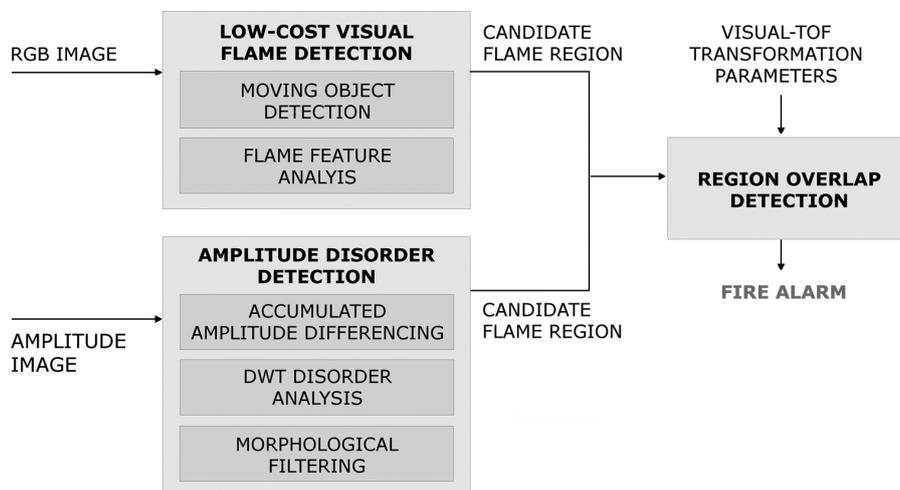


Figure 3.27: General scheme of the TOF-VISUAL based flame detector.

Some types of TOF cameras, e.g., the OptriCam [107], already offer both TOF sensing and RGB capabilities and their visual and TOF images are already registered. The majority of TOF cameras, however, still does not have this RGB capabilities. As such, a visual-TOF registration, i.e., the calculation of the transformation parameters (Section 3.3), is necessary.

3.6.2 Experimental results of TOF-visual flame detector

Analogously as for the evaluation of the indoor TOF based flame detector (Section 2.6), several realistic fire and non-fire experiments were performed to illustrate the potential use of the outdoor visual-TOF flame detector. An example of these experiments, i.e., the Christmas tree fire, is shown in Fig. 3.28. In order to test the detection range of the proposed multi-sensor detector, the distance between the sensors and the fire is also varied during the experiments.

As the results in Table 3.4 show, robust flame detection can be obtained with the proposed multi-sensor visual-TOF image processing. Compared to the VFD detection results, i.e., an average detection rate of 88% and an average false positive rate of 4%, the outdoor detector, with its average detection rate of 92% and no false positive detections, performs better. By further inspecting the results, one can also see that increasing the distance between the cameras and the fire source, does not much influence the detection results. For example, the detection rate of the outdoor wood fire test at 22 meters is around 89%, which is quasi as good as the 91% of the straw fire test at 7 meters and the 95% of

the Christmas tree fire at 10m. It seems that it is more the type of burning material that influences the detection results, i.e., some materials burn much more rapidly than others and will show the discriminative flame characteristics much faster. The results also show that, compared to the indoor detector, the average detection rate of the outdoor detector is a little lower. This can mainly be attributed to the fact that the resolution of the TOF camera, for the moment, is too low to detect small objects over long distances. Very small flames (e.g., in the beginning of the fire) are, as such, not detected.

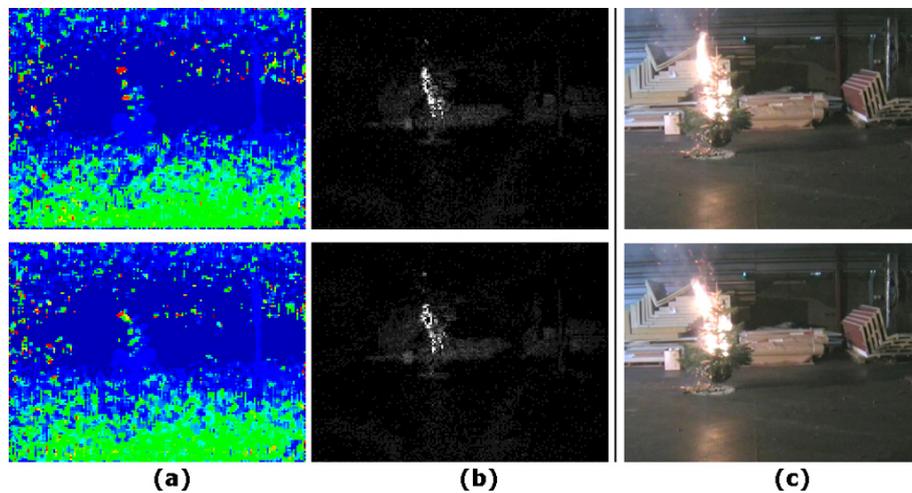


Figure 3.28: Christmas tree experiment: (a) TOF depth map and (b) corresponding amplitude image; (c) ordinary video.

Table 3.4: Performance evaluation of outdoor visual-TOF fire detection.

Video sequence (distance)	# fire frames (GT)	# detected fire frames	# false positive detections	flame detection rate
Outdoor wood fire (22m)	1230	1090	0	0.89
Christmas tree fire (10m)	815	771	0	0.95
Outdoor straw fire (7m)	460	419	0	0.91
Non-fire activities (> 10m)	0	0	0	-

3.7 Conclusions

The majority of the SOTA systems in video based fire detection are plagued by a number of difficulties in real-world scenes, e.g., lighting, reflections and noise. Many of these difficulties are caused by limitations due to the type of sensors used. Until today, most work still concentrates on systems that operate purely on visual inputs and largely ignores other sensor modalities. However, despite the progress made down this track, the goal of accurate fire detection in cluttered environments, such as car parks, remains unsolved.

To overcome the ‘visual’ limitations of current VFD systems, multiple types of sensors, like a color, thermal and/or depth camera, can be used by applying the multi-modal registration and detection techniques described in this chapter. The proposed multi-modal flame and smoke detectors take advantage of the different kinds of information represented by thermal, visual and/or depth images in order to accurately detect flames and smoke. By fusing the multi-modal modalities and using the strengths of each medium, fire detection is done more accurately and with fewer false detections. Merging information from multi-modal sensors has, as such, proven to be a win-win.

In order to combine the information in a multi-modal setup, the corresponding objects in the scene need to be registered. The proposed silhouette based registration algorithm analyses the contours and the correlation of visual, thermal and/or TOF FG silhouettes. First, the rotation is computed using silhouette contour extraction and circular cross correlation. Next, contour scaling is used to estimate the thermal-visual scale factor. Finally, the translation vector is estimated by maximizing the binary correlation. The geometric parameters found during this registration phase are further used by each of the multi-modal detectors to coarsely map the visual, thermal and/or depth images.

To evaluate our system, the proposed multi-modal flame and smoke detectors are tested on several fire and non-fire experiments and are compared to their single-sensor alternatives. We have shown objectively that each of the detectors outperform the single-sensor detectors and adhere to all the relevant requirements: object-based automatic calibration/registration, low number of false alarms, no missed detections and fast warning/alarming with different levels of detection. Due to the low-cost of the proposed techniques, such as the silhouette coverage analysis and the visual silhouette disorder analysis (which is only performed if smoke warning is given), the multi-modal detectors are also less computational expensive as many of the existing individual detectors. This makes them suitable for real-time operation, for example, in a CCTV environment.

Unfortunately, due to the unavailability of one of the cameras, we have not (yet) been able to compare the proposed multi-modal detectors, e.g., the visual-TOF and LWIR-visual flame detector, against each other. Based on their individual experiments, however, it is expected that they will give similar performance.

Future work on multi-modal fire analysis will mainly focus on investigating the benefit of using different infrared spectral cameras. Xenics [27], i.e., one of our research partners, already started research on the technical aspect in this direction [143]. For further evaluation of the work presented in this chapter, Xenics has also made arrangements within the FireSense [109] project to cooperate in their future experiments. This gives us the opportunity to test our multi-modal algorithms on a broader scale.

The author's work on multi-modal fire detection and image registration using visual, infrared and depth imaging led to the following publications.

- Steven Verstockt, Chris Poppe, Sofie Van Hoecke, Charles-Frederik J. Hollemeersch, Bart Merci, Bart Sette, Peter Lambert, and Rik Van de Walle. Silhouette-based multi-sensor smoke detection: coverage analysis of moving object silhouettes in thermal and visual registered images. *Machine Vision and Applications* journal, DOI 10.1007/s00138-011-0359-3, August 2011.
- Steven Verstockt, Charles-Frederik J. Hollemeersch, Chris Poppe, Peter Lambert, Rik Van de Walle, Sofie Van Hoecke, Bart Merci, and Bart Sette. Multi-sensor fire detection by fusing visual and LWIR flame feature. *ICGST International Journal on Graphics, and Image Processing*, 10(6): 43–50, December 2010.
- Steven Verstockt, Sofie Van Hoecke, Tarek Beji, Bart Merci, Benedict Gouverneur, A. Enis Cetin, Pieterjan De Potter, and Rik Van de Walle. A multi-modal video analysis approach for car park fire detection. *Special issue in Fire Safety Journal: Car Park Fire Safety*, submitted July 2011.
- Steven Verstockt, Sofie Van Hoecke, Pieterjan De Potter, Peter Lambert, Charles-Frederik J. Hollemeersch, Bart Merci, Bart Sette, and Rik Van de Walle. Multi-modal time-of-flight based fire detection. *Special issue in Multimedia Tools and Applications: Analysis and Retrieval of Events/Actions and Workflows in Video Streams*, submitted March 2011.

- Steven Verstockt, Pieterjan De Potter, Peter Lambert, Sofie Van Hoecke, and Rik Van de Walle. Multi-sensor fire detection using visual and time-of-flight imaging. In *Proc. 1st IEEE International Workshop on Advances in Automated Multimedia Surveillance for Public Safety, International Conference on Multimedia and Expo (ICME)*, July 2011.
- Steven Verstockt, Nele Tilley, Bart Merci, Charles-Frederik J. Hollemeersch, Bart Sette, Sofie Van Hoecke, Peter Lambert, and Rik Van de Walle. Future directions for video fire detection. In *Proc. 10th international IAFSS symposium*, June 2011.
- Steven Verstockt, Chris Poppe, Pieterjan De Potter, Sofie Van Hoecke, Charles-Frederik J. Hollemeersch, Peter Lambert, and Rik Van de Walle. Silhouette coverage analysis for multi-modal video surveillance. In *Proc. 29th Progress In Electromagnetics Research Symposium*, pages 1279–1283, March 2011.
- Benedict Gouverneur, Steven Verstockt, Guy Gielis, Stefan Nemeth, Thomas Bocquet, Jan Stynen, Jan Vermeiren The development of a multiband system for early detection of wildlife fires and indoor search and rescue operations. In *Proc. SPIE defence, security and sensing*, April 2011

Chapter 4

Multi-view fire analysis

The main focus of this chapter is on the development of a video fire analysis framework which can be used for video driven fire-spread forecasting. To the best of our knowledge, no such framework yet exists. Our main contribution consists of a virtual sensor grid based on multi-view 3D plane slicing, the use of dynamic camera maps and spatial and temporal 3D filters, which extend existing 2D concepts. Using the framework, the location of the fire, its size, its propagation and its direction can accurately be estimated. The proposed multi-view localization techniques have been tested thoroughly on fire and non-fire video sequences and have shown to work. Furthermore, preliminary experiments show the feasibility of video driven fire-spread forecasting.

4.1 Introduction

In order to actually understand and interpret the fire, detection is not enough. It is also important to have a clear understanding of the fire development and the location of the fire. Where did the fire start? What is the size of the fire? What is the direction of smoke propagation? The answer to each of these questions plays an important role in safety analysis and fire fighting/mitigation, and is essential in assessing the risk of escalation. Unfortunately, most video-based fire alarm systems still just ring the bells, i.e., they only detect the presence of fire. Even though the majority of these systems consist of several cameras monitoring the same scene, the analysis is usually carried out separately on each of the camera's sequences. However, by combining the detection results of each of these single-view cameras and analyzing them together, more accurate detection and localization of smoke and flames can be achieved and valuable fire characteristics can be detected at the early stage of the fire. These characteristics, in turn, can be used for fire-spread forecasting.

Being able to model and forecast the fire can help emergency services to work more efficiently and save lives. However, the calculations with current modeling techniques still take too long and valuable time is often lost. Using the multi-view fire analysis framework proposed in this chapter, which is able to give real-time information about the state of the environment, these zone model-based predictions of the future state can be improved and accelerated. By combining the information about the fire from models and real-time data an estimate of the fire can be produced that is better than could be obtained from using the model or the data alone. This is the final goal of video based fire forecasting, of which the proposed framework is the first part. The second part, i.e., linking the modeling and the real-time detection, is performed by our fire engineering research partners [144].

The first part of this chapter (Section 4.2) focuses on the state-of-the-art methods and tools for video fire analysis and discusses their advantages and limitations. The results of these first approaches are still limited and interpretation of the provided information is not straightforward. As such, the main goal of our work is to provide a more valuable video fire analysis tool, i.e., our novel multi-view fire analysis framework. The second part of this chapter (Section 4.3) presents a global description of the proposed framework. By fusing the low-cost video fire detection results of multiple cameras into a grid of virtual sensor points, i.e., the FireCube, valuable fire characteristics are detected at the early stage of the fire.

Next, in Section 4.4, a more detailed description is given on homography-based multi-view plane slicing, i.e., one of the main components of the framework. The proposed plane slicing technique merges the single-view detection results of the multiple cameras by homographic projection onto multiple horizontal and vertical planes, which slice the scene. Subsequently, Section 4.5 explains how a 3D grid of virtual sensor points, called the FireCube, is created at the crossings of these slices. Using this grid and subsequent spatial and temporal 3D clean-up filters (Section 4.6), information about the location of the fire, its size and its direction of propagation can be instantly extracted from the video data. This is further discussed in Section 4.7. In Section 4.8 we also briefly introduce the concept of video driven fire spread forecasting.

A subjective, i.e., visual confirmation, and objective evaluation (Section 4.9) shows that the proposed multi-view fire localization framework is able to accurately detect and localize the fire. It is found that two cameras are already sufficient to achieve a dimension accuracy of 90% and a position accuracy of 98%. By further increasing the number of cameras, it is even possible to achieve a dimension accuracy of $\pm 96\%$ and a position accuracy of $\pm 99\%$.

The experiments also shows that increasing the number of cameras has a positive effect on the detection rate, as was expected. Finally, Section 4.11 finishes this chapter and lists the conclusions.

4.2 State-of-the art in video fire analysis

Only recently, a few approaches have been proposed in literature which are capable of providing additional information on the fire circumstances, such as size and location. Yasmin [47] proposes a dynamic programming (DP) matching algorithm, which analyzes the set of contour pixels of subblocked binarized images from consecutive frames. For each of the contour pixels, the DP matching generates a displacement vector. These vectors are further analyzed by histogram analysis to obtain the orientation with the highest number of displacement vectors, i.e., the global direction of the smoke. Despite the fact that this single-view approach offers some interesting insights, the output is very limited for further analysis, since it is restricted to four directions in 2D. Furthermore, it seems impossible to perform valuable fire analysis with one single camera, since a lot of crucial information can be missed. For example, when the propagation of the flames and/or smoke is not well aligned with the camera view, analysis of growing size and propagation becomes very complicated.

The system of Martinez-de Dios et al. [145] analyzes visual and IR movies of a propagating fire front in order to supply the time evolutions of the fire front shape and position, flame inclination angle, height and base width. As secondary outputs their system also provides the fire front rate of spread (RoS) and a 3D graphical model of the fire front that can be rendered from any virtual view. The experimental setup of the system is illustrated in Fig. 4.1. The 3D model of the instantaneous fire front is constructed in real-time, based on measures of the fire front base (rear and leading edges position), and flame height and inclination angle. This graphical representation can be rendered from any point of view, simulating the image obtained by a virtual camera. Subjective evaluation of the 3D fire model viewed by camera3 (Fig. 4.1d) and the true image of camera3 (Fig. 4.1e), shows that their system achieves good performance. Although the system is directly aimed at, and is only demonstrated in laboratory experiments on a flat burn table, the authors indicate that it can also be extended to field applications [146, 147]. However, as both the laboratory and the field tests require a frontal camera view (the camera axis is perpendicular to the fire front) and a lateral camera view (the camera axis is parallel to the fire front), questions do also arise about the system's real-world applicability.

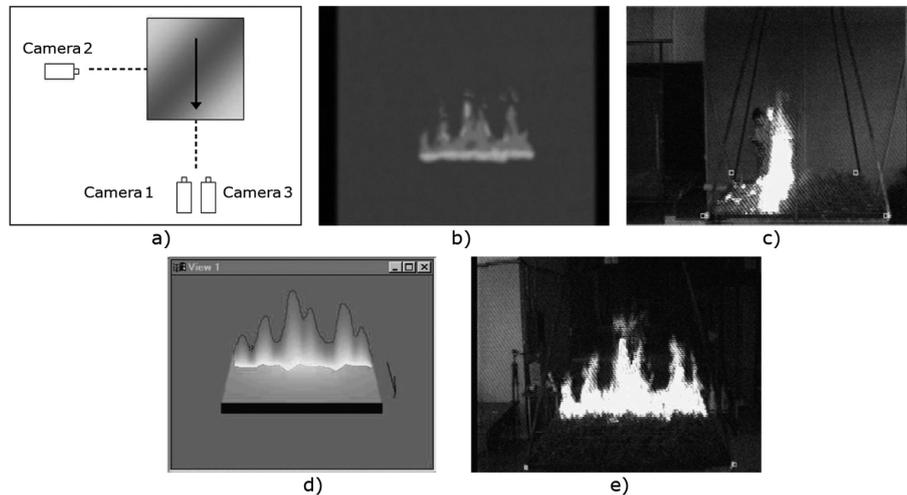


Figure 4.1: Video fire analysis system by Martinez-de Dios et al. [145]. Test with linear fire front, 30 s after ignition: a) camera configuration; b) image of Camera1 (infrared, frontal); c) image of Camera2 (visual, lateral); d) 3D fire model based on the images of Cameras1 and 2, viewed by Camera3 (virtual image); and e) true image of Camera3 (visual, frontal).

Furthermore, in order to estimate certain fire characteristics, such as the leading edge of the fire front, an observer must supply additional information about the fire ‘circumstances’ (e.g., the type of ignition), which further limits its suitability for fire analysis in real-world environments.

Similar lab experiments to those of Martinez-de Dios et al. are also discussed in [148]. In this work, Pastor et al. present a method for the fast and accurate calculation of the RoS by processing single-view infrared images. In order to calculate the RoS (\sim flame front’s position as a function of time) the correspondence between the coordinate system of the image (expressed in pixels) and the real coordinate system (expressed in meters) is needed. They propose to estimate this correspondence, i.e., the homography matrix, using the direct linear transformation (DLT) algorithm. The same technique is used by the novel video fire analysis framework proposed in this chapter. A drawback of the system of Pastor et al., from our point of view, is that it is based on an application for linear flame fronts that are generated on flat surfaces with known dimensions. Although technical guidelines are given for the extrapolation of their method to experimental scenarios on a larger scale, they discuss themselves some problems/bottlenecks related to the system’s applicability in a real-world environment. Furthermore, the assumption of a frontal view is again seen as a limitation.

Far more interesting than the prior approaches is the stereo vision work of Akhloufi and Rossi [58, 149, 150]. This approach, which is illustrated in Fig. 4.2, uses two stereo-vision cameras to track the fire spread in 3D space. First, a color based segmentation is used to extract the fire regions from both cameras. These regions are then further analyzed by feature point detection and matching between the two camera images. Feature points (also called interest points or keypoints in literature) are locations in the image where the signal changes two-dimensionally, such as corners and line intersections, as well as locations where the texture varies significantly [151]. Based on the feature point matching, 3D fire points are computed using stereo correspondence. Finally, a 3D ellipsoid is fitted for volume reconstruction and for the computation of fire characteristics such as spread dynamics, local orientation and heading direction.

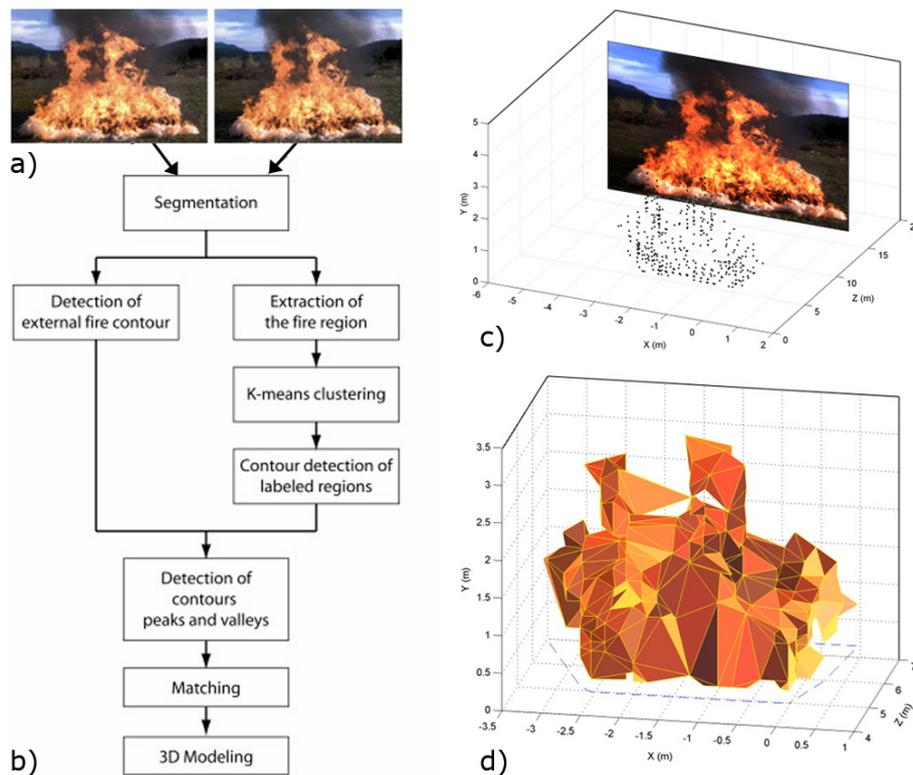


Figure 4.2: Video fire analysis system by Akhloufi and Rossi [58, 149, 150]: a) input (stereo-vision) video sequences; b) general scheme of 3D modeling framework; c) 3D position of corresponding 'feature points'; d) 3D surface reconstruction of the fire.

Although it is already possible to extract some valuable fire development information by means of this stereo-vision based technique, questions arise about its applicability in real-time scenarios without a priori knowledge of the fire. Furthermore, we believe that a grid-based approach is more appropriate than the ellipsoid modeling technique for fire development analysis, since interpretation and temporal analysis of the latter is not straightforward.

Despite the limited results of the discussed video fire analysis approaches, the results from existing ordinary multi-view object analysis approaches, such as the people and vehicle trackers in [152–154], are already very promising and their basics are also appropriate for video fire analysis. The majority of these works relies on homographic projection [23] of camera views, which also forms the basis of our framework.

4.3 Global description of the framework

Single-view VFD algorithms, such as the ones discussed in Chapter 2, are able to accurately detect flames and smoke. However, due to visibility problems, such as occlusion, and due to limitations in $2D \rightarrow 3D$ reconstruction, crucial information on the location, size, and propagation of the fire is hard to retrieve. On the other hand, this information is of great importance for a better understanding of the fire. In order to retrieve these valuable fire characteristics, a new multi-view localization framework is proposed that detects the 3D position and volume of the fire in an accurate manner. Fig. 4.3 presents the global architecture of this framework.

Using the proposed localization framework, information about the fire location and (growing) size can be generated very accurately and quickly. First, the framework detects the fire, i.e., smoke or flames, in each single view. In order to do this, the framework uses the low-cost flame and smoke detectors which are discussed in Chapter 2. As an alternative, an appropriate single-view smoke or flame detector can also be chosen out of the numerous approaches already proposed in literature [1]. Secondly, the single-view detection results of the available cameras are projected by homography [23] onto horizontal and vertical planes which slice the scene. For optimal performance it is assumed that the camera views overlap, such that each position is seen by at least two cameras. Overlapping multi-camera views provide elements of redundancy, i.e., each point is seen by multiple cameras, this way helping to minimize ambiguities like occlusions and improving the accuracy in the determination of the position and size of the flames and smoke. Next, the multi-view plane slicing algorithm averages the multi-view detection results in each of the hori-

zontal and vertical planes. This step is a 3D extension of Arsic's multiple plane homography [155]. Then, a 3D grid of virtual multi-camera sensors, i.e., the FireCube, is created at the crossings of these planes.

At each sensor point of the 3D FireCube, the detection results of the horizontal and vertical planes that cross in that point are analyzed and only the points with stable detections are further considered as candidate fire or smoke. Finally, 3D spatial and temporal filters clean up the grid and remove the remaining noise. The filtered grid can then be used to extract the smoke and fire location, information about the growing process and the direction of propagation. The novel aspects in the proposed framework are the 3D grid analysis and the spatial and temporal 3D filtering, which extends existing 2D filter concepts. In the next sections, the major components, that make up the framework, are described more in detail.

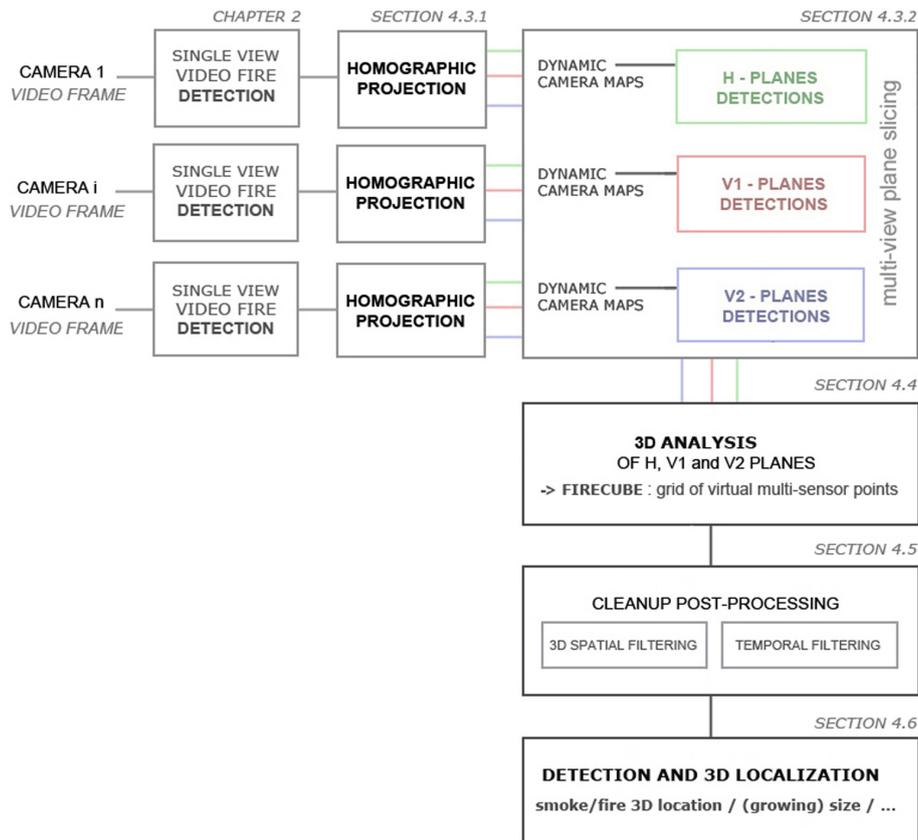


Figure 4.3: Global architecture of multi-view localization framework for 3D smoke and flame analysis.

4.4 Homography-based multi-view plane slicing

4.4.1 Homographic projections

The core of our multi-view localization framework is the planar homography constraint [152–154], which is translated into the 3D structure of the FireCube, as explained further. The constraint combines foreground likelihood information, i.e., the probability of a pixel in the image belonging to the foreground (FG), from different views to determine the locations of the FG objects on a plane in a reference coordinate system. The constraint implies that only pixels corresponding to the plane locations will consistently warp, under homographies of the reference plane, to the same FG location in the common reference view. Points that do not meet this assumption are mapped, i.e., are projected, to a skewed location on the reference plane.

Fig. 4.4, in which one person is viewed in one position by three cameras, illustrates the concept of the homography constraint on a ground plane. As can be seen, the parts of the person’s body (white FG pixels) that touch the ground are consistently projected to the ground plane location of the person in the common reference view. Other parts, for example the person’s legs, are projected to skewed locations and do not overlap. This can best be seen in the common reference view, which combines the projections of the three camera views to the reference ground plane. It must be pointed out that the homography constraint is not limited to the real ground plane. The constraint also applies to the homographic projection of virtual planes parallel and orthogonal to the real ground plane, i.e., the planes that form the basis of the FireCube. The creation of these planes is based on basic geometry and is fairly straightforward. Before going into details on this, the basics of homographic projection, which forms the basis of our multi-view plane slicing, are discussed.

To project the FG views, i.e., the detected/moving parts of the single-view camera VFD images, to a common plane, the homography matrix Hom of each camera is needed:

$$Hom = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix}. \quad (4.1)$$

The Hom matrix maps the plane in the camera view to the plane in the common view by transforming the coordinates of the points using the projection parameters h_{ij} ($i, j = 1..3$). The matrix can be calculated offline using techniques such as the 4-point-based Direct Linear Transform (DLT) [23] or even self-calibration methods can be used [156]. The proposed framework uses the former method.

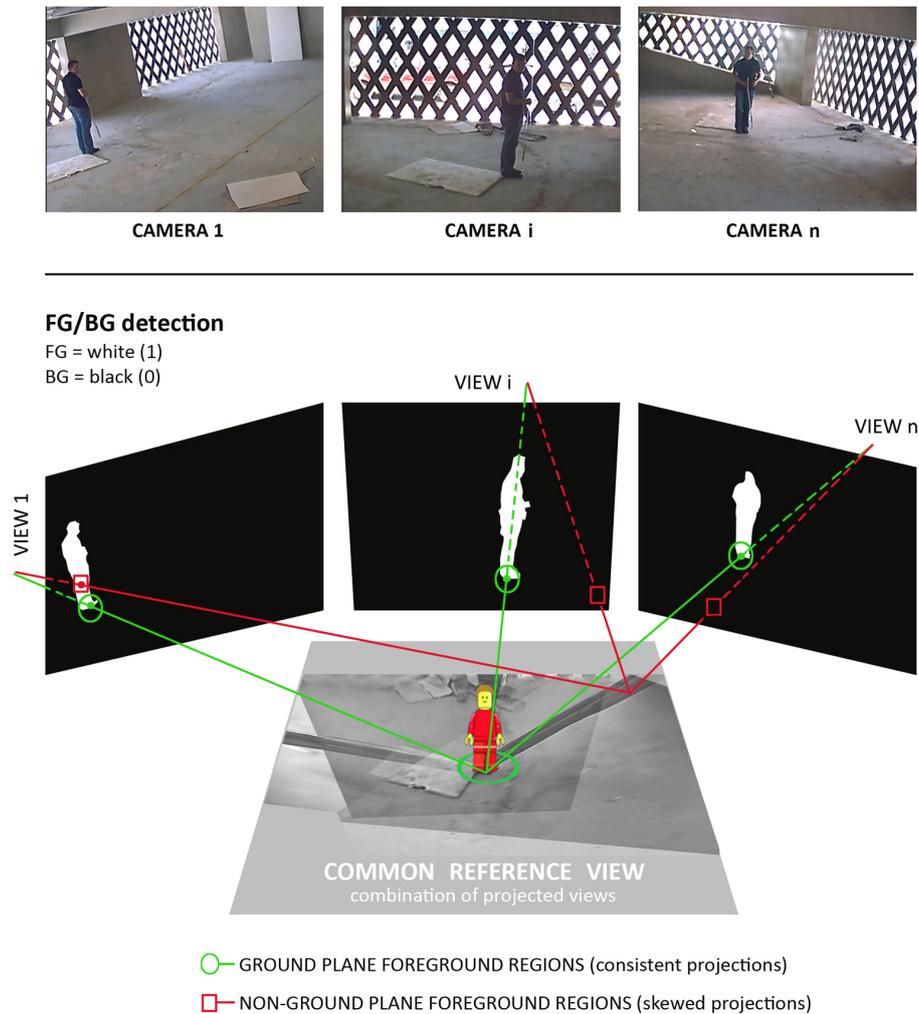


Figure 4.4: Homography constraint.

During the camera calibration a set of four 2D to 2D correspondences $x_i \leftrightarrow x'_i$ between the views and the ground plane are selected for each camera and the homography matrix Hom is calculated by DLT, in the same way as is explained in the work of Hartley and Zisserman [23]. An example of the camera calibration for one of our cameras is shown in Fig. 4.5. A person is placed at four predefined positions, i.e. a square of one by one meter, and the coordinates x_i of his feet, i.e., the lowest pixels, are used as calibration points.

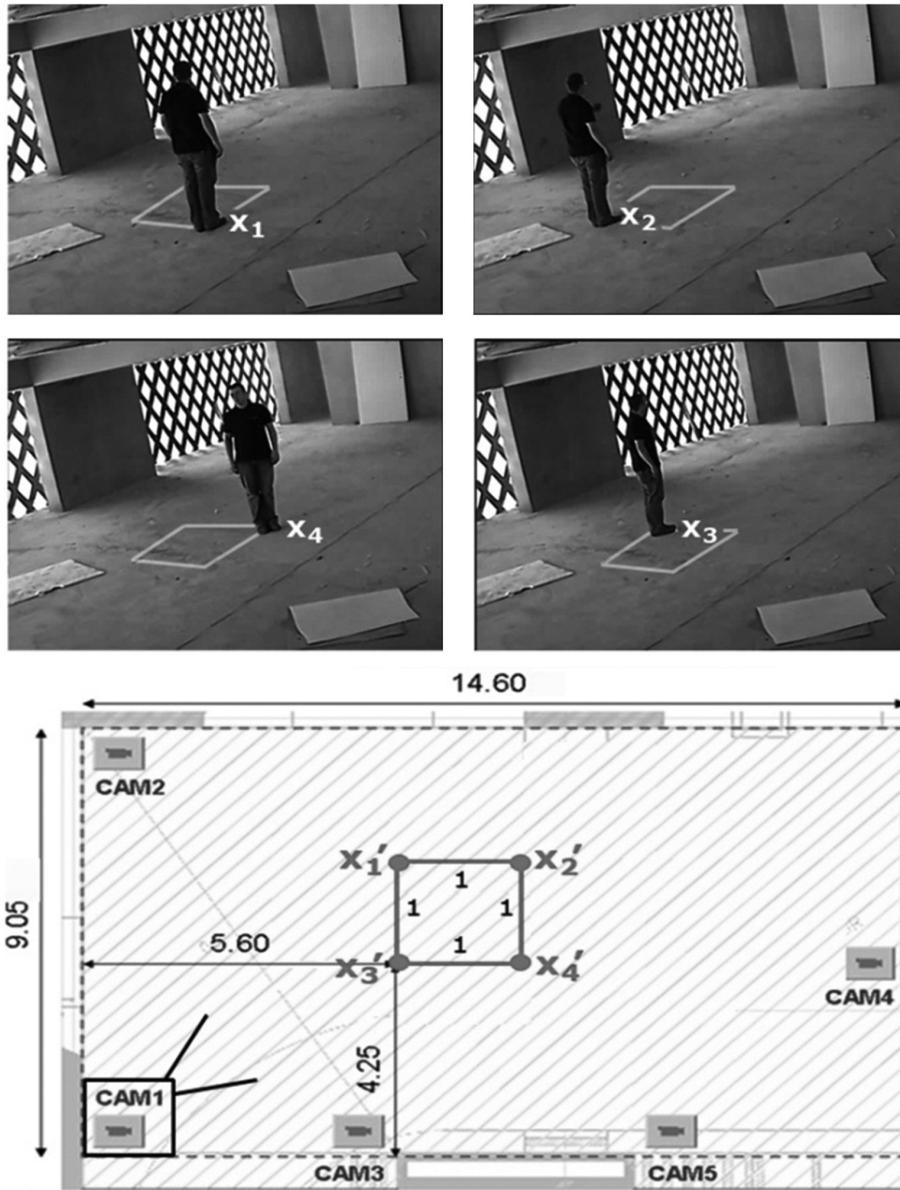


Figure 4.5: 4-point-based DLT homography estimation of camera 1. A person is placed at four predefined positions, i.e. a square of one by one meter, and the coordinates x_i of his feet are used as calibration points.

In order to map the flame and smoke risk values of each point $[x, y]$ onto the point $[x', y']$ in the common reference plane, the homographic projection of single-view VFD results uses the *Hom*-matrix of the selected view (Eq. 4.1) and the transformations given by:

$$\begin{aligned} x' &= \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}} \\ y' &= \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}}. \end{aligned} \quad (4.2)$$

In order to know the overall flame and smoke risk value of $[x', y']$, the average risk value of all the mappings on $[x', y']$ is taken.

4.4.2 Multi-view plane slicing

By detecting the presence of high flame and smoke risk values on different virtual planes orthogonal and parallel to the ground plane, the precise 3D locations of the fire can be retrieved. In order to do this, the homography matrices of these virtual planes need to be known. Selecting calibration points for each of these planes is too time-consuming and error-prone, so that a technique is needed to automatically generate the homography of these planes. From the few horizontal multiple plane strategies that have recently been proposed in literature, the computationally low complex multilayer homography by Arsic et al. [155] is the most interesting. Other techniques, such as the plane slicing of Khan and Shah [157] and Lai and Yalmaz [158], can also be used, but since their calibration and height and distance estimation require more computational work, the prior one is chosen and extended to 3D in this work.

As in [155], the computational effort for creating the multiple plane homographies is kept at a minimum. Starting from the eight calibration points shown in Fig. 4.6, the homography is computed for the six reference planes connecting these points, i.e., two horizontal and four vertical planes. All other homographies, for planes parallel to the calibration planes, are computed by basic geometry. For example, the homography of a horizontal plane parallel to the ground plane at height z is estimated using:

$$\begin{aligned} p_{x,z} &= (p_{x,1} - p_{x,2})z/z_2 + p_{x,1} \\ p_{y,z} &= (p_{y,1} - p_{y,2})z/z_2 + p_{y,1}, \end{aligned} \quad (4.3)$$

where $p_{x,1}$ and $p_{y,1}$, and, $p_{x,2}$ and $p_{y,2}$, respectively are the coordinates of the calibration points p_1 and p_2 in the planes at height z_1 and z_2 , i.e., the ground plane and the head plane respectively. In our experiments the calibration points were selected by detecting the lowest and highest pixels of a person's feet and head at some predefined positions, i.c. a square of one by one meter, as is illustrated in Fig. 4.5.

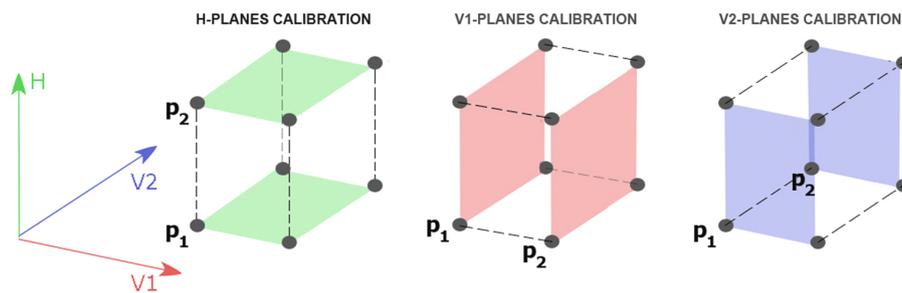


Figure 4.6: Multi-plane camera calibration (2x 4-point-based DLT homography estimation). Starting from eight calibration points, the homography is computed for the six reference planes connecting these points.

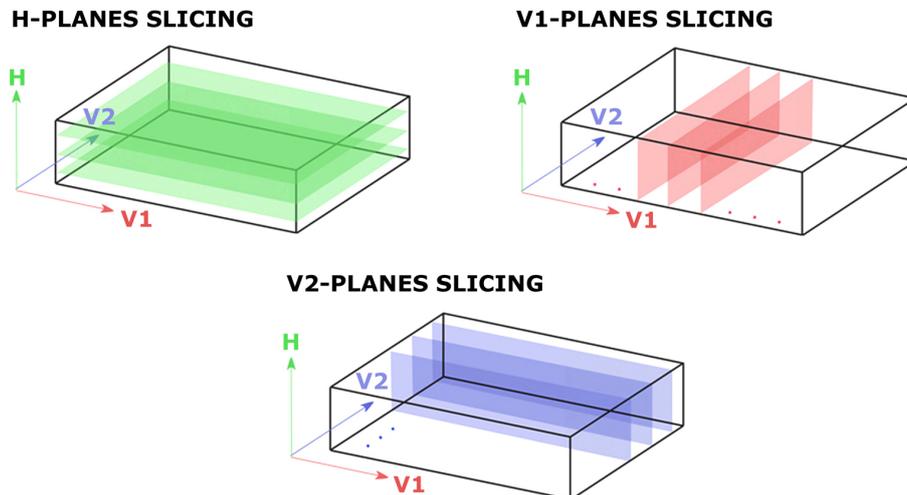


Figure 4.7: Plane slicing in horizontal and vertical directions. Each of the 'virtual' cuts represents an averaged detection plane, i.e., a combination of the projected multi-view detection results to that plane.

Similar to the calculation of the homography of a horizontal plane parallel to the ground plane, the multi-view localization algorithm calculates the homography for any vertical plane orthogonal to the ground plane. As soon as the homography of each view to the common plane is known, the overall flame and smoke risk value on that plane is calculated by averaging the projected risk values.

As is illustrated in Fig. 4.7, the plane slicing virtually cuts the scene in the horizontal and vertical directions. Each of these cuts represents an averaged detection plane, i.e., a combination of the projected multi-view detection results to that plane. By observing these detection planes, it is already possible to obtain an estimate for the fire/smoke location and size, but in order to ensure accurate localization and to automatically provide easy-interpretable data for motion analysis, further processing of these plane detections is needed. For this reason, the 3D FireCube is created, as described in Section 4.5. Before going into detail on the FireCube, however, it is important to discuss the use of dynamic camera maps, which facilitate the analysis of plane slices.

4.4.3 Dynamic camera maps

Since fire/smoke regions are not always visible in all views, or only partly visible, the detection/localization will be influenced by the number of cameras having the fire/smoke position under surveillance. By making a camera map that contains the number of cameras that are able to monitor the specific position, appropriate detection criteria can be determined. Fig. 4.8 shows the creation of one of the camera maps that are used in our experiments.

The higher the value of a position on the dynamic camera map, the more cameras that monitor that position. Special events, such as camera unavailability and tampering [159], can be automatically detected and could also be taken into account by dynamically updating the map. For example, when an object blocks one of the cameras' field of view or one of the cameras is broken/switched of, the binary mapping of this camera view can easily be subtracted from the dynamic camera map and the camera will (temporally) not participate in the video fire analysis.

A dynamic camera map can also be very useful during installation of a multi-view camera system. Using the map, the optimal system configuration, i.e., number of cameras, placement, etc., can easily be found. Furthermore, they can also be a great tool for surveillance operators, for example, in order to know which region is mapped by which camera.

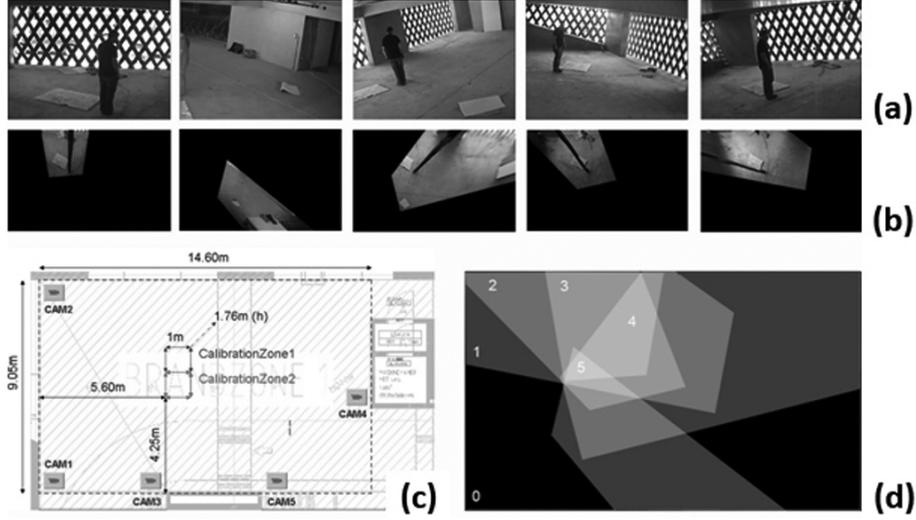


Figure 4.8: Camera mappings to (real) ground plane: (a) input video sequences; (b) homography mappings to ground plane; (c) real ground plane; (d) BW-map of number of views for each location by combination of mappings from (b).

4.5 Grid analysis

At the crossings of the horizontal planes $Plane_H$ and vertical planes $Plane_{V1}$ and $Plane_{V2}$, a grid, i.e., the FireCube (Fig. 4.9), is formed. The proposed FireCube consists of virtual 3D sensor points (x, y, z) at which the detection results, i.e., the corresponding smoke and flame risk value P_{smoke} and P_{flames} , of $Plane_H$, $Plane_{V1}$ and $Plane_{V2}$ are analyzed using:

$$FireCube[x, y, z] = \begin{cases} FG, & \text{if } \overline{P_{flames}[x', y', z']} \geq t_{flames} \\ & \text{or } \overline{P_{smoke}[x', y', z']} \geq t_{smoke} \\ BG, & \text{otherwise.} \end{cases} \quad (4.4)$$

Sensor points for which the average of its smoke or flame risk values is higher than the t_{smoke} or t_{flames} threshold are labeled as FG , i.e., fire/smoke. Based on our experiments, a t_{smoke} of 0.6 and a t_{flames} of 0.7 are found the best values.

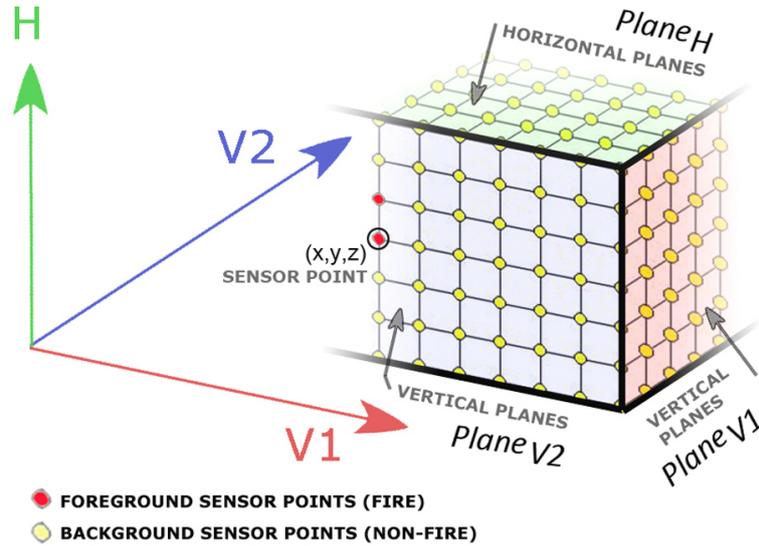


Figure 4.9: Grid accumulation of horizontal and vertical planes.

The number of horizontal and vertical planes depends on the desired localization accuracy. This accuracy is related to the distances between the planes and the dimension of the scene. The higher the number of planes, the more accurate the localization will be, but the more the computational cost will increase. However, the cost of increasing the number of planes is limited, since the calibration does not change and the homography matrices for the new planes can be calculated off-line. The only additional cost at run-time will be the projection of the detection results to the new planes (which possibly can be accelerated with additional hardware). In order to reduce this trade-off between spatial accuracy and execution time, one may also think of enlarging the temporal window between consecutive detections. However, the influence of such frame skipping enhancement on the overall performance of the FireCube is not investigated, but can be part of future work.

4.6 Clean-up filtering

For higher detection robustness, the localization framework also includes a clean-up filtering step in the spatial and temporal domain. In the spatial domain, it filters out noisy FG points in the FireCube using a set of weighted 3D median filters and fills up holes in FG object regions using a 3D filling operator. The temporal filtering on its turn removes FG object regions which

do no overlap with one of the detected regions in the previous or subsequent FireCubes. This implies that a FG object must be detected in at least two FireCubes of consecutive frames (in time) to be further considered as FG smoke or fire.

4.6.1 Spatial filtering

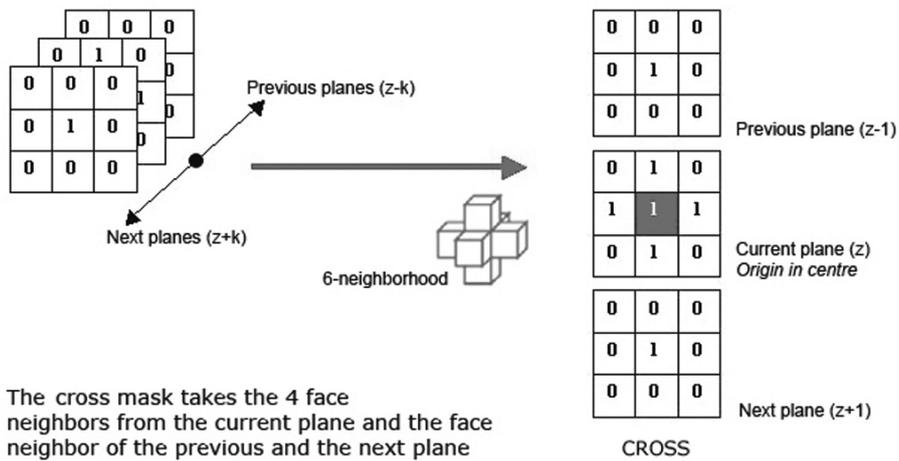
Spatial 3D filters have already successfully been applied in many biomedical applications. For example, in [160] the 3D morphological opening filter is used as a kind of neighborhood operator to enhance the spatial data in medical images. Another example is the 3D median filter, which is used to denoise magnetic resonance images [161]. Although these filters perform well in the referred work, they place too high constraints on the smoke points in the large spaced FireCube's grid. For this reason, a novel, less strict, filter is developed.

The set of weighted 3D-median filters collects the binary results of six different weighted median filters and selects the median value, i.e., the middle value, in the sorted results. The 'local' filters consist of a sliding 3-by-3-by-3 binary mask, centered at a central point. The mask is a binary matrix with ones on positions where points must be taken into account and zeros on positions where the value of the points does not affect the filter result. In order to be labeled as candidate smoke, more than half of the considered points should be non-zero. The six filters respectively use a horizontal, two vertical, a diagonal, a crossed and a singular mask. Fig. 4.10 gives an overview of these masks.

The combination of weighted 3D-median filters reduces the effect of outlying data points, i.e., noise, and preserves the edges at the object region boundaries. A clarifying example of how the weighted 3D-median filters work is given in Fig. 4.11. As this example illustrates, only smoke points on the grid which pass at least three of the six weighted median filters remain smoke; other points are removed. It is also important to mention that mirrored boundary pixels are used to facilitate the filter operation at the FireCube boundaries. This is a common mechanism for boundary processing in many filter operations.

The median filters denoise the FireCube, but do not resolve the problem of possibly large holes in the remaining object regions, which can occur due to similarity with the background or due to too large homogeneous regions in the moving object. Since the dimension of the median filters is 3-by-3-by-3, holes with a larger dimension will not be filtered out completely. In order to accurately estimate the size of the objects and to analyze the objects' motion, these holes must, however, be filled. This is the task of the 3D filling operator.

The 3D filling operator is a global filter which operates on the entire Fire-Cube at once. The filling operator first pushes the BG points, touching the boundary of the FireCube, onto a stack of BG boundary points. Next, the 3D filling iteratively removes a BG point from the stack and pushes its 6-face BG neighbors, which are not yet investigated, onto the stack. When there are no BG points left on the stack, the removal stops. The remaining BG points, which have not been pushed onto the stack, are labeled as holes and filled up in the FireCube, by relabeling them as FG. An example of this hole filling procedure is shown in Fig. 4.12.



The cross mask takes the 4 face neighbors from the current plane and the face neighbor of the previous and the next plane

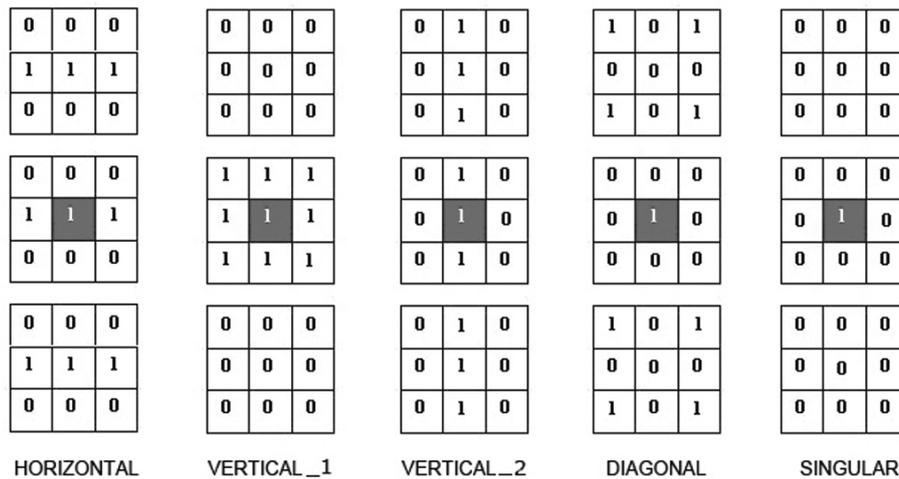


Figure 4.10: Masked windows for weighted 3D-median filtering.

One could think that instead of filtering in 3D, the spatial clean-up post-processing could be done on a 2D plane-slice level. However, 2D filtering operations have a critical drawback when used in a 3D context. A single grid point, existing on a slice in 2D, might be interpreted as noise, while in 3D this grid point might be a tip of a fire/smoke object, not to be eliminated. Hence, the use of a 3D noise removal filter and a 3D filling operator provides a better technique for more reliable localization. As will be shown in Section 4.9, experimental results confirm this hypothesis. By using 2D filtering operators the dimension accuracy decreases with almost 3% and the position accuracy with 2%.

4.6.2 Temporal filtering

After filtering out noisy points and filling up holes in the spatial domain, a 3D temporal filtering is applied on consecutive FireCube grids to remove temporal misdetections. This last clean-up operator filters out object regions that only appear once in one of the consecutive grids $FireCube_{t-1}$, $FireCube_t$, and $FireCube_{t+1}$. Since the 3D spatial structure of the object region may vary over consecutive frames, the detection does not search for a perfect match. In fact, it checks if the 3D bounding box BB_m of the m^{th} object region in $FireCube_t$ overlaps with one of the bounding boxes BB of the regions in the previous $FireCube_{t-1}$ and the following $FireCube_{t+1}$ (Eq. 4.5). If no overlap exists, the filter relabels the region as BG :

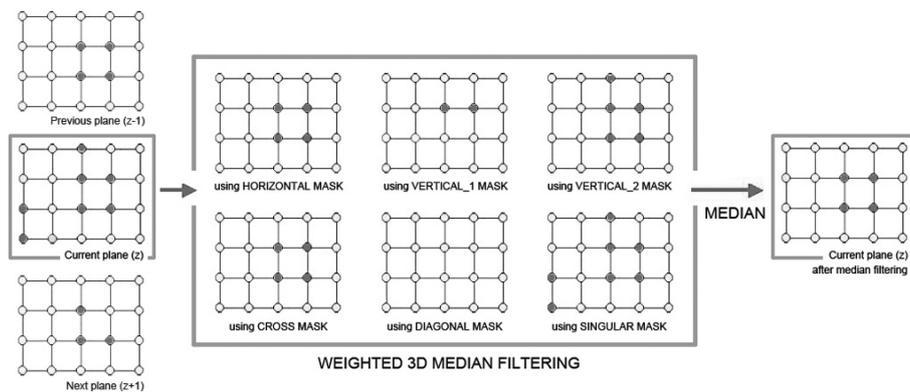


Figure 4.11: Example of 'local' weighted 3D-median filtering.

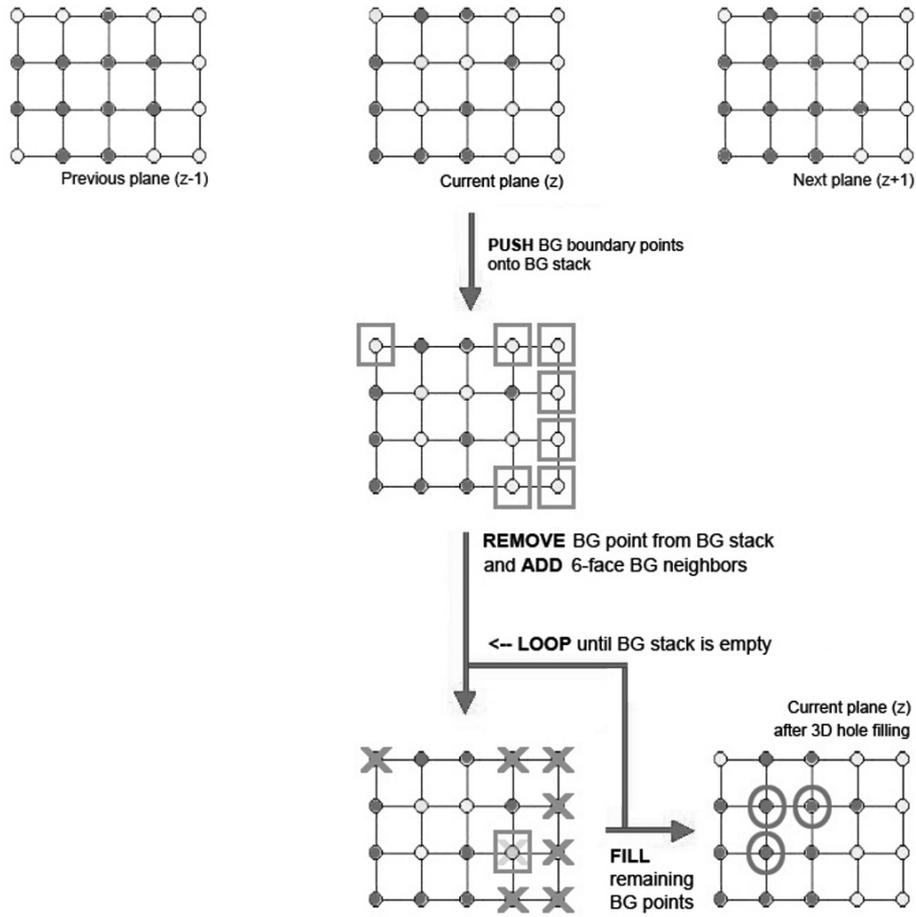


Figure 4.12: ‘Global’ 3D hole filling example.

$$\text{FireCube}_t[BB_m] = \begin{cases} FG, & \text{if } \text{FireCube}_t[BB_m] \cap \text{FireCube}_{t-1}[BB_i] \neq \emptyset \\ & \text{or } \text{FireCube}_t[BB_m] \cap \text{FireCube}_{t+1}[BB_i] \neq \emptyset \\ & \forall i \neq m \\ BG, & \text{otherwise.} \end{cases} \quad (4.5)$$

As the example in Fig. 4.13 shows, temporal misdetections are easily removed using this 3D overlap filter. At the end only temporal stable fire/smoke objects remain.

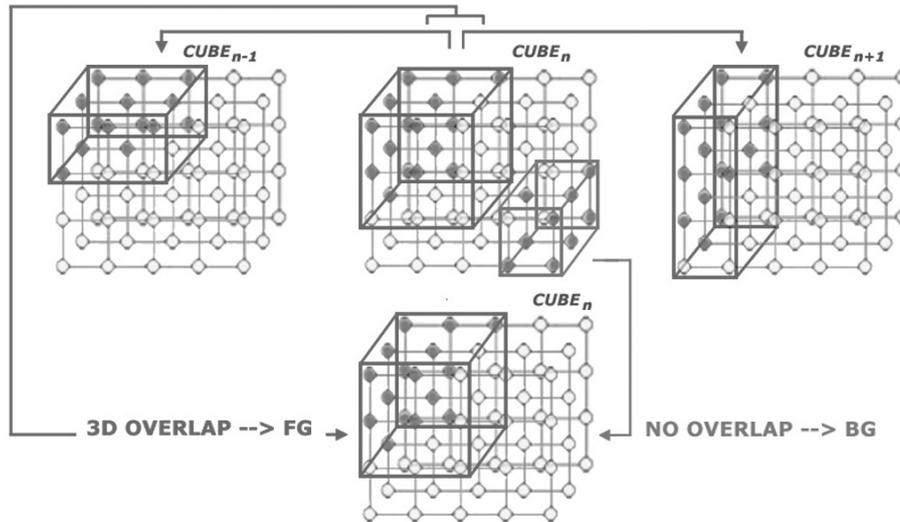


Figure 4.13: Example of temporal 3D overlap filter.

The proposed fusion methodology of the FireCube and its subsequent clean-up steps eliminate the majority of the errors created by misclassification during single-view detections. By combining the detections of multiple views and analyzing them in multiple planes in horizontal and vertical directions, a more accurate detection and localization is achieved. Further analysis makes it possible to produce highly valuable 3D fire development information, as is discussed in the next section.

4.7 Fire and smoke development analysis

The last phase in the proposed localization framework allows generating fire and smoke development information, i.e., easily interpretable flame and smoke characteristics that are of great value in evaluating the fire, such as location, volume, height, orientation and growing size.

4.7.1 Retrieval of fire and smoke characteristics

Different approaches can be used to determine the fire and smoke characteristics. First of all, for the simplicity of the model, one could think of estimating them using the 3D bounding box of the FG region, which is the smallest rectangular box containing all the FG points. Using this bounding box, the volume of the fire region can be equated to the volume of the box, and the position can be set equal to the centroid of the box. Further, the height of the box can be used as

smoke layer height estimation, one of the most important determinants [162] in evaluating the fire risk. Although this bounding box approach seems appealing at first sight, the box can become too large for particular shapes of fire regions, and as such, the characteristics will incorrectly describe the region.

A more effective approach is to describe the fire regions using the 3D convex hull, i.e., the smallest convex polyhedron that contains all the fire points, which has more degrees of freedom and delivers a more accurate estimation. Several algorithms have been proposed in literature to find the convex hull of a set of 3D points, such as the gift wrap [163], the divide and conquer [163], or the quick hull method [164]. Each of these methods are able to accurately describe convex 3D point clouds. However, their computational cost is very high and there are situations in which the convex hull will be too large to describe the fire region. For this reason, another technique is proposed, namely an adaptive bounding box algorithm.

The proposed adaptive bounding box algorithm splits up the bounding box of the fire/smoke region into 2 parts: the fire/smoke plume box and the smoke layer box (Fig. 4.14). This strategy of splitting up the fire region in multiple zones is also used in many other fire-related applications, e.g., fire zone modeling [165]. The splitting starts with a horizontal shrinking of the original bounding box. As long as more than ten percent of the box is filled with BG points, the box is iteratively reduced in the upside direction. The residual box of this reduction process forms the smoke layer box. Subsequently, the FG points that have been removed during the shrinking are merged into a new box, i.e., the fire/smoke plume box. In order to detect multiple fire sources, the proposed adaptive bounding box algorithm needs to be changed slightly. For example, connected component labeling can be performed on the resulting FG blobs, after smoke layer extraction. Each of the resulting blobs can then be merged in a separate plume box.

The smoke layer box and the plume box are very useful instruments for further analysis of the fire. First of all, the spatial characteristics of both boxes, such as the volume, the height and the central position, closely approximate valuable real-world fire characteristics. For example, the centroid of the floor plane of the plume box is a good indication for the location of the fire seat. In addition, temporal analysis of these characteristics produces valuable fire development information, such as growing size and propagation direction. An approximation of the growing size can be retrieved by temporal analysis of the ratios between the current volume of the box and its volume in the previous video frames. An approximated propagation direction can be calculated using a 3D variant of motion history images [166].

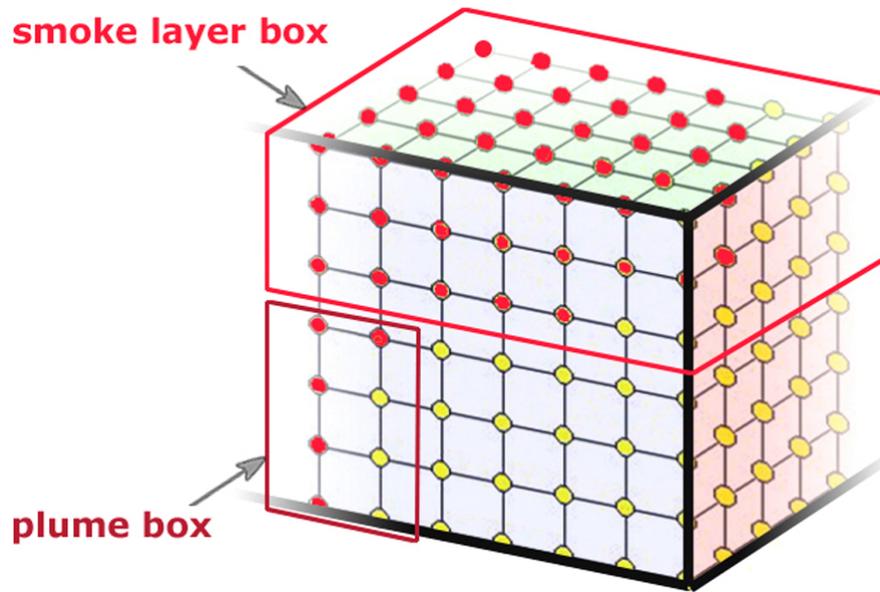


Figure 4.14: Adaptive bounding box strategy for splitting up fire region in plume and smoke layer box.

4.8 Video driven fire spread forecasting

Fire spread forecasting is about predicting the further evolution of a fire, in the event of the fire itself. In the world of fire research, research on this topic is only just starting to emerge [167]. Based on their common use in fire modeling, computational fluid dynamics (CFD) calculations [168] look interesting for fire forecasting at first sight. These are three-dimensional simulations where the rooms of interest are subdivided into a large amount of small cells (Fig. 4.15a). In each cell, the basic laws of fluid dynamics and thermodynamics (conservation of mass, total momentum and energy) are evaluated in time. These types of calculations result in quite accurate and detailed results, but they are costly, especially in calculation time. As such, CFD simulations do not seem to be the most suitable technique for fast fire forecasting. Zone models [168], on the other hand, seem more appropriate to perform this task.

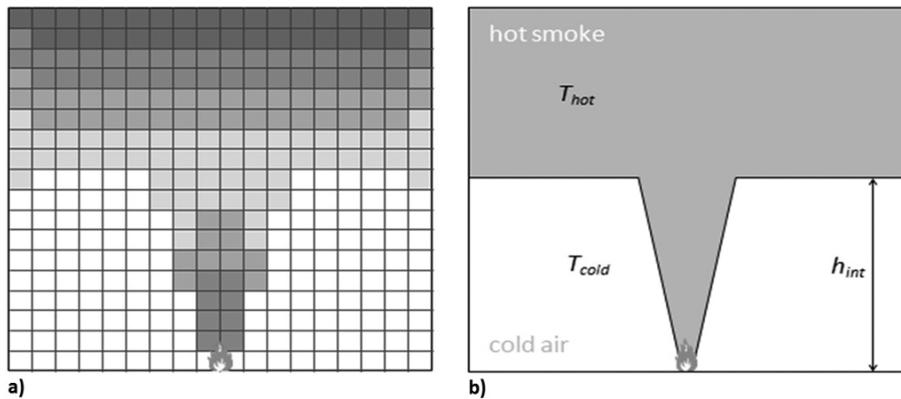


Figure 4.15: Fire modeling techniques: a) Computational Fluid Dynamics (CFD) model; b) zone model.

In a zone model, the environment is subdivided into two main zones. The smoke of the fire is in the hot zone. A cold air layer exists underneath this hot zone (Fig. 4.15b). The interface between these two zones is an essentially horizontal surface. The height of the interface (h_{int}) and the temperature of the hot (T_{hot}) and cold (T_{cold}) zones vary as function of time. These calculations are simple in nature. They rely on a set of experimentally derived equations for fire and smoke plumes. It usually takes between seconds and minutes to perform this kind of calculations, depending on the simulated time and the dimensions of the room or building. Therefore, it is much better suited for fire forecasting than the use of CFD calculations [169].

The real aim of our fire forecasting research is to use measured data from the fire, e.g., obtained by sensors or video images in the room of interest, in order to replace or correct the model predictions [170, 171]. This process of data assimilation is illustrated in Fig. 4.16, which summarizes our future plans for video driven fire forecasting [144]. As can be seen in the graph, the model predictions of the smoke layer depth (\sim zone model interface h_{int}) are corrected at each correction point. This correction uses the measured smoke characteristics from our fire analysis framework. The further in time, the closer the model begins to match the future measurements and the more accurate predictions of future smoke layer height become. A preliminary experiment on video driven fire forecasting is discussed hereafter. Real-time smoke height and fire size measurements of a sofa fire are compared to zone-model measurements in order to guide and accelerate the modeling.

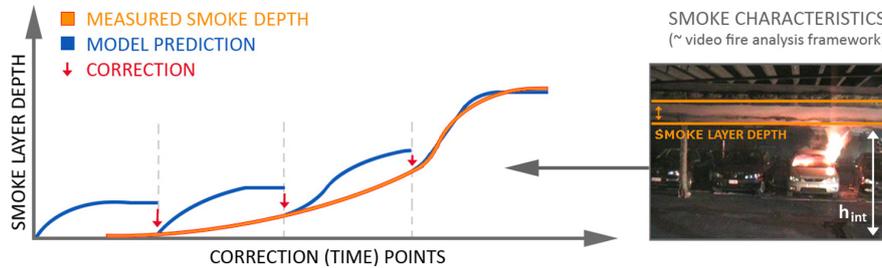


Figure 4.16: Data assimilation: video driven fire forecasting (\sim FireGrid [170]). The FireCube gives information about the state of the fire, and using this information, zone model-based predictions of the future state can be improved and accelerated.

The proposed video driven fire forecasting is a prime example of how video-based detectors will be able to do more than just generate alarms. Detectors can give information about the state of the environment, and using this information, zone model-based predictions of the future state can be improved and accelerated. By combining the information about the fire from models and real-time data an estimate of the fire can be produced that is better than could be obtained from using the model or the data alone. Important to note is that not only flame and smoke information is useful to efficiently forecast and fight the fire, but also other information about the monitored scene can be of high importance. For example, a broken window can influence the fire growth. Most of this data can also be delivered by an intelligent video surveillance system. However, this is out of the scope of this dissertation.

4.8.1 Experimental setup

As a preliminary experiment on video driven fire-spread forecasting, we have performed an ISO 9705 room-corner test [172] in which a single seat sofa is put on fire. The setup of this test is shown in Fig. 4.17. The proposed experiment is based on a recently developed data assimilation methodology [171, 173], which main objective is to assist emergency response (in case of a fire) based on real time information. The under-lying idea is similar to Numerical Weather Predictions (NWP), be it on different time and length scales. In our work at hand, the methodology is extended to an ISO-room fire with open door. The data used for assimilation is provided by a video camera. In the future, our fire engineering research partners will also use thermocouple data to measure the average upper layer temperature. The video camera was placed at the door level to estimate the smoke layer depth and the fire size, i.e., the width and the height of the flames. This real-time information is compared to the zone-model measurements in order to guide and accelerate the modeling of the sofa fire.

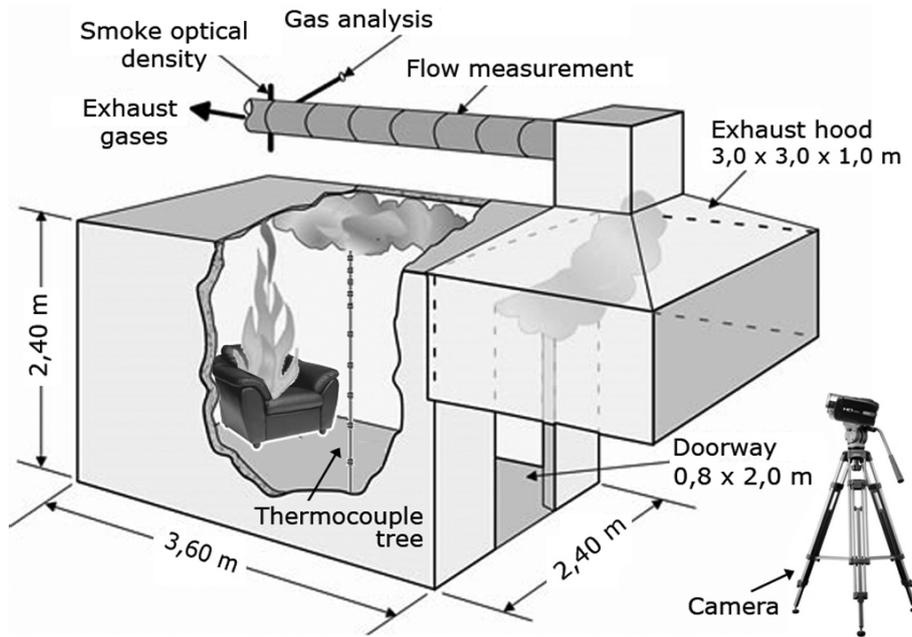


Figure 4.17: ISO room-corner test: sofa fire with open door. A video camera was placed at the door level to estimate the smoke layer height. Thermocouples can be placed inside the room to measure the average upper layer temperature. For the moment, however, thermocouple data is not used.

The two-zone model was used as the assimilating model for our video driven fire forecasting, because of its simplicity and efficiency in the calculation of tenability conditions in the early stages of a compartment fire. More information on this model is already given in Section 4.8 and a detailed description on how the forecasting (\sim real-time evolution of the fire) is performed using this model is given in [144]. In short, Fig. 4.18 summarizes the concept of video driven fire forecasting. Further on, we will only focus on how the estimation of the smoke layer height and the fire size is performed, i.e., the first part of the video driven fire forecasting (\sim sensor readings), and how close these estimations match the zone model results.

4.8.2 Single view video fire analysis

Due to the lack of multi-view camera images, the estimation of the smoke layer height and the fire size was performed using a novel single view video fire analysis technique. Instead of using the proposed multi-view framework,

a commonly used technique for the determination of the smoke layer interface height [174], which relies on the second derivative of the temperature profile, was translated into a novel video analysis approach. To the best of our knowledge, this is the first time that a technique from the ‘fire world’ is applied to video analysis.

A general scheme of the proposed algorithm is shown in Fig. 4.19. At start-up of the system, multiple lines with high energy are automatically selected in the video images. These lines show high similarity with the thermocouples that are used for temperature profile analysis. In our experiment, four different ‘energy lines’ are selected. The energy values E^{line_N} of these lines are calculated using the same discrete wavelet transform (DWT) based function as in Section 2.3.2. Also here, the energy is evaluated blockwise dividing E^{line_N} in blocks $[x, y]$ of arbitrary size, and summing up the square contribution of each high-frequency, i.e., high detail, wavelet subimage within each block. In our set-up, a block-size of 8-by-8 pixels was experimentally found the most appropriate value. Depending on the environmental characteristics, e.g., the size of the room and the camera position, a larger or smaller block-size can be needed. However, no method yet exists for the appropriate block-size estimation. Further experiments will investigate if model-based ‘validations’ can help in finding such an (automatic) estimator. Subsequently, the energy profile EP^{line_N} is constructed by blockwise normalization of all the energy values in the energy line. Some exemplary energy profiles are shown in Fig. 4.20. It can be noticed that these energy profiles show high similarity with temperature profiles that are used in fire engineering.

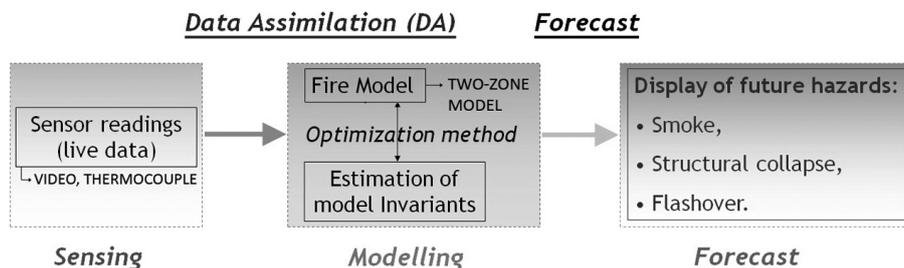


Figure 4.18: General scheme of fire forecasting concept. The methodology consists mainly of assimilating real-time data collected on the fire. The main parameters of this (assimilated) model are then estimated, and a real-time evolution of the fire (i.e., forecast) is displayed.

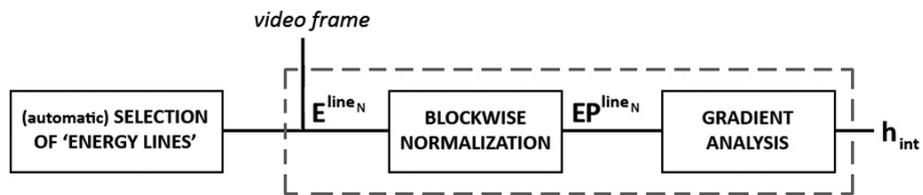


Figure 4.19: General scheme of single view video fire analysis algorithm for smoke layer height (h_{int}) estimation.

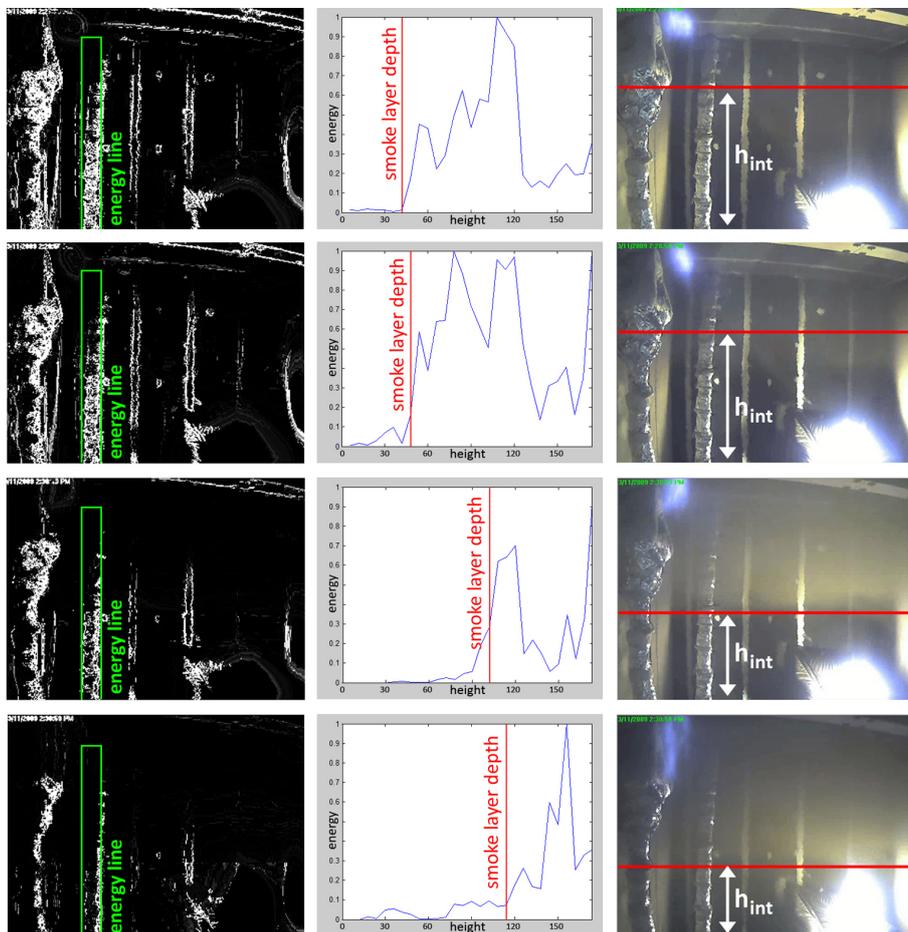


Figure 4.20: Exemplary smoke height estimations of 'burning sofa test': energy line (left), energy profile (middle) and input frame + detected h_{int} (right).

In order to find the smoke layer depth, the gradient ∇EP^{line_N} of the energy profile is analyzed. The first point at which the gradient shows a high increase, i.e., exceeds the experimentally found threshold $t_{\nabla}(= 0.2)$, is labeled as the smoke layer depth. Further (model-based) sensitivity analysis is needed to verify if this is the optimal threshold. By subtracting this smoke layer depth from the height of the room, the smoke layer interface h_{int} can easily be retrieved. In Fig. 4.20, some exemplary smoke layer height estimations of our ‘burning sofa test’ are shown. For each of the input video frames, one of the energy lines and its energy profile is given. Furthermore, the detected smoke layer height corresponding with this energy line is visualized in the video frame. A subjective, i.e., visual, evaluation of these exemplary frames already shows that the proposed approach is effective in detecting the smoke layer height. However, further objective evaluation is needed to confirm this.

For the detection of the fire size, i.e., the flame width and height, we use the visual flame detector proposed in Section 2.3.1. Based on the detected flame width W_f and height/length L_f , an indirect estimation Q_c of the convective heat release rate (HRR) of the fire is calculated. In fire engineering this is known as the Heskestads correlation [175]:

$$Q_c = \left(\frac{L_f + 1.02W_f}{0.235} \right)^{5/2}. \quad (4.6)$$

4.9 Experimental results

In order to test the localization and the detection performance of the multi-view localization framework, several fire experiments were performed in a car park. In the first experiment, the proposed framework detects valuable characteristics of smoke generated by a smoking machine. In this experiment, only visual validation of the detection results is performed, because real ground truth, i.e., the exact position of the smoke location, is (quasi) impossible to retrieve. In the second experiment, the location of a pool fire is detected, using the framework, and compared to the real positions and dimension of the pool fire in the car-park, i.e., the exact ground truth. Pool fires consist of a liquid fuel with a horizontal surface of a specified area. They are widely used for testing in fire engineering research. The third experiment investigates the influence of the number of cameras on the detection performance of the proposed and the investigated state-of-the-art detectors.

All the experimental sequences were acquired by Linksys WVC2300 cameras, which support 640x480 MPEG-4 video format at 30 frames per second. Single-view fire detection results, i.e., the input for the localization framework, were retrieved by using the flames and smoke detector proposed in Section 2.3. Since the framework is independent of the type of VFD, also other detectors [1] can be used here. As such, it is even possible to integrate other types of sensors, such as the IR and TOF based fire detectors proposed in this dissertation. It must also be pointed out here that, in order to reduce the computational cost, a lot of work is done offline, before the detection is switched on. For example, the calibration and the multi-plane homography estimation are some of the time-critical parts which are performed during startup of the system. As such, only the projections and FireCube analysis is performed at runtime, making the framework suitable for real-time analysis.

4.9.1 Experiment 1: smoking machine test (subjective evaluation)

In this experiment, the proposed algorithm detects the location, the growing size and the propagation direction of smoke generated by a smoking machine. Exemplary shots of these experiments are shown in Fig. 4.21, where the upper (a-c) and the lower images (d-f) are three different camera views of the test sequences at frame 4740 and 5040 respectively.

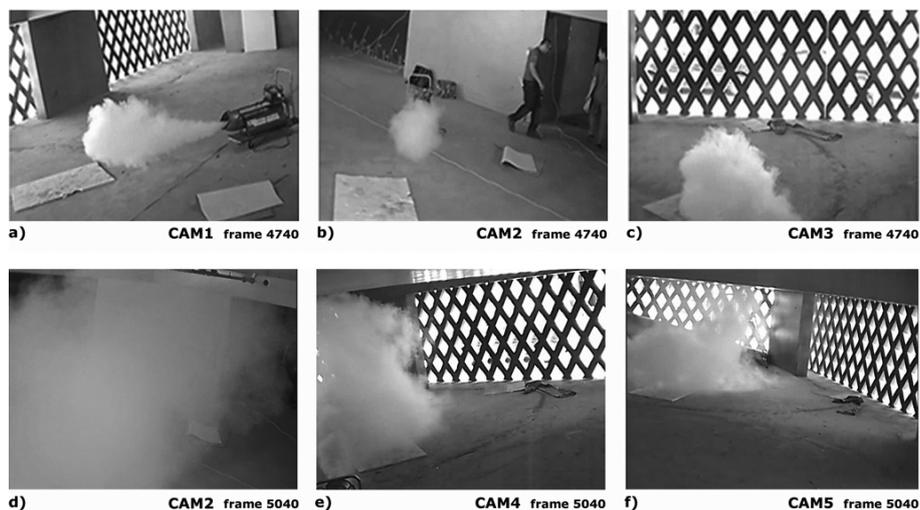


Figure 4.21: Smoking machine experiments.

As can be seen in the 3D model in Fig. 4.22, and the back-projections of the 3D results in Fig. 4.23, the framework is able to detect the location and the dimension of the smoke regions. In Fig. 4.22 the smoke regions are represented by the dark gray 3D boxes, which are bounded by the minimal and maximal horizontal and vertical FG slices. As a reference, also the bounding box of the smoking machine is visualized in the 3D model.

Even if a camera view is partially or fully occluded by smoke, like for example in frame 5040 of CAM2 (Fig. 4.21d), the framework localizes the smoke region, as long as it is visible from the other views. Based on the detected 3D smoke boxes, the framework generates the spatial smoke characteristics, i.e., the height, width, length, centroid position, and volume of the smoke region. These characteristics are also shown in Fig. 4.22. By analyzing this information over time, the growing size and the propagation direction can be estimated. Both can be very helpful in fighting the fire.

The back-projections (Fig. 4.23) of the 3D smoke regions to the camera views, using inverse homography, show that the multi-view slicing approach produces plausible and acceptable results. Due to the fact that no 3D ground truth data of this smoke test is available, only this kind of visual validation is possible. However, based on this visual validation it is already possible to state that, contrary to existing fire analysis approaches [13] which deliver a rather limited 3D reconstruction, the FireCube outputs valuable 3D information about the fire development. To further improve the evaluation process, also comparison with fire zone modeling [27, 30] can be investigated, as is done in Section 4.9.4.

4.9.2 Experiment 2: pool fire test (objective evaluation)

Contrarily to the evaluation method in the first experiment, which is mainly a subjective (visual) validation, this second experiment objectively evaluates the localization performance of the proposed framework. In order to do this, the location of a pool fire is detected, using the framework, and compared to the real positions and dimension of the pool fire in the car-park, i.e., the exact ground truth.

Fig. 4.24 shows some exemplary images from the pool fire test. These images are taken by four different cameras, placed at each side of the car park. The dimension of the pool fire is 2.00 by 4.00 meters and its position, i.e., the centroid, is at (14.00m, 15.00m). Since the exact position and the dimension of the pool fire is known, the output of the framework can easily be evaluated against these values. Table 4.1 and Table 4.2 list the results of this evaluation for increasing number of cameras.

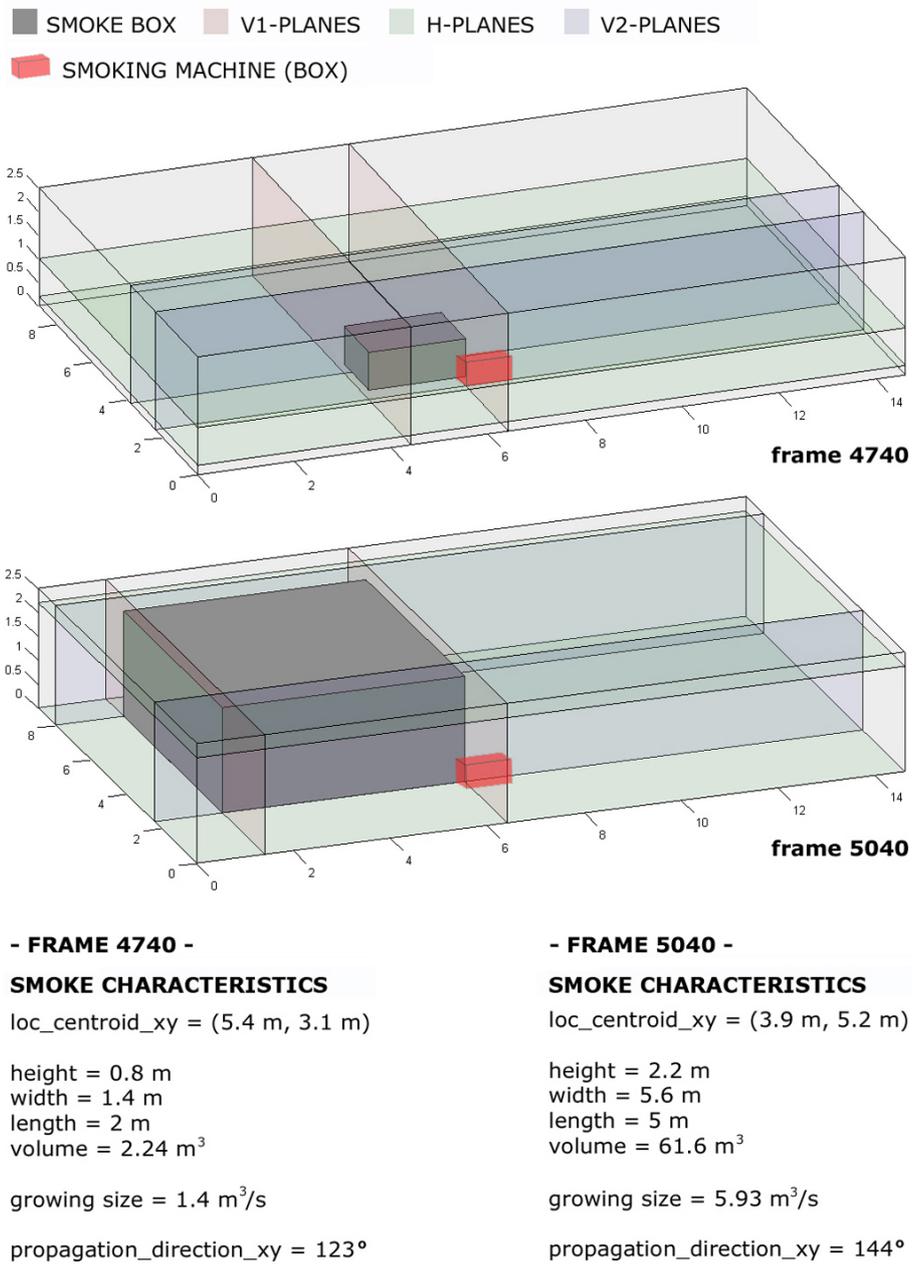


Figure 4.22: Plane slicing-based smoke box localization.

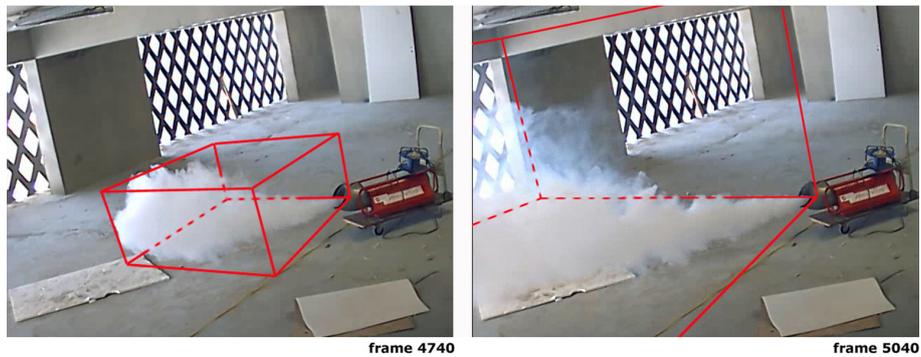


Figure 4.23: Back-projection of 3D smoke box results into camera view CAM1.



Figure 4.24: Pool fire experiments.

The results in Table 4.1 and Table 4.2 show that two cameras are already sufficient to achieve a dimension accuracy of 90% and a position accuracy of 98%. By further increasing the number of cameras it is even possible to achieve a dimension accuracy of 96% and a position accuracy of 99%. As such, it is shown that the proposed framework is able to accurately detect the position and the dimension of the flames, which are two valuable fire characteristics.

Table 4.1: Dimension accuracy of the proposed localization framework for increasing number of cameras (pool fire test).

# cameras	dimension		dimension accuracy	
	<i>GT: 2.00m by 4.00m</i>		3D filtering	2D filtering
	width	height		
1	1.67m	3.49m	0.85	0.81
2	1.88m	3.42m	0.90	0.88
3	1.92m	3.84m	0.96	0.93
4	1.93m	3.87m	0.96	0.94

Table 4.2: Position accuracy of the proposed localization framework for increasing number of cameras (pool fire test).

# cameras	position		position accuracy	
	<i>GT: (14.00m,15.00m)</i>		3D filtering	2D filtering
	x	y		
1	13.48m	15.64m	0.98	0.95
2	13.82m	15.79m	0.98	0.95
3	13.80m	15.38m	0.99	0.97
4	13.85m	15.41m	0.99	0.98

Important to mention is that adding a fifth camera will not necessarily lead to 100% dimension/position accuracy, since there is no kind of linear relationship between the number of cameras and the accuracy. In most cases it will of course improve, but how much is difficult to say. This, for example, will depend on the camera positions and its detection accuracy.

In order to show the effectiveness of the proposed 3D clean-up filtering (Section 4.6), Table 4.1 and Table 4.2 also contain the results of a 2D filter variant of the 3D median and hole filling filters. These results show that the use of a 3D noise removal filter and a 3D filling operator provides a better and more reliable technique for localization. By using the 3D filtering operators, the dimension accuracy increases on average with almost 3% and the position accuracy with 2%.

4.9.3 Experiment 3: influence of the number of cameras on the detection performance (objective evaluation)

Experiment 2 revealed that the number of cameras has a positive influence on the localization performance of the framework, especially on the dimension accuracy. This experiment investigates if the number of cameras causes a similar effect on the detection performance of the proposed and investigated state-of-the-art detectors. The detection results for increasing number of cameras are listed in Table 4.3. It is important to mention that the combination of the detection results of multiple cameras is performed by using the FireCube.

As the results in Table 4.3 show, the average gain of using two cameras instead of one is only 1%. However, the gain of using four cameras instead of one is already 3%. As such, this experiment shows that the number of cameras has a positive effect on the detection rate, not only for our approach but also for other state-of-the-art detectors. For more dynamic scenes, with a higher number of occlusions, or for other types of fire, e.g., in which the flames are more occluded by smoke, the gain of using multiple cameras is even expected to be bigger.

4.9.4 Experiment 4: video driven fire-spread forecasting

A temporal plot of all the smoke layer height detection results of the ‘burning sofa test’ is shown in Fig. 4.25. For each of the four (automatically detected) energy lines, the evolution of the detected smoke layer height is presented. As these results show, the evolution is quasi similar for all the energy lines. Only near the end of the test, small differences can be noticed between Line4 and the other energy lines. Most probably, these artifacts arise due to the presence of high-energy flames within the low-energy smoke region of Line4. In order to cope with this problem, the smoke layer height is averaged over all the energy lines. It is also expected that in environments with a more energy-rich background, less such artifacts will occur.

Comparing the mean smoke layer height to the ‘expected’ height from the zone model reveals, i.e., objectively shows, that the proposed method performs well in estimating the smoke layer height. This is shown in Fig. 4.26. Only at the very beginning of the sofa fire, the video and the modeling data show some significant differences. The reason for these differences is that the modeling assumed the sofa started directly to burn. Afterwards, the video recordings revealed that the fire was started with a wooden crib and that the sofa was not really burning in the early beginning of the video sequences. As this was assumed in the model, this resulted in a video/modeling mismatch.

Table 4.3: Influence of the number of cameras on the detection performance of the proposed smoke and flame detection method (Method 1), the combined method based on the flame detector by Celik et al. [96] and the smoke detector described in Toreyin et al. [36] (Method 2), and combination of the feature-based flame detection method by Borges et al. [45] and the smoke detection method by Xiong et al. [40] (Method 3).

video sequence (# frames)	# fire frames <i>ground truth</i>	# detected fire frames			# false positive frames			detection rate ¹		
		method			method			method		
		1	2	3	1	2	3	1	2	3
Smoking machine (1733)	923	834	654	789	9	34	52	0.89	0.67	0.80
	2 CAMERAS	843	668	781	12	23	48	0.90	0.70	0.79
	3 CAMERAS	858	681	795	10	20	51	0.92	0.72	0.81
	4 CAMERAS	860	707	806	6	21	54	0.93	0.74	0.81
Pool fire (2260)	1844	1665	1634	1618	0	0	0	0.90	0.89	0.88
	2 CAMERAS	1671	1674	1635	4	0	7	0.90	0.91	0.88
	3 CAMERAS	1702	1662	1640	0	0	3	0.92	0.90	0.89
	4 CAMERAS	1702	1674	1651	0	0	0	0.92	0.91	0.90

* $detection\ rate = (\#\ detected\ fire\ frames - \#\ false\ alarms) / \#\ fire\ frames$

Based on future experiments, planned in 2012, further (model-based) sensitivity analysis on the t_{∇} energy threshold will be performed to optimize the ‘fitting’ of the video and the modeling data.

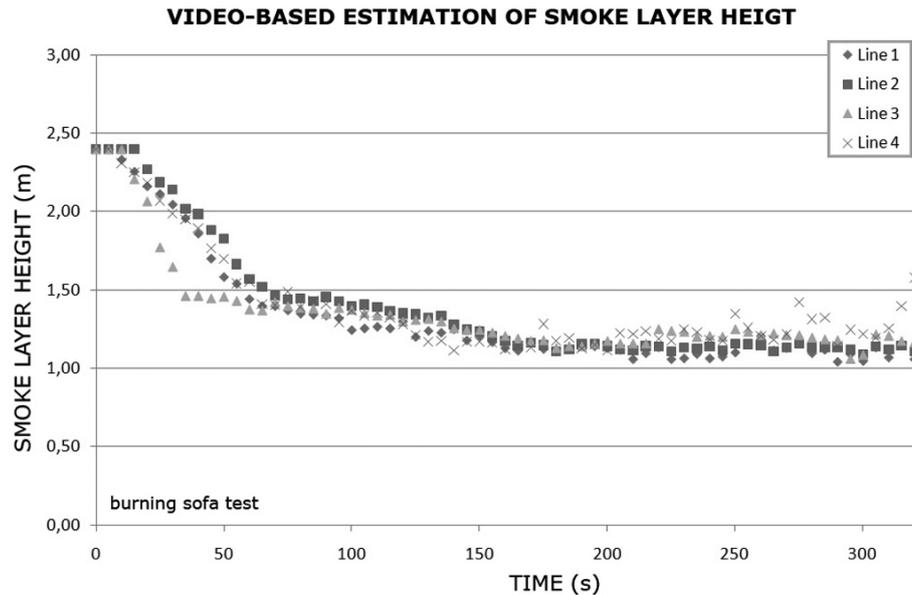


Figure 4.25: Video-based estimation of smoke layer height for 4 energy lines.

Besides ‘forecasting’ the smoke layer height, estimating the fire size is also an essential aspect of Fire Hazard Analysis (FHA). As discussed in Section 4.8.2, the fire size (\sim HRR) can be estimated directly from the flame width and height using the Heskestads correlation (Eq. 4.6). A temporal plot of the HRR of the ‘burning sofa test’ is shown in Fig. 4.27. Again, comparison with the ‘expected’ HRR of the zone model reveals that the proposed method performs well in estimating the fire size. After a relatively short growth period, the measured HRR is constant, around 98 kW. Therefore, in a first forecast, a constant fire size can be assumed. Starting from 200s the fire size is no longer constant and starts growing. Therefore, for example, the data collected between 200s and 260s (\sim assimilation window) can be taken into account to calculate the fire growth factor, as is explained in the ongoing work of our research on video driven fire-spread forecasting [144, 169, 176]. Other fire characteristics, such as a forecast of the temperature, can also be estimated using the techniques described in these works.

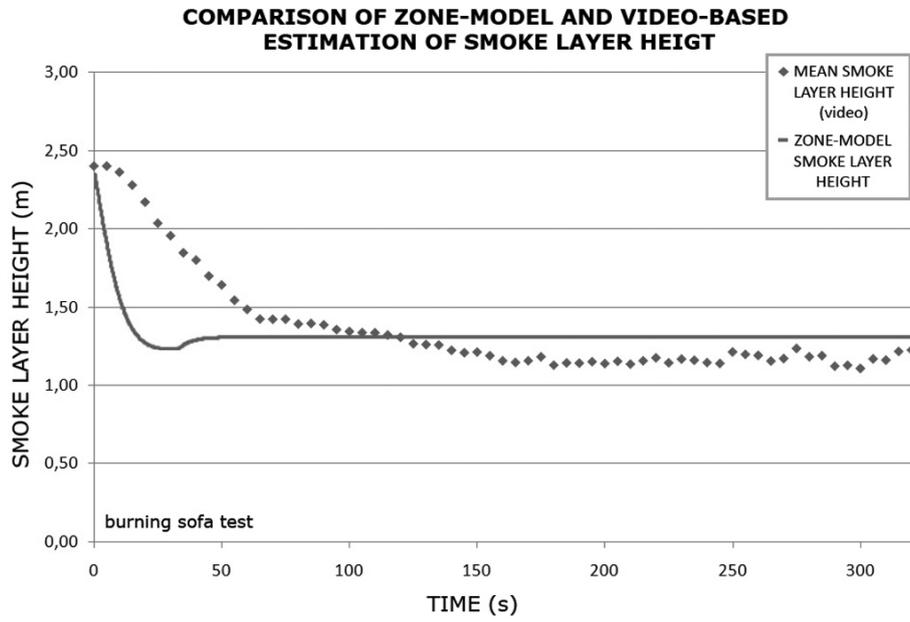


Figure 4.26: Comparison of zone-model and video-based estimation of h_{int} .

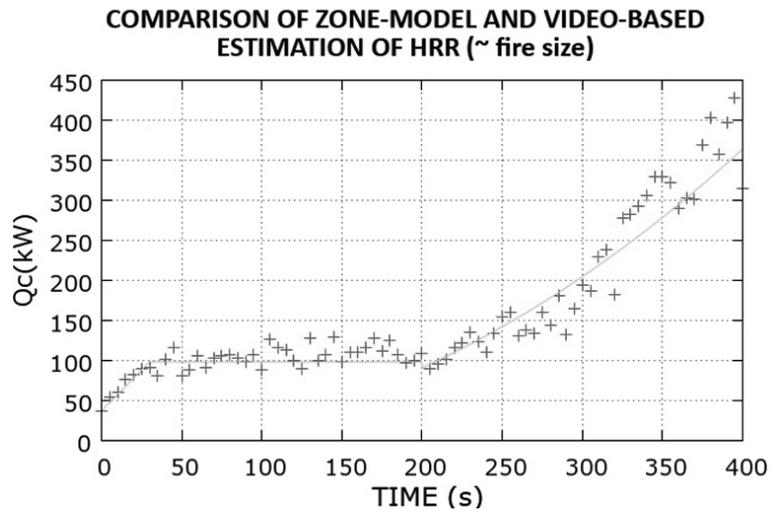


Figure 4.27: Comparison of zone-model and video-based estimation of HRR.

4.10 General remarks and future improvements

In order to further improve the proposed fire analysis framework, future work will mainly focus on two topics: the use of sensor feedback/fine-tuning and the investigation/evaluation of combining video-based sensors with conventional/traditional fire detectors. Related to this, some general remarks about our intelligent fire detection system (proposed in Fig. 1.1) are also discussed in the following two subsections.

4.10.1 Sensor feedback and sensor fine-tuning

Within our framework, the sensors have no direct effect on each other. Each sensor stands on its own, delivering video, thermal, depth or amplitude information without taking into account feedback from the other neighboring sensors. Possibly, the framework can be further enhanced if the available information and intelligence of all sensors can be fed back to the sensors to optimize their functionality. Sensor feedback can, for example, be used to improve the BG modeling within each of the video detectors. Currently, their BG model is updated solely on their own detection information. Furthermore, detection criteria (thresholds) can be linked on other sensors 'knowledge' about the environment. In recent experiments on video-driven fire forecasting, we have used thermocouples for sensor feedback, in order to evaluate the video-based smoke layer height.

Related to sensor feedback, the framework can also use environmental characteristics (such as geo-data and meteorological info) to select optimal sensor parameters. Off-site and on-site metadata, i.e., side info about the environment, can be gathered/evaluated and used for real-time fine-tuning of the sensor data. Both ideas, i.e., 'sensor feedback' and 'context-based sensor fine-tuning' are currently included in an EU-FP7 proposal on mega-fires in which we participate for the sensor WP.

4.10.2 Combination with conventional/traditional sensors

One can also think of combining information from, on one hand, video sensors and, on the other hand, conventional/traditional non-video fire sensors to further improve the reliability of the framework or to reduce its financial and/or computational cost. However, the added value of 'cheaper' point sensors will be low, due to the several limitations (Section 1.2.2) related to these type of detectors (especially when they are used in large open spaces). Combining volume sensors, like the Open-Area Smoke Imaging Detection (OSID) beam detector of Xtralis [177], with video looks the most interesting.

OSID can help, for example, in reducing steam/dust false alarms, which are still difficult to detect with multi-modal video sensors. Similarly, video can help to solve water/rain problems of OSID. This combination of OSID beam detectors and multi-modal video will be (further) investigated within the context of the car park fire project [28] of Prof. Bart Merci. Furthermore, within the EU-FP7 project on mega-fires we also plan to further focus on this.

4.11 Conclusions

The detection of valuable fire characteristics, e.g., fire size, smoke propagation and smoke layer height, is an important but difficult problem in fire analysis. Most fire alarm systems only detect the presence of fire and are not able to model fire evolution. Only recently a few video based approaches have been proposed which are capable of providing additional information on the fire circumstances. The results of these single-view fire analysis approaches, however, are still limited and interpretation of the provided information is not straightforward. In order to provide a more valuable video fire analysis tool, a multi-view fire analysis framework has been proposed in this chapter. The novel framework fuses fire detection results of multiple single-view cameras, providing gains over ‘classical’ single-view fire analysis techniques.

The proposed 3D fire localization framework relies on 3D extensions to homographic plane slicing. The framework merges single view VFD results of multiple cameras by homographic projection onto multiple horizontal and vertical planes which slice the scene under surveillance. At the crossings of these slices, a 3D grid of virtual sensor points is created. At each of these points, the detection results of the crossing planes are accumulated and compared to a dynamic camera map, which is also one of the novel aspects in our work. Using this grid and subsequent spatial and temporal 3D clean-up filters, information about 3D location, size, and propagation of the fire can be extracted.

The novel 3D grid, the use of dynamic camera maps, and the spatial and temporal 3D filters, which extend existing 2D concepts, provide accurate localization and enable more reliable fire analysis. Objective and subjective experimental results, in which the flames and smoke development in a car park is analyzed, confirm these findings and indicate that the proposed multi-view fire localization framework is able to accurately detect and localize the fire. Two cameras are already sufficient to achieve a dimension accuracy of 90% and a position accuracy of 98%. By further increasing the number of cameras, it is

even possible to achieve a dimension accuracy of 96% and a position accuracy of 99%. Furthermore, the experiments show that increasing the number of cameras to monitor the scene has a positive effect on the detection rate. For example, the gain of using four cameras instead of one is 3%.

The valorization potential of the proposed multi-view localization framework is very high, as the (generic) framework can easily be adapted to other applications. In order to analyze the behavior/evolution of other object types, only the single-view detectors need to be changed. For example, in [178] the generic framework is used to extract the body model of a person. Related to fire fighting, the framework could for example also be used to localize and rescue people in a burning building.

On top of the multi-view localization framework, this chapter also presented how the output from our video fire analysis techniques can be used for video driven fire forecasting. Although this is ongoing research, first experiments already show good results. The experiments revealed that the evolution of the detected smoke layer height closely matches the expected evolution from the computational zone model. Furthermore, the video based detection of the flame height and width allowed us to estimate the convective heat release rate (HRR) and to detect a change in the trend of the fire. Again, comparison with the expected HRR of the zone model revealed that the proposed method performs well in estimating the fire size and the fire trend.

Finally, it is important to mention that the proposed single-view algorithm for the detection of the smoke layer height is mainly based on a fire engineering technique. To the best of our knowledge, this is the first time that a technique from the fire world is applied to video analysis. Based on the ‘positive’ result of this first multidisciplinary interchange of techniques, we still see plenty of room for further research on the interplay between both worlds. In the near future, for example, experiments are planned in which fire modeling data will be used to optimize/fine tune the video-based estimation of fire characteristics in our fire analysis framework.

The author’s work on multi-view fire analysis and video driven fire spread forecasting led to the following publications.

- Steven Verstockt, Sofie Van Hoecke, Nele Tilley, Bart Merci, Bart Sette, Peter Lambert, Charles-Frederik J. Hollemeersch, and Rik Van de Walle. FireCube: a multi-view localization framework for 3D fire analysis. *Fire Safety Journal*, 46(5):262 – 275, July 2011.

- Steven Verstockt, Sofie Van Hoecke, Nele Tilley, Bart Merci, Bart Sette, Peter Lambert, Charles-Frederik J. Hollemeersch, and Rik Van De Walle. Hot Topics in Video Fire Surveillance. In *Video Surveillance*, pages 443–458. InTech, 2011.
- Steven Verstockt, Nele Tilley, Bart Merci, Charles-Frederik J. Hollemeersch, Bart Sette, Sofie Van Hoecke, Peter Lambert, and Rik Van de Walle. Future directions for video fire detection. *Accepted for publication in proc. 10th international IAFSS symposium*, June 2011.
- Steven Verstockt, Glenn Van Wallendael, Sarah De Bruyne, Chris Poppe, Peter Lambert, and Rik Van de Walle. Cube : a multi-view localization framework for 3D object analysis in computer vision applications. In *Proc. Computer Graphics International 2010 (CGI)*, pages 1–4, June 2010.
- Steven Verstockt, Sarah De Bruyne, Chris Poppe, Peter Lambert, and Rik Van de Walle. Multi-view object localization in H.264/AVC compressed domain. In *Proc. 6th IEEE International conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 370–374, September 2009.
- Tarek Beji, Steven Verstockt, Rik Van de Walle, and Bart Merci. Numerical Prediction of Compartment Fires using Real Time Video Analysis. *Fire Technology*, submitted August 2011.
- Tarek Beji, Steven Verstockt, Rik Van de Walle, and Bart Merci. Prediction of smoke filling in large volumes by means of data assimilation based numerical simulations. *Journal of Fire Sciences*, submitted October 2011.
- Tarek Beji, Steven Verstockt, Rik Van de Walle, and Bart Merci. Numerical simulations to assist real-time smoke control management in large spaces using data assimilation. *Accepted for publication in proc. 7th mediterranean combustion symposium (MSC7)*, September 2011.
- Tarek Beji, Steven Verstockt, Rik Van de Walle, and Bart Merci. Inverse Zone Modelling based Numerical Fire Predictions using Real Time Video and Thermocouple Information. *Accepted for publication in proc. 10th international IAFSS symposium*, June 2011.
- Steven Verstockt, Bart Merci, Bart Sette, Peter Lambert, and Rik Van de Walle. Hot topics in video fire analysis. In *Newsletter of the International Association of Fire Safety Science*, 29: 14–15, August 2010.

Chapter 5

Conclusions

In this dissertation, we have focused on three research topics that concern the video based detection and analysis of fire in large and open spaces. First, we have dealt with the detection of flames and smoke using a single visual, thermal or depth video sensor. Subsequently, we investigated the benefit of multi-modal fire detection using a combination of these sensors. Finally, we explored the possibilities of multi-view fire analysis in order to perform video driven fire forecasting. In the following three sections, we present our contributions on each of these topics. Finally, we point out directions for future work and give an answer to the research question given in Chapter 1.

5.1 Video fire detection

The growing demand for security has given rise to the increased use of video surveillance systems in recent years. Surveillance cameras are rapidly appearing in- and outdoor in all sort of places. This has highlighted various problems such as the fact that it is practically impossible for surveillance operators to keep a constant watch on the video from multiple cameras. Hence, there is need for intelligent video content analysis to support the operators by asking for attention only when unwanted behavior occurs. This is also the reason why, over the last decades, a considerable amount of research has been conducted concerning the intelligent detection, recognition and tracking of ordinary moving objects, such as people and vehicles. Recently, video processing techniques for automatic fire and smoke detection, i.e., the first topic of this dissertation, have also become a hot topic in video surveillance research. The several vision-based detection algorithms that have been proposed in literature have lead to a large amount of video fire detection (VFD) algorithms that can be used to detect the presence of fire at an early stage.

In order to detect the fire, i.e., smoke and/or flames, many VFD algorithms start from video processing building blocks that have already been used successfully in many other video surveillance application domains and extend them with some fire specific feature analysis. Although this already gives satisfactory test results, real-life experiments show that developing a reliable VFD system in the visible spectrum is a tough challenge. Our main contribution to this problem is the introduction of other types of video sensors operating in other spectral bands. Contrary to many other research approaches, the proposed optimizations for the detection of flames and smoke are more in the breadth than in the depth direction. Instead of dealing with ever more complex visual fire detection algorithms, the focus of our first research topic (Chapter 2) is on investigating the benefit of infrared (IR) and time-of-flight (TOF) image processing for fire detection. The latter one, i.e., TOF fire detection, has not been covered by related work until now.

Subjective and objective evaluation (Section 2.3.3, 2.4.4, 2.5.4 & 2.6) on fire and non-fire video sequences revealed that each of the proposed detectors is able to detect smoke or flames at an acceptable level of accuracy. Depending on the environment characteristics, however, one type of detector outperforms the other and vice versa. Based on this fact, we (still) state that only by using multi-modal VFD a 'better' detector can be achieved providing high accuracy under all circumstances. This hypothesis can also be inferred from Table 5.1, which summarizes the pro and contra of each of the proposed visual, TOF, LWIR, and multi-modal smoke/flame detectors and compares them to each other. The table is mainly based upon tests/data provided in this dissertation. As the same evaluation criteria are used as in Table 1.1, the novel video-based detectors can also easily be compared to the 'traditional' detector types which were discussed in the introduction chapter of this book. Each of these criteria is evaluated on a scale ranging from -- to ++, indicating the detector type its weaknesses and strengths respectively.

As indicated in Table 5.1, the purchase price of a visual and TOF camera is still (an order of magnitude) cheaper than a LWIR camera. Furthermore, it is logical that a multi-modal detector is more expensive than the single sensors of which it is composed. As sensor costs are decreasing, however, the purchase price should not be a major criteria in the selection of the 'best' detector. Similarly, as the response time for each of the detectors is within less than a second after the smoke/flames appear in the field of view (fov) of the camera, this time-related criteria also shouldn't have much impact on the detector choice. More 'important' criteria are the sensitivity and the false alarm resistance. As can be seen in the table, the multi-modal detectors score best on these facets.

Table 5.1: Comparison of visual, TOF, LWIR, and multi-modal smoke/flame detectors.

Suitable for poor lighting conditions	-	+	+	+	+	+
Suitable for outdoor use	+	+	-	+	+	+
Suitable for large open spaces	+	+	-	+	+	+
False alarm resistance	+	+	+	+	+	+
Sensitivity	+	+	+	+	+	+
Response time	+	+	+	+	+	+
Installation cost	+	+	+	-	-	+
Purchase price	+	-	+	-	-	+
Detector type	Visual flame/smoke detector	LWIR flame detector	TOF flame/smoke detector	Visual-LWIR smoke detector	Visual-LWIR flame detector	Visual-TOF flame detector

The experiments throughout this dissertation revealed that the average gain of using a multi-modal detector is almost 4% on the detection rate (\sim sensitivity), with an average false alarm reduction of 3%. From the individual detectors, the LWIR and TOF detector slightly outperform the visual detectors in our experiments. However, as each of them has its own specific limitations, a better detector can only be achieved by using multi-modal VFD. A drawback of the multi-modal detectors (and the LWIR detector) is that their installation cost (e.g., registration/calibration and temporal alignment) is higher. However, as manufacturers start to combine multiple sensors within the same device, this will soon no longer be an issue. Finally, multi-modal detectors have also shown more suitable for large open spaces, outdoor use and poor lighting conditions.

5.2 Multi-modal fire detection

The majority of single-sensor VFD systems, such as those presented in Chapter 2, are plagued by a number of difficulties in real-world scenes, e.g., lighting, reflections and noise. Many of these difficulties are caused by limitations due to the type of sensors used. To overcome these sensor-related limitations, multiple types of sensors, like a color, IR and/or TOF camera, can be combined using the multi-modal detection techniques described in Chapter 3. The proposed multi-sensor flame and smoke detectors take advantage of the different kinds of information represented by thermal, visual and/or depth images in order to accurately detect the fire. By fusing the multi-modal modalities and using the strengths of each medium, high accuracy is achieved under all circumstances, providing gains over ‘classical’ single-sensor VFD. Merging information from multi-modal sensors has, as such, proven to be a win-win (\sim Table 5.1).

An important aspect when developing multi-modal algorithms is the registration of the multi-modal images. The goal of registration is to establish geometric correspondence between the multi-sensor images so that they may be transformed, compared, and analyzed in a common reference frame. The proposed silhouette based registration algorithm analyses the contours and the correlation of visual and thermal silhouettes of moving objects. First, the rotation is computed using silhouette contour extraction and circular cross correlation. Next, contour scaling is used to estimate the thermal-visual scale factor. Finally, the translation vector is estimated by maximization of binary correlation. The geometric parameters found during this registration phase are further used by each of the multi-modal detectors to coarsely map the visual, thermal and/or depth images.

Based on experiments on different challenging fire and non-fire sequences, it was found objectively that each of the detectors adhere to all the relevant requirements: object-based automatic calibration/registration, low number of false alarms, no missed detections and fast warning/alarming with different levels of detection. Due to the low-cost of the proposed techniques, such as the silhouette coverage analysis and the visual silhouette disorder analysis (which is only performed if smoke warning is given), the multi-modal detectors are also less computational expensive as many of the existing individual detectors. Furthermore, as the ‘work’ is limited, real-time requirements imposed by the fire industry are met.

To conclude, the visual-LWIR and visual-TOF flame and smoke detectors have shown that the relevance and usefulness of multi-modal video analysis cannot be ignored. But neither can the reality that single-sensor analysis will continue to dominate at least the next couple of years. The downsides of multiple types of sensors, e.g., the extra sensor costs, the (re)calibration/registration and the setup time, still put up barriers for a potential breakthrough of multi-modal video surveillance (at least in the short term). However, despite these ‘downsides’, it is still our strong belief that low cost algorithms running on multiple sensors will soon start to take over the ever more complex single-sensor algorithms that are proposed in most publications today. The fusion of multi-modal data will definitely become the keyword in video surveillance in the near future.

5.3 Multi-view fire analysis

In Chapter 4 we tackled the domain of video-based fire analysis. As we found that single-view processing is not enough in order to actually understand and interpret the fire, we decided to combine the detection results of multiple single-view cameras and analyze them together. In this way, more accurate detection and localization of smoke and flames is achieved and valuable fire characteristics are detected at the early stage of the fire.

The main part of our fire analysis contribution consists of a novel multi-view fire analysis framework which fuses low-cost video fire detection results of multiple cameras into the FireCube. Using the FireCube, the location of the fire, its size, its propagation and its direction can accurately be estimated. The proposed multi-modal detection and multi-view localization techniques have been tested thoroughly on fire and non-fire video sequences and have proven to provide gains over ‘classical’ single-view fire analysis techniques.

Two cameras are already sufficient to achieve a dimension accuracy of 90% and a position accuracy of 98%. By further increasing the number of cameras, it is even possible to achieve a dimension accuracy of 96% and a position accuracy of 99%. Furthermore, the experiments show that increasing the number of cameras to monitor the scene has a positive effect on the detection rate. The gain of using four cameras instead of one is 3%.

The video-based detection of valuable fire characteristics, e.g., fire size and smoke layer height, also opened the door to video driven fire forecasting. Being able to model and forecast the fire can help emergency services to work more efficiently and save lives. However, the calculations with existing modeling techniques still take too long and valuable time is often lost. Using the multi-view fire analysis framework, which is able to give real-time information about the state of the environment, these zone model-based predictions of the future state can be improved and accelerated. By combining the information about the fire from models and real-time data, an estimate of the fire can be produced that is better than could be obtained from using the model or the data alone. This is also shown by our preliminary experiments, which will be continued in the ongoing ‘data assimilation’ work of one of our colleagues [144]. To conclude, we can say that the research in video driven fire forecasting is only just starting to bloom and a wide-spread adoption and efficient use can be predicted in the near future.

5.4 Future work

Future work directly in the line of this dissertation is evaluating the performance of the different types of cameras and algorithms in more challenging and real fire environments. As most of our recordings were performed on rather controlled fires/environments, we did (mainly due to security reasons) not yet investigate the impact of real fires on the camera technology and the detection accuracy. This is definitely something that needs further investigation, as was already discussed in Section 2.4.5. Another aspect that needs further research is the investigation/evaluation of feature alternatives and more advanced strategies to combine these feature values. This is already partially addressed in Section 2.7 as one of the strategies to further improve the proposed detection algorithms.

Related to the performance-related question of the different types of cameras and algorithms, the evaluation and testing of video based fire detection and analysis systems should also be standardized in the future. For the moment, several types of testing and evaluation are found in literature, which

complicates the comparison of techniques. Commercial fire detection systems mainly use the NFPA 72 standard [4], which was written to provide requirements for the installation, performance, testing, inspection, and maintenance of a fire alarm. We think it would be a good idea to develop a video variant of this standard, which can be used to easily evaluate VFD techniques against each other and against other types of fire detectors.

Another point of interest is that, within the proposed fire analysis framework, the sensors have no direct effect on each other. Each sensor stands on its own, delivering video, thermal, depth or amplitude information without taking into account feedback from the other neighboring sensors. Possibly, the fire analysis framework can be further enhanced if the available information and intelligence of all sensors can be fed back to the sensors to optimize their functionality. Related to sensor feedback, the framework can also use environmental characteristics (such as geo-data and meteorological info) to select optimal sensor parameters. Both future research topics, i.e., the use of sensor feedback/fine-tuning and the investigation/evaluation of combining video-based sensors with conventional/traditional fire detectors, are discussed in more detail in Section 4.10.

Similar to current efforts at Bilkent University [108], it could also be interesting to extend the proposed VFD techniques to PTZ based VFD. To monitor a large area, many cameras are often needed. For several video surveillance applications, e.g., wildfire detection and monitoring of large car parks, this is unaffordable. Instead of using multiple cameras, these applications mostly use a dynamic PTZ camera. One of the drawbacks of PTZ, however, is that the position of everything in the scene can change between consecutive image acquisitions, which complicates for example the background modeling and object tracking/analysis. This makes it, for the moment, difficult to predict how the proposed algorithms will behave in a PTZ setting and how costly (required) adaptations for PTZ fire detection will be. A recent literature survey did also not reveal any research in this direction.

Besides the research on PTZ based VFD, a more thorough investigation of the fire detection capabilities of the different IR spectral bands (NIR/SWIR, MWIR, LWIR) seems also interesting as a future work. Related to this, we also plan to investigate the benefit of fusing different infrared spectral cameras into a multi-modal infrared detector. Xenics, i.e., one of our research partners, already started research on the technical aspect in this direction [143]. Furthermore, extensions can be made to the fire analysis framework to further optimize the video driven fire forecasting. We plan, for example, to incorporate other non-video sensors, such as thermocouple data, into the FireCube.

Finally, perhaps the most interesting topic for further research is to combine the knowledge from the fire and the video world. As a first step in this direction, Section 4.9.4 proposes an algorithm for the detection of the smoke layer height, which is mainly based on a fire engineering technique. To the best of our knowledge, this is the first time that a technique from the ‘fire world’ is applied to video analysis.

Based on the ‘positive’ result of our first multidisciplinary interchange of techniques, we still see plenty of room for further research on the interplay between both the video and fire world. Related to this, future research should also focus on further closing the gap between both worlds. In the field of interfacing and communicating the fire data, for example, there still exists a lot of research potential. We see the work in this dissertation as a first step to bridge the gap and to stimulate the interaction with the fire world.

5.5 Answer to the research question

We end this dissertation with an answer to the different aspects of the research question given in Chapter 1. The central question of this thesis was:

‘Can we develop an algorithm to timely and accurately detect/analyze the fire in large and open public places and can we use the extracted fire characteristics for video driven fire forecasting?’.

First of all, we can **‘develop an algorithm to detect the fire’**. Several video based smoke and flame detection algorithms existed even before our research started. It is the second aspect: **‘in large and open public places’**, that makes the problem difficult. Videos from this kind of surveillance scenes, e.g., car parks, shopping malls and atria, often contain difficulties, such as changing/limited illumination, shadows and noise, which makes the detection error-prone. For such scenes, our research advances the state of the art, for example by introducing low-cost multi-modal fire detectors which are able to cope with many of the sensor-related ‘limitations’. The main benefit of (f)using multi-modal image data is that unreliably extracted parts from one sensor might be reliably extracted from the other sensor. By using the strengths of each medium, fire detection is done more accurately and with fewer false detections.

As was already mentioned above, the presented work also fulfilled the **‘timely and accurately’** aspect of the research question. The several experiments in this dissertation show a significant accuracy improvement, specifically due to the combination of low-cost flame and smoke features. By focusing on ‘low-cost’ fire features, we are able to keep the processing cost low, i.e., to ensure real-time detection. As soon as smoke or flames appear in the field of view of the camera, fire alarm is given. Whether the proposed detectors are already accurate enough for real-world operation is a more difficult question. As the majority of our experiments were performed in a ‘controlled’ environment, we can only guarantee that the detectors are accurate enough in the less difficult surveillance scenes, where not too many real-world objects appear simultaneously. Difficult scenes, with high moving object densities for example, will possibly decrease the detection accuracy.

Related to the ‘timely and accurately’ aspect of the research question, we would like to remark again that expecting automatic video surveillance to be able to detect and analyze everything without false alarms or missed detections and with only one simple configuration is unrealistic. To be honest, we believe this will never be possible. As such, the techniques proposed in this dissertation must not be seen as ‘the’ ultimate fire detection tools. They must be seen as a complement to the existing techniques.

We see the video-based detection and analysis as a tool for operators, which reduces/simplifies a lot of their work. Based on the visual, LWIR, TOF and other sensors input, the operator/fire expert can then take a decision. In the end, an operator will (mostly) be able to better estimate the risk of the fire (based on his knowledge of the environmental context, his experience, etc.). In Fig. 5.1 we show some examples of situations in which video fire detection will have problems to see the difference between risky and non-risky fires.

The fourth aspect of the question concerned the **‘fire analysis’** capabilities of a video system. Contrarily to fire detection, the amount of literature in this direction was limited when we started our research. Even today, most video video-based fire alarm systems just ring the bells, i.e., they only detect the presence of fire and are not able to model fire evolution. Even though the majority of these systems consist of several cameras monitoring the same scene, the analysis is usually carried out separately on each of the camera’s video sequences. However, by combining the detection results of each of these single-view cameras and analyze them together, as is done in our work, more accurate detection and localization of smoke and flames is achieved and valuable fire characteristics are detected at the early stage of the fire.

The fifth and last part of the research question was: ‘**can we use the extracted fire characteristics for video driven fire forecasting**’. To the best of our knowledge, we are the first who have tackled this subject. The work described on video driven fire spread forecasting is a step in the direction of an application aiding firefighters in assessing the fire risk more efficiently.



Bengal light, Barcelona, 2011



a) Bradford Stadium Fire, England, 1985



Will cameras see the difference?
→ can they estimate the risk?



Figure 5.1: VFD problems: will cameras see the difference?

However, the problem on ‘video driven fire forecasting’ is certainly not completely solved. The work gives a solid basis for the extraction of fire characteristics and how this information can be used to estimate the future state of the fire. The next step is to further develop/improve the data assimilation methodology, e.g., by incorporating thermocouple data, which is ongoing work of one of our research partners.

Being able to answer each aspect of the research question, we hope to have convinced the reader that multi-modal and multi-view fire analysis has its advantages in safety analysis and fire fighting/mitigation, and is essential in assessing the risk of escalation. Furthermore, it is important to stress that many of the investigated aspects are not limited to fire detection. They can easily be adapted to other application domains, e.g., multi-modal object recognition and traffic analysis, which can benefit from the new insights in this dissertation. As such, the results in this thesis are not only of scientific importance for fire detection, but also for video surveillance in general. Based on this and on the fact that the video surveillance market is growing rapidly, it is our belief that the proposed contributions will only increase in value in the coming years.

I can see clearly now, the 'fire' is gone!
- Johnny Cash & Steven Verstockt -

Publications

Journal papers

1. Steven Verstockt, Chris Poppe, Sofie Van Hoecke, Charles-Frederik J. Hollemeersch, Bart Merci, Bart Sette, Peter Lambert, and Rik Van de Walle. Silhouette-based multi-sensor smoke detection: coverage analysis of moving object silhouettes in thermal and visual registered images. *Machine Vision and Applications*, DOI 10.1007/s00138-011-0359-3, July 2011.
2. Steven Verstockt, Sofie Van Hoecke, Nele Tilley, Bart Merci, Bart Sette, Peter Lambert, Charles-Frederik J. Hollemeersch, and Rik Van de Walle. FireCube: a multi-view localization framework for 3D fire analysis. *Fire Safety Journal*, 46(5):262 – 275, July 2011.
3. Steven Verstockt, Charles-Frederik J. Hollemeersch, Chris Poppe, Peter Lambert, Rik Van de Walle, Sofie Van Hoecke, Bart Merci, and Bart Sette. Multi-sensor fire detection by fusing visual and LWIR flame feature. *ICGST International Journal on Graphics, and Image Processing*, 10(6): 43–50, December 2010.

Submitted journal papers

1. Steven Verstockt, Sofie Van Hoecke, Tarek Beji, Bart Merci, Benedict Gouverneur, A. Enis Cetin, Pieterjan De Potter, and Rik Van de Walle. A multi-modal video analysis approach for car park fire detection. *Special issue in Fire Safety Journal: Car Park Fire Safety*, submitted July 2011.
2. Steven Verstockt, Sofie Van Hoecke, Pieterjan De Potter, Peter Lambert, Charles-Frederik J. Hollemeersch, Bart Merci, Bart Sette, and Rik

Van de Walle. Multi-modal time-of-flight based fire detection. *Special issue in Multimedia Tools and Applications: Analysis and Retrieval of Events/Actions and Workflows in Video Streams*, submitted March 2011.

3. Tarek Beji, Steven Verstockt, Rik Van de Walle, and Bart Merci. Numerical Prediction of Compartment Fires using Real Time Video Analysis. *Fire Technology*, submitted August 2011.
4. Tarek Beji, Steven Verstockt, Rik Van de Walle, and Bart Merci. Prediction of smoke filling in large volumes by means of data assimilation based numerical simulations. *Journal of Fire Sciences*, submitted October 2011.

Book chapters

1. Steven Verstockt, Sofie Van Hoecke, Nele Tilley, Bart Merci, Bart Sette, Peter Lambert, Charles-Frederik J. Hollemeersch, and Rik Van De Walle. Hot Topics in Video Fire Surveillance. In *Video Surveillance*, pages 443–458. InTech, 2011.

Papers in conference proceedings

1. Steven Verstockt, Ioannis Kypraios, Pieterjan De Potter, Chris Poppe, and Rik Van de Walle. Wavelet-based multi-modal fire detection. In *Proc. 19th European Signal Processing Conference (EUSIPCO)*, September 2011.
2. Steven Verstockt, Pieterjan De Potter, Peter Lambert, Sofie Van Hoecke, and Rik Van de Walle. Multi-sensor fire detection using visual and time-of-flight imaging. In *Proc. 1st IEEE International Workshop on Advances in Automated Multimedia Surveillance for Public Safety, International Conference on Multimedia and Expo (ICME)*, July 2011.
3. Steven Verstockt, Nele Tilley, Bart Merci, Charles-Frederik J. Hollemeersch, Bart Sette, Sofie Van Hoecke, Peter Lambert, and Rik Van de Walle. Future directions for video fire detection. In *Proc. 10th international IAFSS symposium*, June 2011.

4. Steven Verstockt, Chris Poppe, Pieterjan De Potter, Sofie Van Hoecke, Charles-Frederik J. Hollemeersch, Peter Lambert, and Rik Van de Walle. Silhouette coverage analysis for multi-modal video surveillance. In *Proc. 29th Progress In Electromagnetics Research Symposium*, pages 1279–1283, March 2011.
5. Steven Verstockt, Sofie Van Hoecke, Onciny Meyfrootd, and Rik Van de Walle. Assistive smartphone for people with special needs : the personal social assistant. In *Proc. Medicine 2.0 : Social Media and Web 2.0 in Health, Medicine and Biomedical Research*, pages 69, November 2010.
6. Steven Verstockt, Alexander Vanoosthuysse, Bart Merci, Nele Tilley, Bart Sette, Charles-Frederik J. Hollemeersch, Peter Lambert, and Rik Van de Walle. Performance evaluation framework for vision-based fire detection. In *Proc. 12th International conference on Fire Science and Engineering (Interflam 2010)*, pages 257–268, July 2010.
7. Steven Verstockt, Alexander Vanoosthuysse, Sofie Van Hoecke, Peter Lambert, and Rik Van de Walle. Multi-sensor fire detection by fusing visual and non-visual flame features. In *Proc. 4th International Conference on Image and Signal Processing*, pages 333–341, June 2010.
8. Steven Verstockt, Glenn Van Wallendael, Sarah De Bruyne, Chris Poppe, Peter Lambert, and Rik Van de Walle. Cube : a multi-view localization framework for 3D object analysis in computer vision applications. In *Proc. Computer Graphics International 2010 (CGI)*, pages 1–4, June 2010.
9. Steven Verstockt, Rudy Dekeerschieter, Alexander Vanoosthuysse, Bart Merci, Bart Sette, Peter Lambert, and Rik Van de Walle. Video fire detection using non-visible light. In *Proc. 6th International seminar on Fire and Explosion Hazards (FEH-6)*, pages 49–50, April 2010.
10. Steven Verstockt, Peter Lambert, Rik Van de Walle, Bart Merci, and Bart Sette. State of the art in vision-based fire and smoke detection. In *Proc. 14th International Conference on Automatic Fire Detection (AUBE)*, pages 285–292, September 2009.
11. Steven Verstockt, Sarah De Bruyne, Chris Poppe, Peter Lambert, and Rik Van de Walle. Multi-view object localization in H.264/AVC compressed domain. In *Proc. 6th IEEE International conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 370–374, September 2009.

12. Steven Verstockt, Sebastiaan Van Leuven, Rik Van de Walle, Elmar Dermaut, Steven Torelle, and Wouter Gevaert. Actor recognition for interactive querying and automatic annotation in digital video. In *Proc. 13th IASTED international conference on Internet and Multimedia Systems and Applications*, pages 149–155, Augustus 2009.
13. Steven Verstockt, Bart Merci, Peter Lambert, and Rik Van de Walle. Video processing techniques for early fire detection. In *Proc. International WildFire Management conference*, June 2009.
14. Steven Verstockt, Davy Decoo, Daute Van Nieuwenhuysse, Filip De Pauw, and Rik Van de Walle. Assistive Smartphone for people with special needs : the Personal Social Assistant. In *Proc. 2nd Conference on Human System Interactions*, pages 328–334, May 2009.
15. Steven Verstockt, Peter Lambert, and Rik Van de Walle. Feature extraction for localized CBIR : what you click is what you get. In *Proc. 4th International Conference on Computer Vision Theory and Applications*, pages 373–376, February 2009.
16. Sofie Van Hoecke, Steven Verstockt, Koen Samyn Tunnel Simulator for Traffic Video Detection In *Proc. 3rd International Conference on Advances in System Simulation (SIMUL)*, November 2011.
17. Tarek Beji, Steven Verstockt, Rik Van de Walle, and Bart Merci. Numerical simulations to assist real-time smoke control management in large spaces using data assimilation. In *Proc. 7th mediterranean combustion symposium (MSC7)*, September 2011.
18. Tarek Beji, Steven Verstockt, Rik Van de Walle, and Bart Merci. Inverse Zone Modelling based Numerical Fire Predictions using Real Time Video and Thermocouple Information. In *Proc. 10th international IAFSS symposium*, June 2011.
19. Benedict Gouverneur, Steven Verstockt, Guy Gielis, Stefan Nemeth, Thomas Bocquet, Jan Stynen, Jan Vermeiren The development of a multiband system for early detection of wildlife fires and indoor search and rescue operations. In *Proc. SPIE defence, security and sensing, April 2011*
20. Pieterjan De Potter, Steven Verstockt, Ioannis Kypraios, Chris Poppe, Peter Lambert, and Rik Van de Walle. Automatic Available Seat Counting In Public Rail Transport Using Wavelets. In *Proc. 53rd International Symposium ELMAR-2011*, September 2011.

21. Pieterjan De Potter, Christophe Billiet, Chris Poppe, Brecht Stubbe, Steven Verstockt, Peter Lambert, and Rik Van de Walle. Available Seat Counting in Public Rail Transport. In *Proc. 29th Progress In Electromagnetics Research Symposium*, pages 1294–1298, March 2011.
22. Chris Poppe, Steven Verstockt, Sarah De Bruyne, and Rik Van de Walle. Multi-camera analysis of soccer sequences. In *Proc. 7th IEEE International conference on Advanced Video and Signal Based Surveillance (AVSS 2010)*, pages 26–31, Augustus 2010.
23. Sarah De Bruyne, Chris Poppe, Steven Verstockt, Peter Lambert, and Rik Van de Walle. Estimating motion reliability to improve moving object detection in the H.264/AVC domain. In *Proc. IEEE International Conference on Multimedia and Expo (ICME)* pages 330-333, June 2009.

Other publications

1. Steven Verstockt, Bart Merci, Bart Sette, Peter Lambert, and Rik Van de Walle. Hot topics in video fire analysis. In *Newsletter of the International Association of Fire Safety Science*, 29: 14–15, August 2010.
2. Steven Verstockt, Peter Lambert, Bart Merci, Rik Van de Walle, and Bart Sette. Video fire detection going faster than fire. In *Annual report and newsletter of European Group of Organisations for Fire Testing, Inspection and Certification*, 8:12–13.

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 *Proceedings of the 14th International Conference on Automatic Fire Detection*, volume 2, pages 285–292, September 2009.
- [2] A.P.E. Cote. *Fundamentals of fire protection: Automatic Fire detection and alarm system*. National Fire Protection Association, Quincy, MA, 2004.
- [3] W.D. Davis and K.A. Notarianni. NASA fire detection study. Technical report, National Institute of Standards and Technology, March 1996.
- [4] NFPA. NFPA 72: National fire alarm and signaling code. Technical report, National Fire Protection Association, 2010.
- [5] D.P. Nolan. *Handbook of Fire and explosion protection engineering principles (2nd ed.): Fire and Gas Detection and Alarm Systems*. Elsevier, Oxford, UK, 2011.
- [6] Sense-WARE. *Flame and Gas detection: Comparison Table Automatic Fire Detectors*. <http://www.flame-detection.net/flame-detection/?download=Comparison-Table-Fire-Detection.pdf>.
- [7] Somati. *Somati: protecting today's future*. <http://www.somati.be/>.
- [8] M.J. Jr. Karter. U.S. fire loss for 2007. Technical report, National Fire Protection Association, 2007.
- [9] G. Hadjisophocleous, J. Ouyang, G. Ding, and Z. Liu. Study of video image fire detection systems for protection of large industrial applications and atria. In *Proceedings of Suppression and Detection Research and Applications (SUPDET) A Technical Working Conference*, pages 1–13, February 2010.
- [10] W. Kuffner. *Method of Determining Smoke Detector Spacing in High Ceiling Applications*. PhD thesis, Carleton University, Canada, 2009.
- [11] D. Gottuk. Video image detection system installation performance criteria. Technical report, Fire Protection Research Foundation, October 2008.

- [12] A. Khadivi and M. Hasler. Fire detection and localization using wireless sensor networks. In *Proceedings of First International Conference on Sensor Applications, Experimentation, and Logistics*, pages 16–26, September 2009.
- [13] E. Zervas, A. Mpimpoudis, C. Anagnostopoulos, O. Sekkas, and S. Hadjiefthymiades. Multisensor data fusion for fire detection. *Information Fusion*, 12(3):150–159, July 2011.
- [14] W. Hu, T. Tan, L. Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on systems, man and cybernetics*, 34(3):334–352, August 2004.
- [15] P.J. Withagen. *Object detection and segmentation for visual surveillance*. PhD thesis, Intelligent Autonomous Systems (IAS), University of Amsterdam, 2006.
- [16] Y. Wang, D.M. Krum, E.M. Coelho, and D.A. Bowman. Contextualized videos: Combining videos with environment models to support situational understanding. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1568–1575, Nov/Dec 2007.
- [17] Y. Dedeoglu, B.U. Toreyin, U. Gudukbay, and A.E. Cetin. Silhouette-based method for object classification and human action recognition in video. In *Proceedings of ECCV Workshop on HCI*, pages 64–77, May 2006.
- [18] L. Wang, W. Hu, and T. Tan. Recent developments in human motion analysis. *Pattern Recognition*, 36(3):585–601, March 2003.
- [19] G.L. Foresti, C.S. Regazzoni, and P.K. Varshney. *Multisensor surveillance systems: the fusion perspective*. Kluwer, Massachusettes, 2003.
- [20] G. Mohammadi, F. Dufaux, T.H. Minh, and T. Ebrahimi. Multi-view video segmentation and tracking for video surveillance. In *Proceedings of SPIE Mobile Multimedia/Image Processing, Security, and Applications*, pages –, April 2009.
- [21] L. Jiao, Y. Wu, G. Wu, E.Y. Chang, and Y.-F. Wang. Multimedia systems. *Anatomy of a multicamera video surveillance system*, 10(2):144–163, August 2004.
- [22] D.M. Hansen, P.T. Duizer, S. Park, T.B. Moeslund, and M.M. Trivedi. Multi-view video analysis of humans and vehicles in an unconstrained environment. In *Proceedings of the 4th International Symposium on Advances in Visual Computing (ISVC '08)*, pages 428–439, December 2008.
- [23] R. Hartley and A. Zisserman. *Estimation - 2D projective transformations, Multiple view geometry in computer vision (2nd ed.)*. Cambridge University Press, 2004.
- [24] S.J. Krotosky and M.M. Trivedi. Mutual information based registration of multimodal stereo videos for person tracking. *Computer Vision and Image Understanding*, 106(2-3):270–287, May 2007.

- [25] B.C. Arrue, A. Ollero, and J.R. Martinez de Dios. An intelligent system for false alarm reduction in infrared forest-fire detection. *IEEE Intelligent Systems*, 15(3):64–73, May 2000.
- [26] G. Pieri and D. Moroni. Active video surveillance based on stereo and infrared imaging. *EURASIP Journal on Advances in Signal Processing*, 2008(380210):1–8, January 2008.
- [27] Xenics. *Infrared solutions: situational firefighting*. <http://www.xenics.com/>.
- [28] Fire Safety Engineering (Bart Merci). *Fire Safety and Explosion Safety in Car Parks*. <http://www.carparkfiresafety.be/>.
- [29] W. Phillips, M. Shah, and N. da Vitoria Lobo. Flame recognition in video. *Pattern recognition letters*, 23(1-3):319–327, January 2002.
- [30] F. Gomez-Rodriguez, S. Pascual-Pena, B. Arrue, and A. Ollero. Smoke detection using image processing. In *Proceedings of 4th International Conference on Forest Fire Research & Wildland Fire Safety*, pages 1–8, November 2002.
- [31] F. Gomez-Rodriguez, B.C. Arrue, and A. Ollero. Smoke monitoring and measurement using image processing - application to forest fires. In *Proceedings of SPIE AeroSense 2003: XIII Automatic Target Recognition*, pages 404–411, April 2003.
- [32] T.-H. Chen, P.-H. Wu, and Y.-C. Chiou. An early fire-detection method based on image processing. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, volume 3, pages 1707–1710, October 2004.
- [33] C.B. Liu and N. Ahuja. Vision based fire detection. In *Proceedings of 17th International Conference on Pattern Recognition (ICPR)*, volume 4, pages 134–137, August 2004.
- [34] G. Marbach, M. Loepfe, and T. Brupbacher. An image processing technique for fire detection in video images. *Fire Safety Journal*, 41(4):285–289, June 2006.
- [35] B.U. Toreyin, Y. Dedeoglu, U. Gudukbay, and A.E. Cetin. Computer vision based method for real-time fire and flame detection. *Pattern Recognition Letters*, 27(1):49–58, January 2006.
- [36] B.U. Toreyin, Y. Dedeoglu, and A.E. Cetin. Contour based smoke detection in video using wavelets. In *Proceedings of European Signal Processing Conference (EUSIPCO)*, September 2006.
- [37] T. Celik, H. Ozkaramanli, and H. Demirel. Fire and smoke detection without sensors: image processing based approach. In *Proceedings of 15th European Signal Processing Conference (EUSIPCO)*, pages 1794–1798, September 2007.

- [38] Z. Xu and J. Xu. Automatic fire smoke detection based on image visual features. In *Proceedings of International Conference on Computational Intelligence and Security Workshops*, pages 316–319, December 2007.
- [39] T. Celik, H. Demirel, H. Ozkaramanli, and M. Uyguroglu. Fire detection using statistical color model in video sequences. *Journal of Visual Communication and Image Representation*, 18(2):176–185, January 2007.
- [40] Z. Xiong, R. Caballero, H. Wang, A.M. Finn, M. A. Lelic, and P.-Y. Peng. Video-based smoke detection: possibilities, techniques, and challenges. In *Proceedings of Suppression and Detection Research and Applications (SUPDET) A Technical Working Conference*, 2007.
- [41] B. Lee and D. Han. Real-time fire detection using camera sequence image in tunnel environment. In *Proceedings of International Conference on Intelligent Computing*, pages 1209–1220, August 2007.
- [42] S. Calderara, P. Piccinini, and V. Cucchiara. Smoke detection in video surveillance: a mog model in the wavelet domain. In *Proceedings of 6th International Conference in Computer Vision Systems(ICVS)*, pages 119–128, May 2008.
- [43] P. Piccinini, S. Calderara, and R. Cucchiara. Reliable smoke detection system in the domains of image energy and color. In *Proceedings of International Conference on Image Processing*, pages 1376–1379, October 2008.
- [44] F. Yuan. A fast accumulative motion orientation model based on integral image for video smoke detection. *Pattern Recognition Letters*, 29(7):925–932, May 2008.
- [45] P.V.K. Borges, J. Mayer, and E. Izquierdo. Efficient visual fire detection applied for video retrieval. In *Proceedings of 16th European Signal Processing Conference (EUSIPCO)*, August 2008.
- [46] X. Qi and J. Ebert. A computer vision based method for fire detection in color videos. *International Journal of Imaging*, 2(S09):22–34, Spring 2009.
- [47] R. Yasmin. Detection of smoke propagation direction using color video sequences. *International Journal of Soft Computing*, 4(1):45–48, 2009.
- [48] J. Gubbi, S. Marusic, and M. Palaniswami. Smoke detection in video using wavelets and support vector machines. *Fire Safety Journal*, 44(8):1110–1115, November 2009.
- [49] J. Chen, Y. He, and J. Wang. Multi-feature fusion based fast video flame detection. *Building and Environment*, 45(5):1113–1122, May 2010.
- [50] O. Gunay, K. Tasdemir, B.U. Treyin, and A.E. Cetin. Fire detection in video using lms based active learning. *Fire Technology*, 46(3):551–577, 2010.
- [51] I. Kolesov, P. Karasev, A. Tannenbaum, and E. Haber. Fire and smoke detection in video with optimal mass transport based optical flow and neural networks. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 761–764, September 2010.

- [52] B.C. Ko, K.H. Cheong, and J.Y. Nam. Early fire detection algorithm based on irregular patterns of flames and hierarchical bayesian networks. *Fire Safety Journal*, 45(4):262–270, June 2010.
- [53] R.A. Gonzalez-Gonzalez, V. Alarcon-Aquino, O. Starostenko, R. Rosas-Romero, J.M. Ramirez-Cortes, and J. Rodriguez-Asomoza. Wavelet-based smoke detection in outdoor video sequences. In *Proceedings of the 53rd IEEE Midwest Symposium on Circuits and Systems (MWSCAS)*, pages 383–387, August 2010.
- [54] P.V.K. Borges and E. Izquierdo. A probabilistic approach for vision-based fire detection in videos. *IEEE Transactions on circuits and systems for video technology*, 20(5):721–731, May 2010.
- [55] D. Van Hamme, P. Veelaert, W. Philips, and K. Teelen. Fire detection in color images using markov random fields. In *Proceedings of Advanced Concepts for Intelligent Vision Systems (ACIVS)*, volume 2, pages 88–97, December 2010.
- [56] T. Celik. Fast and efficient method for fire detection using image processing. *ETRI Journal*, 32(6):881–890, December 2010.
- [57] F. Yuan. Video-based smoke detection with histogram sequence of lbp and lbpv pyramids. *Fire Safety Journal*, 46(3):132–139, April 2011.
- [58] L. Rossi, M. Akhloufi, and Y. Tison. On the use of stereo vision to develop a novel instrumentation system to extract geometric fire fronts characteristics. *Fire Safety Journal (Forest Fires)*, 46(1-2):9–20, January 2011.
- [59] IBBT. *ISYSS project: Intelligent SYstems for Security and Safety*. <http://www.ibbt.be/en/projects/overview-projects/p/detail/isyss>.
- [60] R.T. Collins, A.J. Lipton, and T. Kanade. A system for video surveillance and monitoring. In *Proceedings of American Nuclear Society (ANS) Eighth International Topical Meeting on Robotics and Remote Systems*, 1999.
- [61] S. Calderara, P. Piccinini, and R. Cucchiara. Vision based smoke detection system using image energy and color information. *Machine Vision and Applications*, pages 1–15, May 2010.
- [62] J. Yang and R.S. Wang. A survey on fire detection and application based on video image analysis. *Video Engineering*, 2006(8):92–96, 2006.
- [63] G. Doretto, A. Chiuso, Y.N. Wu, and S. Soatto. Dynamic textures. *International Journal of Computer Vision*, 51(2):91–109, 2003.
- [64] T. Amiaz, S. Fazekas, D. Chetverikov, and N. Kiryati. Detecting regions of dynamic texture. In *Proceedings of International Conference on Scale-Space and variational methods in Computer Vision (SSVM)*, pages 848–859, May 2007.

- [65] D. Chetverikov and R. Peteri. A brief survey of dynamic texture description and recognition. In *Proceedings of 4th International Conference on Computer Recognition Systems*, pages 17–26, May 2005.
- [66] B.U. Toreyin, Y. Dedeoglu, A.E. Cetin, S. Fazekas, D. Chetverikov, T. Amiaz, and N. Kiryati. Dynamic texture detection, segmentation and analysis. In *Proceedings of ACM International Conference on Image and Video Retrieval (CIVR)*, pages 131–134, July 2007.
- [67] S. Fazekas, T. Amiaz, D. Chetverikov, and N. Kiryati. Dynamic texture detection based on motion analysis. *International Journal of Computer Vision*, 82(1):48–63, April 2009.
- [68] S. Fazekas and D. Chetverikov. Analysis and performance evaluation of optical flow features for dynamic texture recognition. *Signal Processing: Image Communication, Special Issue on Content-Based Multimedia Indexing and Retrieval*, 22(7-8):680–691, August 2007.
- [69] R. Péteri, M. Huskies, and S. Fazekas. *DynTex: A comprehensive database of dynamic textures (2006)*. <http://projects.cwi.nl/dyntex/>.
- [70] R.C. Gonzales and R.E. Woods. *Digital Image Processing (3rd ed.)*. Prentice-Hall, New Jersey, 2002.
- [71] J. Han and B. Bhanu. Fusion of color and infrared video for moving human detection. *Pattern Recognition*, 40(6):1771–1784, June 2007.
- [72] R. Vandersmissen. Night-vision camera combines thermal and low-light-level images. *Photonik International*, 2008(2):2–4, Augustus 2008.
- [73] 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*, 41(4):315–320, June 2006.
- [74] B.U. Toreyin, R.G. Cinbis, Y. Dedeoglu, and A.E. Cetin. Fire detection in infrared video using wavelet analysis. *SPIE Optical Engineering*, 46(6):1–9, June 2007.
- [75] I. Bosch, S. Gomez, R. Molina, and R. Miralles. Object discrimination by infrared image processing. In *Proceedings of the 3rd International Work-Conference on The Interplay Between Natural and Artificial Computation (IWINAC)*, pages 30–40, June 2009.
- [76] O. Gunay, K. Tasdemir, B.U. Treyin, and A.E. Cetin. Video based wildfire detection at night. *Fire Safety Journal*, 44(6):860–868, August 2009.
- [77] S. Verstockt, R. Dekeerschieter, A. Vanoosthuysse, B. Merci, B. Sette, P. Lambert, and R. Van de Walle. Video fire detection using non-visible light. In *Proceedings of the 6th International Seminar on Fire and Explosion Hazards*, April 2010.

- [78] H. Hügli and T. Zamofing. Pedestrian detection by range imaging. In *Proceedings of 2nd International Conference on Computer Vision Theory and Applications*, pages 18–22, March 2007.
- [79] R. Tanner, M. Studer, A. Zanolini, and A. Hartmann. People detection and tracking with ToF sensor. In *Proceedings of the 5th International Conference on Advanced Video and Signal Based Surveillance*, pages 356–361, September 2008.
- [80] A. Bevilacqua, L.D. Stefano, and P. Azzari. People tracking using a time-of-flight depth sensor. In *Proceedings of IEEE International Conference on Video and Signal Based Surveillance*, pages 89–95, November 2006.
- [81] A. Grassi, V. Frolov, and F. Leon. Information fusion to detect and classify pedestrians using invariant features. *Information Fusion*, In Press:1–9, 2010.
- [82] F. Tombari, L. Di Stefano, S. Mattoccia, and A. Zanetti. Graffiti detection using a time-of-flight camera. In *Proceedings of 10th International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 645–654, October 2008.
- [83] A. Bleiweiss and M. Werman. Fusing time-of-flight depth and color for real-time segmentation and tracking. In *Proceedings of DAGM 2009 Workshop on Dynamic 3D Imaging*, pages 58–69, September 2009.
- [84] L. Sabeti, E. Parvizi, and Q.M.J. Wu. Visual tracking using color cameras and time-of-flight range imaging sensors. *Journal of Multimedia*, 3(2):28–36, June 2008.
- [85] D.W. Hansen, R. Larsen, and F. Lauze. Improving face detection with ToF cameras. In *Proceedings of the IEEE International Symposium on Signals, Circuits & Systems*, pages 225–228, July 2007.
- [86] S. Meers and K. Ward. Head-pose tracking with a time-of-flight camera. In *Proceedings of Australian Conference on Robotics and Automation*, pages 1–7, December 2008.
- [87] P. Breuer, C. Eckes, and S. Müller. Hand gesture recognition with a novel infrared time-of-flight range camera - a pilot study. In *Proceedings of the Mirage 2007, 3rd International Conference on Computer vision/computer graphics collaboration techniques*, pages 247–260, March 2007.
- [88] A.D. Wilson. Using a depth camera as a touch sensor. In *Proceedings of ACM International Conference on Interactive Tabletops and Surfaces (ITS)*, pages 69–72, November 2010.
- [89] A. Leone, G. Diraco, C. Distanti, P. Siciliano, M. Malfatti, L. Gonzo, M. Grassi, A. Lombardi, G. Rescio, P. Malcovati, V. Libal, J. Huang, and G. Potamianos. A multi-sensor approach for people fall detection in home environment. In *Proceedings of ECCV Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*, pages 1–12, October 2008.

- [90] T. Schamm, M. Strand, T. Gump, R. Kohlhaas, J. M. Zollner, and R. Dillmann. Vision and ToF-based driving assistance for a personal transporter. In *Proceedings of 14th International Conference on Advanced Robotics*, pages 1–6, June 2009.
- [91] S. Vacek, T. Schamm, J. Schroder, and R. Dillmann. Collision avoidance for cognitive automobiles using a 3D PMD camera. In *Proceedings of 6th IFAC Symposium on Intelligent Autonomous Vehicles*, pages 1–6, September 2007.
- [92] A. Dorrington, C. Kelly, S. McClure, A. Payne, and M. Cree. Advantages of 3D time-of-flight range imaging cameras in machine vision applications. In *Proceedings of 16th Electronics New Zealand Conference (ENZCon)*, pages 95–99, November 2009.
- [93] D-Imager. *3D image sensor*. <http://panasonic-electric-works.net/D-IMager/>.
- [94] S. Verstockt, A. Vanoosthuysse, S. Van Hoecke, P. Lambert, and R. Van de Walle. Multi-sensor fire detection by fusing visual and non-visual flame features. In *Proceedings of International Conference on Image and Signal Processing*, pages 333–341, June 2010.
- [95] 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. *Fire Safety Journal*, 46(5):262–275, July 2011.
- [96] T. Celik and H. Demirel. Fire detection in video sequences using a generic color model. *Fire Safety Journal*, 44(2):147–158, May 2008.
- [97] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66, January 1979.
- [98] F. Shafait, D. Keysers, and T.M. Breuel. Efficient implementation of local adaptive thresholding techniques using integral images. In *Proceedings of SPIE Document Recognition and Retrieval*, January 2008.
- [99] J. Sauvola and M. Pietikainen. Adaptive document image binarization. *Pattern Recognition*, 33(2):225236, February 2000.
- [100] Brandweer Vereniging Vlaanderen. *Gebruik van thermische cameras bij brandbestrijding*. <http://www.brandweervlaanderen.be/>.
- [101] F. Amon, N. Bryner, A. Lock, and A. Hamins. NIST technical note 1499: Performance metrics for fire fighting thermal imaging cameras small and full-scale experiments. Technical report, National Institute of Standards and Technology, July 2008.
- [102] 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*, 47(10):1–13, October 2008.

- [103] F. Bashir and F. Porikli. Performance evaluation of object detection and tracking systems. In *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS2006)*, June 2006.
- [104] S. Verstockt, A. Vanoosthuysse, B. Merci, N. Tilley, B. Sette, C. Hollemeersch, P. Lambert, and R. Van de Walle. Performance evaluation framework for vision-based fire detection. In *Proceedings of 12th International conference on Fire Science and Engineering (Interflam 2010)*, pages 257–268, July 2010.
- [105] S. Verstockt, I. Kypraios, P. De Potter, C. Poppe, and R. Van de Walle. Wavelet-based multi-modal fire detection. In *Proceedings of 19th European Signal Processing Conference (EUSIPCO)*, September 2011.
- [106] S.G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 11(7):674–693, July 1989.
- [107] Optrima. *Optricam: 3D Time-of-Flight Camera Systems*. <http://www.optrima.com/>.
- [108] A.E. Cetin. *Bilkent University - Signal and image processing group*. <http://www.ee.bilkent.edu.tr/~signal/>.
- [109] FireSense project (FP7). *Protection of cultural heritage*. <http://www.firesense.eu/>.
- [110] M.I. Chacon-Murguia and F.J. Perez-Vargas. Thermal video analysis for fire detection using shape regularity and intensity saturation features. In *Lecture Notes in Computer Science: Mexican Conference on Pattern Recognition*, pages 118–126, June 2011.
- [111] J.R. Martinez-de Dios, L. Merino, and A. Ollero. Fire detection using autonomous aerial vehicles with infrared and visual cameras. In *Proceedings of 16th IFAC World Congress*, June 2005.
- [112] H.-M. Chen and P.K. Varshney. Automatic two-stage ir and mmw image registration algorithm for concealed weapons detection. *IEE Proceedings Vision, Image and Signal Processing*, 148(4):209–216, Augustus 2001.
- [113] J. Perez-Jacome and V. Madisetti. Target detection from coregistered visual-thermal-range images. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 2741–2744, April 1997.
- [114] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on medical imaging*, 16(2):187–198, April 1997.
- [115] H.-M. Chen, S. Lee, R.M. Rao, M.-A. Slamani, and P.K. Varshney. Imaging for concealed weapon detection. *IEEE Signal Processing Magazine*, 22(2):52–61, March 2005.

- [116] Y. Benezeth, P.M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger. Human detection with a multi-sensors stereovision system. In *Proceedings of International Conference on Image and Signal Processing*, pages 228–235, June 2010.
- [117] R. Hartley and A. Zisserman. *Epipolar geometry and the fundamental matrix, Multiple view geometry in computer vision (2nd ed.)*. Cambridge University Press, 2004.
- [118] S. Gould, P. Baumstarck, M. Quigley, A. Ng, and D. Koller. Integrating visual and range data for robotic object detection. In *Proceedings of the ECCV Workshop on Multi-camera and Multimodal Sensor Fusion (M2SFA2)*, pages 1–12, October 2008.
- [119] G. Wolberg and S. Zokai. Robust image registration using log-polar transform. In *Proceedings of 7th IEEE International Conference on Image Processing*, pages 493–496, September 2000.
- [120] H. Li, B.S. Manjunath, and S.K. Mitra. A contour-based approach to multisensor image registration. *IEEE Transactions on image processing*, 4(3):320–334, March 1995.
- [121] M. Shah and R. Kumar. *Video registration*. Kluwer Academic Publishers, Dordrecht(NL), 2003.
- [122] M. Irani and P. Anandan. Robust multi-sensor image alignment. In *Proceedings of 6th IEEE International Conference on Computer Vision*, pages 959–966, January 1998.
- [123] B. Zitova and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, October 2003.
- [124] L. Cinque, F. Di Renzo, G.L. Foresti, C. Micheloni, and G. Morrone. Multi-sensor registration for objects motion detection. In *Proceedings of the first ACM international workshop on analysis and retrieval of tracked events and motion in imagery streams (ARTEMIS'10)*, pages 87–92, October 2010.
- [125] Z. Hamici. Real-time pattern recognition using circular cross-correlation: A robot vision system. *International journal of Robotics and Automation*, 21(3):174–183, 2006.
- [126] A. Yamasaki, H. Takauji, S. Kaneko, T. Kanade, and H. Ohki. Denighting: Enhancement of nighttime images for a surveillance camera. In *Proceedings of 19th International Conference on Pattern Recognition (ICPR)*, pages 1–4, December 2008.
- [127] S. Verstockt, C. Poppe, S. Van Hoecke, C. Hollemeersch, B. Merci, B. Sette, P. Lambert, and R. Van de Walle. Silhouette-based multi-sensor smoke detection: coverage analysis of moving object silhouettes in thermal and visual registered images. Submitted to *Machine Vision and Applications*, 2011.

- [128] S. Verstockt, C. Poppe, P. De Potter, S. Van Hoecke, C. Hollemeersch, P. Lambert, and R. Van de Walle. Silhouette coverage analysis for multi-modal video surveillance. In *Proceedings of 29th Progress In Electromagnetics Research Symposium (PIERS)*, pages 1–5, March 2011.
- [129] S. Verstockt, P. Lambert, and R. Van de Walle. Feature extraction for localized cbr : what you click is what you get. In *Proceedings of 4th International Conference on Computer Vision Theory and Applications*, pages 373–376, February 2009.
- [130] V. I. Ponomarev and A. B. Pogrebniak. Image enhancement by homomorphic filters. In *Proceedings of SPIE Applications of Digital Image Processing XVIII*, volume 2564, pages 153–159, July 1995.
- [131] L. Lucchese and S.K. Mitra. Colour image segmentation: a state-of-the-art survey. *Proceedings of the Indian National Science Academy*, 67:207–221, 2001.
- [132] H.F. Ng. Automatic thresholding for defect detection. *Pattern Recognition Letters*, 27(14):1644–1649, October 2006.
- [133] M. Luo, Y.-F. Ma, and H.-J. Zhang. A spatial constrained k-means approach to image segmentation. In *Proceedings of Joint Conference of International Conference on Information, Communications and Signal Processing, and Pacific Rim Conference on Multimedia*, volume 2, pages 738–742, December 2003.
- [134] J. Zhang and J. Hu. Image segmentation based on 2D otsu method with histogram analysis. In *Proceedings of International conference on computer science and software engineering*, pages 105–108, December 2008.
- [135] T.N. Pappas. An adaptive clustering algorithm for image segmentation. *IEEE Transactions on signal processing*, 40(4):901–914, April 1992.
- [136] Z. Liu. *Investigations on Multi-Sensor Image System and Its Surveillance Applications*. PhD thesis, University of Ottawa, Ontario, Canada, 2007.
- [137] L. O’Gorman, M. J. Sammon, and M. Seul. *Binary Image Analysis: contour-based detection, Practical algorithms for image analysis (2nd ed)*. Cambridge University Press, New York, 2008.
- [138] N. Krishnamurthy. *Introduction to computer graphics: Coordinate systems*. Tata McGraw Hill publishing, New Dehli, 2006.
- [139] S. Carpin. Fast and accurate map merging for multi-robot systems. *Autonomous Robots*, 25(3):305–316, June 2008.
- [140] S. Weisberg. *Applied Linear Regression (3rd ed)*. John Wiley & Sons, New Jersey, 2005.
- [141] J. Li, L. Chen, and Y. Cai. Dynamic texture segmentation using fourier transform. *CCSE Modern Applied Science*, 3(9):29–36, September 2009.

- [142] V.J. Traver, M. Mirmehdi, X. Xie, and R. Montoliu. Fast dynamic texture detection. In *Proceedings of 11th European Conference on Computer Vision (ECCV)*, pages 680–693, September 2010.
- [143] B. Gouverneur, S. Verstockt, G. Gielis, S. Nemeth, T. Bocquet, J. Stynen, and J. Vermeiren. The development of a multiband system for early detection of wildlife fires and indoor search and rescue operations. In *Proceedings of SPIE defence, security and sensing*, April 2011.
- [144] T. Beji, S. Verstockt, R. Van de Walle, and B. Merci. Inverse zone modelling based numerical fire predictions using real time video and thermocouple information. In *Proceedings of 10th international IAFSS symposium*, June 2011.
- [145] J.R. Martinez-de Dios, J.C. Andre, J.C. Goncalves, B.Ch. Arrue, A. Ollero, and D.X. Viegas. Laboratory fire spread analysis using visual and infrared cameras. *International Journal of Wildland Fire*, 15(2):175–186, May 2006.
- [146] D.X. Viegas, M.G. Cruz, L.M. Ribeiro, A.J. Silva, A. Ollero, B. Arrue, R. Dios, F. Gomez-Rodriguez, L. Merino, A.I. Miranda, and P. Santos. Gestosa fire spread experiments. In *Proceedings of IV International Conference on Forest Fire Research 2002 Wildland and Fire Safety Summit*, pages 1–13, November 2002.
- [147] J. R. Martinez-de Dios, B. C. Arrue, A. Ollero, L. Merino, and F. Gomez-Rodriguez. Computer vision techniques for forest fire perception. *Image and Vision Computing*, 26(4):550–562, April 2008.
- [148] E. Pastor, A. Agueda, J. Andrade-Cetto, M. Munoz, Y. Perez, and E. Planas. Computing the rate of spread of linear flame fronts by thermal image processing. *Fire Safety Journal*, 41(8):569–579, November 2006.
- [149] M. Akhloufi and L. Rossi. Three-dimensional tracking for efficient fire fighting in complex situations. In *Proceedings of the SPIE Visual Information Processing XVIII*, volume 7341, pages 1–12, April 2009.
- [150] L. Rossi, T. Molinier, M. Akhloufi, Y. Tison, and A. Pieri. A 3D vision system for the measurement of the rate of spread and the height of fire fronts. *Measurement Science and Technology*, 21(10):1–12, August 2010.
- [151] C. Schmid, R. Mohr, and C. Bauckhage. Comparing and evaluating interest points. In *Proceedings of the 6th International Conference on Computer Vision*, pages 230–235, January 1998.
- [152] S.M. Khan and M. Shah. A multi-view approach to tracking people in crowded scenes using a planar homography constraint. In *Proceedings of 9th European Conference on Computer Vision*, pages 133–146, May 2006.
- [153] S. Park and M. Trivedi. Homography-based analysis of people and vehicle activities in crowded scenes. In *Proceedings of the IEEE Workshop on Applications of Computer Vision*, February 2007.

- [154] S. Verstockt, S. De Bruyne, C. Poppe, P. Lambert, and R. Van de Walle. Multi-view object localization in H.264/AVC compressed domain. In *Proceedings of the 6th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 370–374, September 2009.
- [155] D. Arsic, E. Hristov, N. Lehment, B. Hornler, B. Schuller, and G. Rigoll. Applying multi layer homography for multi camera person tracking. In *Proceedings of the 2nd ACM/IEEE International Conference on Distributed Smart Cameras*, pages 1–9, 2008.
- [156] J.-H. Yun and R.-H. Park. Self-calibration with two views using the scale-invariant feature transform. In *Proceedings of the 2nd International Symposium on Visual Computing*, pages 589–598, 2006.
- [157] S. M. Khan and M. Shah. Tracking multiple occluding people by localizing on multiple scene planes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(3):505–519, March 2009.
- [158] P.-L. Lai and A. Yalmaz. Efficient object shape recovery via slicing planes. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1–6, 2008.
- [159] E. Ribnick, S. Atev, O. Masoud, N. Papanikolopoulos, and R. Voyles. Real-time detection of camera tampering. In *Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2006.
- [160] J. Lerdsinmongkol, K. Chaisaowong, S. Roongruangsorakarn, T. Kraus, and T. Aach. Application of 3D morphological operations in the framework of a computer-assisted diagnosis system to construct thorax mask and remove trachea. In *Proceedings of the International Symposium on Biomedical Engineering*, pages 260–265, 2008.
- [161] V. Musoko and A. Prochazka. Non-linear median filtering of biomedical images. In *Proceedings of the 10th Scientific Conference MATLAB2002*, pages 1–5, November 2002.
- [162] NFPA. NFPA 92b: Standard for smoke management systems in malls, atria, and large areas. Technical report, National Fire Protection Association, 2005.
- [163] M. De Berg, O. Cheong, M. Van Kreveld, and M. Overmars. *Convex Hulls, Computational geometry: algorithms and applications (3rd ed.)*. Springer-Verlag, Berlin Heidelberg, 2008.
- [164] C.B. Barber, D.P. Dobkin, and H. Huhdanpaa. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, 22(4):469–483, 1996.
- [165] C. Xiaojun, Y. Lizhong, D. Zhihua, and F. Weicheng. A multi-layer zone model for predicting fire behavior in a fire room. *Fire Safety Journal*, 40(3):267–281, 2005.

- [166] J.W. Davis. Hierarchical motion history images for recognizing human motion. In *Proceedings of the IEEE Workshop on Detection and Recognition of Events in Video*, pages 39–46, 2001.
- [167] G. Rein, C.A. Empis, and R. Carvel. *The Dalmarnock Fire Tests: Experiments and Modelling*. Engineering & Electronics, University of Edinburgh, 2007.
- [168] Society of Fire Protection Engineers. *The SFPE handbook of Fire Protection Engineering*. National Fire Protection Association, Quincy (MA), 2002.
- [169] T. Beji, S. Verstockt, R. Van de Walle, and B. Merci. Submitted to Build and Environment. *Analysis of data assimilation based numerical simulations for prediction of smoke filling in large spaces*, June 2011.
- [170] S. Welch, A. Usmani, R. Upadhyay, D. Berry, S. Potter, and J.L. Torero. *Introduction to FireGrid, The Dalmarnock Fire Tests: Experiments and Modelling*. Engineering & Electronics, University of Edinburgh, 2007.
- [171] W. Jahn. *Inverse Modelling to Forecast Enclosure Fire Dynamics*. PhD thesis, University of Edinburgh, UK, 2010.
- [172] S.E. Dillon. *Analysis of the ISO 9705 room/corner test: simulations, correlations and heat flux measurements*. PhD thesis, University of Maryland, USA - NIST-GCR-98-756, 1998.
- [173] W. Jahn, G. Rein, and J.L. Torero. Forecasting fire growth using an inverse zone modeling approach. *Fire Safety Journal*, 46(3):81–88, April 2011.
- [174] N. Tilley, P. Rauwoens, and B. Merci. Verification of the accuracy of CFD simulations in small-scale tunnel and atrium fire configurations. *Fire Safety Journal*, 46(4):186–193, May 2011.
- [175] G. Heskestad. *Fire plumes, flame height, and air entrainment, SPFE handbook of fire protection (3rd ed.)*. NFPA and SPFE (ed:DiNenno,PJ), 2002.
- [176] T. Beji, S. Verstockt, R. Van de Walle, and B. Merci. Submitted to Fire Technology. *Numerical Prediction of Compartment Fires using Real Time Video Analysis*, August 2011.
- [177] Xtralis. *Open-Area Smoke Imaging Detection (OSID)*. <http://xtralis.com/p.cfm?s=22&p=459>.
- [178] S. Verstockt, G. Van Wallendael, S. De Bruyne, C. Poppe, P. Lambert, and R. Van de Walle. Cube: a multi-view localization framework for 3D object analysis in computer vision applications. In *Proceedings of Computer Graphics International 2010 (CGI)*, pages 1–4, June 2010.



Fireforum Award 2011 voor
het wetenschappelijk eindwerk
rond brandveiligheid



On November 17th 2011, this dissertation won the Fireforum Award 2011 for scientific research on fire safety. More info/photos can be found on the Fireforum website: <http://www.fireforumawards.be/>.



My warm thanks to all the research partners who participated in this study.

