Application of Physics-Based Image Formation Models to Change Detection in The Context of Indoor Workplace Video Surveillance

Mohamed Hamed Ismail Sedky

A Thesis Submitted in partial fulfilment of the Requirements of Staffordshire University For the Degree of

Doctor of Philosophy

Thesis to be kept confidential

Staffordshire University
Faculty of Computing, Engineering and Technology
Stafford
United Kingdom

June 2009

To The Soul of My Father ...

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Mansour Moniri for his expert guidance throughout the course of this dissertation. Since I joined his research group he has prove to be the best supervisor a student could ever hope for, with a passion for good, solid research work, persistence for clear understanding and expression of both the questions, the answers and a level of energy and optimism that never ceases to amaze me. I am sure his influence will serve me well in years to come, and wish I can have as rewarding a relationship as I had with him with all the people I work with in the future.

I would like also to express my deepest gratitude to Dr. Claude Chibelushi for his scientific support with my research process, feedback, and for his friendly concern during this process. I cannot thank him enough for the help, the encouragement, and these discussions that boosted my understanding of the problems every single time.

My deepest gratitude goes to my family, my parents, my sisters and brothers. Finally, to my wife Ange, for her patience, love, continuous support and encouragement, mainly during the absence time, without her it would have been extremely difficult to reach this stage.

Abstract

The aim of this thesis is to investigate the application of physics-based image formation models to change detection in the context of indoor workplace video surveillance.

First, video surveillance applications are reviewed. Based on this review, a new classification of video surveillance applications is proposed and indoor workplace surveillance is chosen as the target application. A new workplace surveillance model is then introduced, which relates the needs of workplace surveillance applications, their implications and the capabilities of video surveillance techniques. Furthermore, a set of requirements for workplace surveillance applications are elicited and a video-based workplace system structure is proposed.

Change detection is then reviewed, and the suitability of using physics-based image formation models to enhance change detection algorithms is investigated. Two physics-based change detection techniques are developed. The foundations of these techniques are advances of colour constancy techniques which extract physical features from the camera output; this approach is unlike other change detection algorithms which use the camera output directly without considering its physical meaning.

The performance of the proposed techniques is measured and compared against the Horprasert algorithm, using objective and computational complexity evaluation methods, where the quality of the change detection is measured using recall and precision measures. The Horprasert algorithm was shown, in an independent study by other researchers, to have the best trade-off between segmentation quality and computational complexity among other state-of-the-art algorithms such as Cavallaro, McKenna and Shen under experimental conditions which covered different lightings and background structures.

The first proposed algorithm is based on an approximated physical model, known as shading model. In this algorithm the illumination variation is modelled using ratios of RGB values of foreground and background pixels, coupled to the Mahalanobis distance to segment foreground objects. Compared to the Horprasert algorithm, the results show that the proposed method increases the recall by 8.85% on average, while the average decrease of the precision is 1.45%. The overall performance is improved by an average increase of 2.7%.

The second proposed algorithm uses a consistent colour descriptor model which

has a physical underpinning, where the full spectrum of surface spectral reflectance is estimated from the camera output; and a real-time correlation between foreground and background surface spectral reflectance is used to segment foreground objects from the static background. The results achieved show that the proposed method increases the recall by 7.45%, the precision by 2.5% on average, while the overall performance is improved by 7.85%, compared to the Horprasert algorithm. The second algorithm achieves more accurate change detection with fewer false alarms, compared to the Horprasert algorithm. This confirms the effectiveness of using the surface spectral reflectance as a cue for change detection.

The second algorithm needs from 3 to 8 background frames, to build its background model, in order to achieve its maximum performance. The first algorithm requires from 9 to 16 background frames, whereas the Horprasert algorithm requires between 9 to 63 frames. This means that the second algorithm can adapt faster to scene variations. This rapid adaptation underpins the robustness of this algorithm. The second algorithm processes one CIF frame in 40 ms on average, which is twice as fast as the Horprasert algorithm and the first algorithm.

The tolerance of an algorithm to threshold variation demonstrates its flexibility. The results show that the performance of the Horprasert algorithm for one sequence declines between -1.75% and -5.67% on average, when using a threshold optimised for a different sequence, and the performance of Algorithm 1 declines between -1.16% and -3.51% on average, however the performance of Algorithm 2 declines only between -0.11% and -0.24% on average. This shows that Algorithm 2 is more flexible and, once optimised for one environment, it can maintain its performance when applied to different environments.

In the context of indoor workplace video surveillance, the results of the comparison between the second algorithm and the other algorithms investigated in this thesis show that it matches better the operational requirements of a surveillance system, in terms of real-time performance and robustness to cope with changes in illumination, scene structure and scene activity, with little or no human intervention. This better match with operational requirements boosts the applicability of the algorithm in real-world scenarios, in line with the 3D workplace surveillance model introduced in this thesis, where technical capability and applicability are important considerations.

Contents

List of Figures

List of Tables

List of Special Symbols

C_f	Costs associated with false alarms
C_m	Costs associated with mis-detections
$E(\lambda)$	Illuminant spectral power distribution
$I(\lambda)$	Reflected light spectral power distribution
N_Q	Camera quantisation noise
N_R	Camera readout noise
N_S	Camera shot noise
$Q_B(\lambda)$	Sensor spectral sensitivity (Blue)
$Q_G(\lambda)$	Sensor spectral sensitivity (Green)
$Q_R(\lambda)$	Sensor spectral sensitivity (Red)
$S(\lambda)$	Object surface spectral reflectance
T	Illuminant colour temperature in ^oK
Ω	Visible wavelength from $400nm$ to $700nm$
λ	Wavelength in m
μ_{DC}	Camera dark current
g	Camera gain
w_d	Geometrical parameter for diffuse reflection
w_s	Geometrical parameter for specular reflection

List of Abbreviations

FN	**description**
ASSA	Active Smart Surveillance Architecture
BRDF	Bi-directional Reflectance Distribution Function
BSIA	British Security Industry Association
BSSA	Basic Smart Surveillance Architecture
CCD	Charge Coupled Device
CCT	Correlated Colour Temperature
CDR	Correct Detection Rate
CIE	Commission Internationale de l'Eclairage
CIF	Common Intermediate Format
CRT	Cathode Ray Tube
EM	Expectation-Maximization
FA	False Alarm
FAR	False Alarm Rate
FGT	Fast Gauss Transform
FP	False Positive

GEM	Generalized Exponential Model
HM	Harmonic Mean
HMM	Hidden Markov Models
ICA	Independent Component Analysis
KDE	Kernel Density Estimate
LCD	Log-Chromaticity Domain
LDD	Linear Dependence Detector
MD	Miss-Detection
ML	Maximum Likelihood
MoG	Mixture of Gaussians
P	Precision
PCA	Principal Component Analysis
PDR	Perturbation Detection Rate
PRF	Precision, Recall and F-measure
PTZ	Pan-Tilt-Zoom
R	Recall
ROC	Receiver Operating Characteristics
SPD	Spectral Power Distribution
SSR	Surface Spectral Reflectance
TN	True Negative
TP	True Positive

"The development of complex surveillance systems is capturing interest of both research and industrial world as there are strong requirements coming from the society in the direction of increasing safety and security in different applications."

C. Regazzoni, V. Ramesh, and G. Foresti [1]

1

Context and Scope of Work

1.1 Context

As one of the most important applications of image processing, understanding and computer vision, computer-based video surveillance has recently received significant attention, especially during the past decade.

Video surveillance can be defined as observation or analysis of a particular site for security and business purposes. A basic closed circuit television CCTV video surveillance system consists of a collection of video cameras, mounted in fixed positions, or on Pan-Tilt-Zoom (PTZ) devices. The video streams are transmitted to a central location (control room), cyclically displayed on one or several video monitors, and recorded. A human operator observes the video, to determine if there is an activity that warrants a response.

The need for analysing video information captured by CCTV cameras has grown and is becoming a crucial issue. This growth is accelerated by the wide range of commercial and law enforcement applications [1], and by advances in signal processing and computer vision techniques, along with the rising uptake of CCTV installations, and the increasing availability of cheap computational power which makes real-time systems for video processing feasible [2, 3].

1.1.1 Video surveillance

The video surveillance market in UK continues to expand, boosted by large government spending. The forecast growth of the market is expected to be boosted by the technological advances of algorithms and systems, combined with further take up of digital CCTV installations and networked solutions. One of the market segments of this expansion is the area of algorithmic components, which system integrators incorporate into their products [2].

With approximately 4.25 million CCTV cameras in the UK, according to the British Security Industry Association (BSIA) [4], only small fractions of the captured videos are ever watched. Some surveys, [5], figure out that camera to screen ratios is between 1:4 and 1:78, and the ratio of operators to screens can be as high as 1:16. For a relatively empty scene, an operator misses 45% of action after 10 minutes and after 22 minutes misses 95% of action. In a busy scene, an operator can see even less action and can virtually never detect a stationary or missing object. Practically, each operator can only monitor 1-4 screens at a time [6], so as few as 3% of cameras are likely to be actively monitored by an operator at any one time [5].

Given the high demand for a solution to the practical problem of insufficient human resources to monitor the ever increasing number of CCTV cameras, a number of video surveillance generations have been evolved. Video surveillance systems can be divided into three main generations, operator-controlled video surveillance, basic automated video surveillance, and smart video surveillance [3].

a. Operator-controlled video surveillance systems

This type is the first generation of video surveillance systems, which are based on analogue signal and image transmission and monitoring. In first-generation systems, visual information is processed entirely by human operators. The operator assigned to surveillance tasks has the role of:

- Performing the surveillance from the central control room by giving attention to the cyclic scan of all CCTV cameras.
- Concentrating its attention to specific cameras only, in case of a predefined, critical, emergency situation.
- Detecting critical situations and activating the prescribed procedures.
- Revealing and managing misbehaviour of anti-intrusion and detected emergencies, by alerting other departments.
- Coordinating between different services.
- Remotely configuring and controlling CCTV cameras.

There is a limit on the number of video signals that an operator can efficiently process, and the performance of a surveillance operator rapidly decreases with the time, as the operator gets tired. An increase in the number of cameras would also correspond to higher surveillance personnel costs. And the fact that the person, who is exposed to surveillance, gets the feeling that his privacy is violated, due to the fact that a human operator is watching him decreases the acceptance of video surveillance by public.

In practice, the video streams are barely monitored or not at all [7]; CCTV footage is often used merely as a record to be examined after an event is known to have occurred.

b. Basic automated video surveillance systems

Second generation, basic automated video surveillance systems attempt to reduce the load on the operator by employing video motion detectors to determine where

there is motion in a given scene. Simple change detection algorithms are used and programmed to generate alarms for a variety of predefined situations [8].

Second-generation systems use the available digital video signal transmission and processing techniques, trying to solve some of the drawbacks of first-generation systems. The common characteristics of this generation is the exploitation of advanced video processing techniques in order to automatically select the most probable part of the video sequence of interest for the surveillance task, to be finally examined by the surveillance operator. The main advantages of second generation systems are:

- Improving the work conditions of human operators, by requesting his/her attention only whenever there is a change in the scene.
- Decreasing the need of more human operators.
- Decreasing the transmission traffic in video surveillance networks, by sending only the relevant change whenever it occurs, instead of sending the full video signal all the time.

Due to the fact that a large amount of motion is not relevant so a more smart automated video surveillance system is required to actually extract the relevant information from specific motion [9].

c. Smart video surveillance systems

Smart video surveillance systems (SVSS) are part of the third generation surveillance systems. They achieve more than motion detection. The common functions of smart video surveillance systems are to detect, classify, track, localise and interpret activities of targets in a given scene [10].

Third generation surveillance systems take advantage of progress in low cost high performance processors and video communications, as well as the advance in video processing and computer vision techniques [11, 13, 15]. The global aim for the majority of these systems is to:

- Extract targets and events in real-time.

- Interpret interesting targets and activities.
- Present a report of such activities to the operator.
- Provide an appropriate automatic response when predefined activities occur.
- Apply efficient storage and rapid retrieval of video data and associated meta-data for later analysis.
- Visualise surveillance video scenes to aid operators to perform monitoring.

The main advantages of smart video surveillance systems are:

- They provide effective transmission of video (possibly via low bandwidth connections from remote sites), by sending an abstract (meta-data) or a report instead of the full video stream.
- They enable better handling of critical situations.
- They enable improved communication and coordination between services.
- They can increase the acceptance of the video surveillance service by the public.

Smart video surveillance systems, can be defined as distributed computer systems connected to a camera network which use image processing and computer vision techniques either for focusing the attention of a human operator or for automatic alarm generation [12]. Smart video surveillance system architectures are discussed in [13] and classified into three different architectures:

1. Basic Smart Surveillance Architecture (BSSA), where the cameras are wired to a central location.
2. Active Smart Surveillance Architecture (ASSA), where the automatic video analysis is used to automatically control cameras, to focus on detected activities or events of interest.
3. Distributed Smart Surveillance Architecture (DSSA), where smart cameras are used to minimize the cost of deployment. These cameras would typically

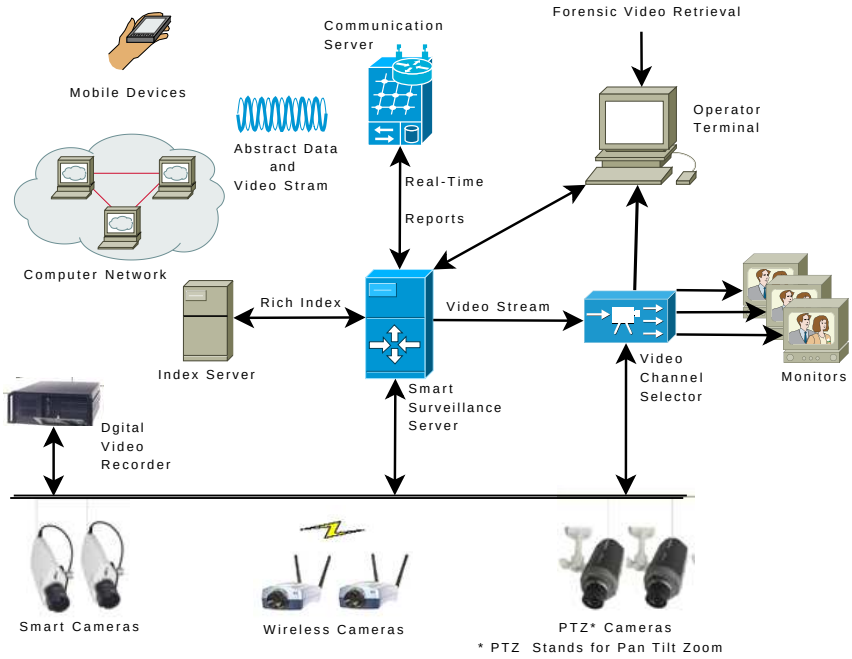


Figure 1.1: Smart video surveillance system architecture

use wireless communication to coordinate with a central camera coordinator. A smart camera would use automatic visual analysis to determine how to use the limited storage and bandwidth to effectively transmit only abstract data (meta-data) to the server.

Figure 1.1 shows a generic SVSS architecture, where the outputs of video cameras are recorded digitally and simultaneously analysed by the smart surveillance server, which produces real time alerts and a rich video index, controls an active camera server, accepts queries from a human operator and sends real-time report to a communication server.

Interpreting the activities of objects and their interactions in a real environment is a difficult task. The smart surveillance server uses video processing, and computer vision technologies in order to automatically segment moving objects (blobs), localise then track the segmented blobs, classify detected objects into semantic categories such as (human, human group, vehicle), identify detected objects if possible, track their positions, and automatically classify their movements, depending on a user query, and insert them into active scene visualisation, as shown in Figure 1.2.

Existing practical CCTV installations rely on human attention and labour.

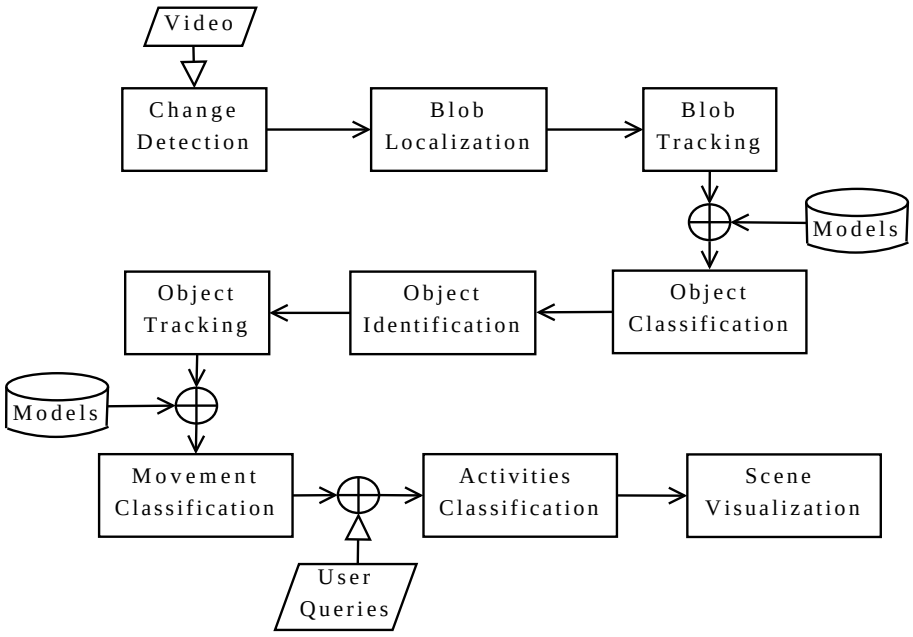


Figure 1.2: Internal structure of smart video surveillance server

Smart video surveillance systems are developed to act as intelligent video viewers to increase the efficiency of surveillance systems by providing real-time alerts and alarms when preset conditions (which represent activities) appear on a video feed. These immediate notifications can greatly improve the response times when critical events occur while reducing the manpower required for watching the monitors.

Smart video surveillance techniques were initially developed for military applications. They then transferred to commercial applications, for the purpose of improving either security or safety. The objective of most smart video surveillance systems was originally security [14], because of the easiness and simplicity that they provide especially when it comes to security surveillance. However in practice video surveillance systems are being used for other purposes. Unintended opportunities offered by video surveillance systems have been discovered; for example, video surveillance systems are also used to some extent, for improving workplace efficiency in terms of workflow, resource utilisation and handling critical situations.

There are practical needs for smart video surveillance techniques, which can integrate with the existing CCTV infrastructure in order to improve its efficiency and effectiveness. A survey of hundreds of US security executives found that systems

which could process the video streams, generated by the spiralling number of CCTV cameras, were among the top items in demand [1].

Smart video surveillance systems are based on video processing and understanding techniques for change detection, object classification, object tracking, movement classification, and activity classification.

Change detection aims to segment moving objects in the scene. It typically provides a focus of attention, for object classification or tracking. The segmentation of moving objects in video streams is not an easy task; a comprehensive modelling of all types of changes is required for reliable object segmentation.

Tracking refers to the association of segmented objects across the timeline. Object tracking adds temporal consistency to sequence analysis. While tracking objects which are moving, it is important to know the nature of the object of interest; object classification can be considered as a standard pattern recognition task. Objects can be classified into semantic categories. Movement classification and activity classification are largely a matter of applying a set of pre-determined spatio-temporal rules, often based on statistical learning techniques, which have been found to correlate to what humans, would interpret as situations of interest, corresponding to threats.

The research community has addressed, with various degrees of success, the technical problems associated with these stages. For example, some smart video surveillance techniques developed by researchers can analyze human motion or model human interactions. However, these techniques have not reached the maturity required for unconstrained real-world applications; their value to the video surveillance industry in the near- and mid-term is minimal.

There is a very broad spectrum of systems proposed in the literature which claim to be useful for smart video surveillance. Unfortunately very few of them are up to the challenges of a real world operating environment, where robustness, efficiency, and flexibility are essential. A particular deficiency of these systems is the inadequate performance of the front-end processing sub-system, which comprises elements such as object detection and segmentation.

[1] New Scientist, 12 July 2003, page 5.

Hence, the literature is preoccupied with a lot more mundane problems at the moment. Chief among these problems is the front-end processing, which comprises elements such as change detection and object segmentation. The reason for the preoccupation with reliable front-end processing is that failures at the segmentation level can propagate downstream of the surveillance processing chain, and result in incorrect outputs [16], which may lead to incorrect responses in critical situations, for example.

1.1.2 Change detection

Change detection module represents the core of an automated video surveillance system, while in a smart video surveillance system the performance of the change detection module affects the efficiency of later video surveillance processes. Although, a smart video surveillance system consists of a number of modules, only few of these modules represent the core of such system. Change detection, object classification and object tracking are the key modules. Change detection algorithms combined with conditional statements (e.g. size, shape) and tracking are used and programmed to generate alarms, for a variety of predefined situations.

Due to the high influence of change detection module effectiveness on the efficiency and robustness of later smart video surveillance modules and the overall system performance, the development of a robust change detection algorithm is a priority.

Change detection algorithms use computer computational power to analyze information associated with patterns of colour or movement contained in the video and associate pixel colour or movement with groups of pixels, which correspond to scene entities, such as objects. This process is called segmentation. The output of this phase is a binary mask, a set of connected pixels which correspond to moving object (foreground).

Proposed change detection techniques in the literature have limited capability to meet the accuracy required by most real-world applications; this limitation is partly due to sensitivity to variation in illumination and noise levels [17]. Robust surveillance systems should maintain adequate performance in the face of difficult

and fluctuating operating conditions, such as shadows and illumination variation. For outdoor scenes, illumination variation could be due to moving clouds, weather changes, or the time of day or night. Although indoor environments are less prone to changes in ambient lighting, indoor lighting changes could be caused by reflections and intermittent light-source occlusion due to a window or door opening, for example.

One of the key steps of change detection module is the selection of a suitable image representation. An important factor in change detection and object segmentation is the choice of the image representation and the transformation that is applied to the raw data in order to obtain the information that are relevant to the specific application domain.

Most of smart video surveillance systems use well-known colour spaces to represent the image, trying to model the variation in illumination and camera noise using statistical modelling techniques. These algorithms still suffer from illumination change and camera noise and the gap appears to be in the image representation used which does not give any consideration to the elements which govern the camera output. Using camera output directly without considering the illumination type, camera sensor characteristics, and the physical meaning behind this output, makes it more difficult to develop robust algorithms. This issue will be discussed in more detail later in Chapter 3.

Change detection approaches may be classified depending on the image representation used as physics-based or non-physics-based. Non-physics based change detection methods use one of the known colour spaces as a cue to model the scene. While the word physics refers to the extraction of intrinsic features about the materials contained in the scene based on an understanding of the underlying physics which govern the image formation. This process is achieved by applying physics-based image formation models which attempt to estimate or eliminate the illumination and/or the geometric parameters in order to extract information about the surface reflectance.

1.1.3 Physics-based image formation models

The colour appearance of a point on a surface depends both on the characteristics of the surface, the scene illumination and on the geometric arrangement of surface, illuminant, and imaging device. For many computer vision applications, the possibility of deriving colour models which are less sensitive to changes in the imaging parameters than raw camera outputs is highly desirable. When recording objects in different pose, illumination conditions, or from a different viewpoint, the segmentation accuracy degrades. To overcome this limitation, physics-based image formation models that are able to represent the effects of changes in imaging factors have been proposed by a number of researchers. Indexing on such models has shown to deliver better recognition results [18].

Conventional video cameras, analogous to a retina, sense reflected light so that colour values are the integration of the product of incident illumination power spectral distribution, object's surface spectral reflectance and camera sensors sensitivities. Humans have the ability to separate the illumination power spectral distribution from the surface spectral reflectance when judging object appearance, such ability is called colour constancy [19].

Researchers in the field of colour constancy [19–22] have done great efforts to extract new physics-based features from the image giving more care to the elements which govern the camera output and new models are introduced. This physical approach, used in colour constancy field, studies colour image formation and estimate or eliminate scene illumination in order to extract object's surface spectral reflectance, taking into consideration camera sensors sensitivities.

The adopted physics-based image formation models to denote the image representation in change detection algorithms as well as their links to different colour spaces used in existing approaches are explained in Chapter 4.

Smart video surveillance literature shows a limited work done toward the use of physics-based image representations; only simple approximated models, discussed in Section 4.3.1, are investigated and implemented [17]. The main gap, in the literature of smart video surveillance, is that the use of such physical models has not been fully investigated.

1.1.4 Applications of video surveillance systems

For years, research into video surveillance was confined to the military domain. The military applications of video surveillance systems include patrolling national borders, measuring the flow of refugees in troubled areas, monitoring peace treaties, and providing secure perimeters around military bases and embassies [23]. Recently, as the technology matured, the attention of researchers turned to practical applications, especially for public and personal security [24–27].

Companies suffer a great deal of damage from shoplifting and vandalism, and managers of retail organisations are increasingly experiencing similar problems. Much of today's practical video surveillance systems are about protecting public, private and business assets.

Smart video surveillance systems offer functionality which caters for crime prevention and surveillance; for example, they are used to ensure the safety of young, elderly and disabled people indoors, in care hostels, in nursing homes and in public places [28].

Video surveillance systems are also used to monitor vehicle traffic flow, detect accidents, monitor congestion, and give statistical reports. They are used in inspection applications, like monitoring of staff, compiling consumer demographics in shopping malls and amusement parks, monitoring livestock, performing engineering surveys, logging routine maintenance tasks at nuclear facilities, inspecting agriculture and counting endangered species.

In sports, smart video surveillance systems are used to improve the performance of players by monitoring and analysing their movements [29]. Tracking the positions of players, and ball, during a match is also often used for improving entertainment experience.

The majority of the literature associates video surveillance with security applications. While the objective of most of these systems was originally security, however in practice they are being used for other purposes. Video surveillance systems are being used daily, for improvement of efficiency of the workplace in terms of work-flow, management of resources, to handle critical situations and coordination of services and information [30]. Smart video surveillance systems

have the ability to contribute in much more practical applications rather than security applications. Where in this thesis the meaning of 'practical' is the consideration of the realistic operating conditions of video surveillance systems, actual requirements of their applications as well as the real capabilities of the techniques used.

One of the main issues related to the application side of video surveillance systems is the one stated by Hannah and Velastin in [5]

> "Researchers have developed a number of sub-systems and partial solutions which go some way towards solving elements of the problem of surveillance. Much progress has been made, but in a piecemeal fashion, and often without reference to the situations in which such systems might actually be used."

Many techniques and approaches which represent partial solutions to the video surveillance problem have been proposed. Although, much progress has been made, however the majority of such contributions do not refer to the application area as well as the practical requirements for such application.

It is important to specify the application area in which a specific module will be designed for, and to link the practical requirements of such application with the technologies proposed in order to be able to deliver practical solutions.

1.1.5 Workplace surveillance

Social scientists are highlighting the unexpected increase of surveillance systems and they are introducing the coming of the 'surveillance society' [31]. Employers argue that workplace surveillance is necessary to increase productivity and efficiency [32]. The 1998 American Management Association report [33] gave the three reasons for using surveillance in the workplace, 1. Performance evaluation, 2. Compliance with laws in regulated industries and 3. Safety and security [34].

A number of surveys in social science [31, 35, 36] state that video surveillance is a vital activity in the offices that are studied, and that both managers and employees recognise the need of monitoring. The second, but very important, conclusion of

these surveys is that the employee attitudes towards monitoring depend largely on the use of the monitoring process output.

There is practical and industrial need for smart video monitoring systems that can make use of the advances of video processing techniques and the existing video surveillance infrastructure, to assist in managing resources in the workplace.

The appeal of using smart video monitoring for workplace applications is for maximizing the utilisation efficiency and effectiveness of resources. Smart video surveillance systems are primarily software based and can offer continued improvement and additional capabilities over time with little or no additional investment in infrastructure [13]. However, workplace surveillance has a number of implications, there are conflicting rights and interests. There is therefore a need for better understanding of the capabilities of smart video surveillance algorithms versus the requirements and implications of workplace surveillance applications. Chapter 2 attempts to relate the requirements of workplace surveillance and its implications with practical smart video surveillance technologies, in order to propose a framework for smart video-based workplace surveillance.

1.2 Scope and objectives of the investigation

This thesis tackles the problem of change detection as a front-end module of a smart video surveillance system. This difficult problem can be formulated as one of representing objects in the scene and segmenting them from the background. The thesis investigates the use of physics-based image formation models in order to extract meaningful physical features suitable to represent the scene. Ultimately, the segmentation should result in a partition corresponding to semantic video objects, which correspond to meaningful things in the real-world. The definition of such semantic video objects depends on the application; it incorporates complex domain specific knowledge. The problem of segmentation is an ill-posed problem; no unique segmentation of a scene exists. In most situations, knowledge about the specific application in which the segmentation is to be used is crucial. The solution needs to deal with issues and challenges related to a specific context. General-purpose

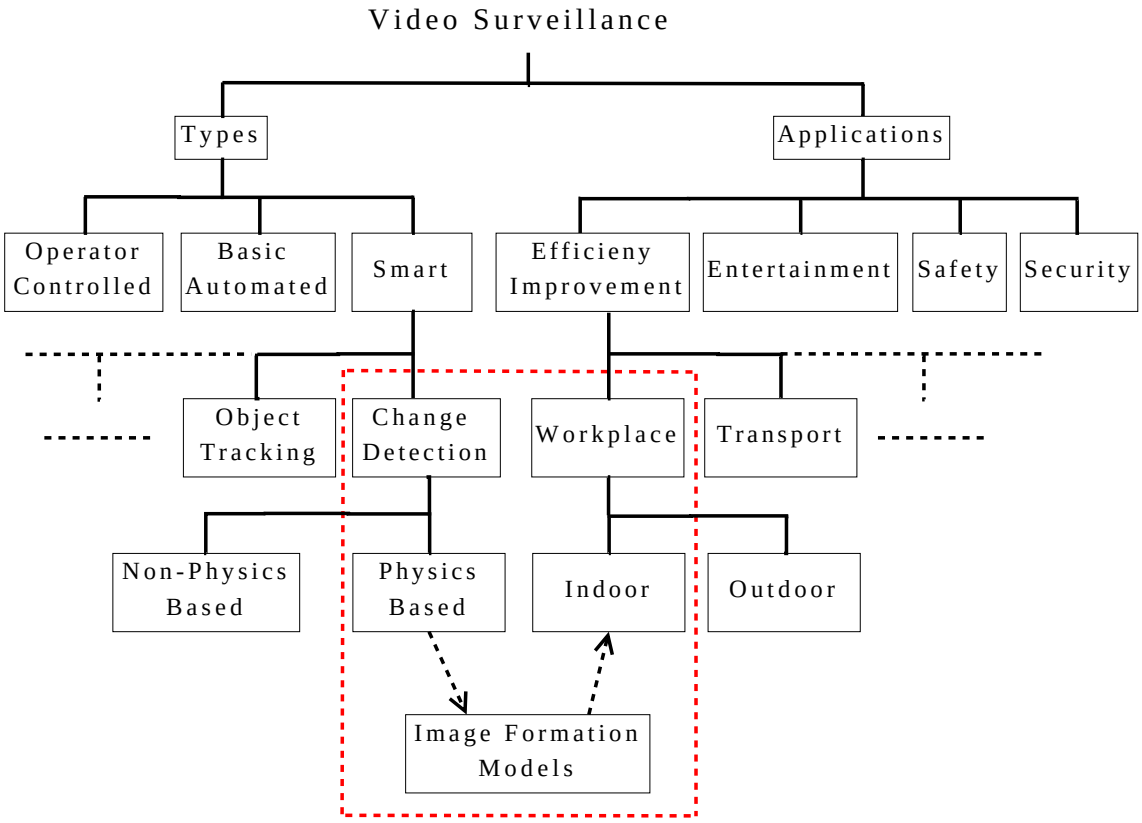


Figure 1.3: Scope of the work

segmentation of video objects is virtually impossible. Typical scenarios of workplace surveillance are confined indoor areas where the camera is near from moving objects. As Figure 1.3 shows, this work specifically investigates the use of physics-based change detection algorithms to fulfil the practical requirements of indoor workplace surveillance applications. The objectives of the investigation are as follows:-

- Study existing smart video surveillance applications.
- Identify promising practical applications, rather than security and safety, which can share the video surveillance infrastructure and make use of the advances in computer vision techniques.
- Study workplace surveillance needs and implications, as well as the link between such issues with smart video surveillance technologies.
- Carry out a detailed study of existing methods and techniques for change detection in smart video surveillance.

- Carry out a detailed study of existing physics-based image formation models and other colour spaces.
- Investigate the use of image formation models for the change detection problem and explore the possibility to introduce new models used in the colour constancy field to solve the change detection problem.
- Develop robust change detection algorithms to fulfil the proposed requirements.
- Evaluate the effectiveness of the proposed change detection algorithms.

1.3 Main contributions

The main contributions presented in this work can be divided into primary and secondary contributions, the primary contributions are as follows:

- A novel change detection algorithm using colour constancy techniques is proposed. The algorithm computationally estimates a consistent physics-based colour descriptor model of surface spectral reflectance from the camera output and then correlates the full-spectrum reflectance of the background and foreground pixels to segment the foreground from a static background.
- A new change detection algorithm, based on a shading model, is proposed. The algorithm uses the reflectance ratio to model the illumination variation and applies the Mahalanobis distance to model the correlation between this ratio triples in order to segment foreground and shadows from a static background.

The secondary contribution is:

- A new classification of practical applications for smart video surveillance is presented. A new workplace surveillance model is introduced which relates the requirements of workplace surveillance applications, and their implications in relation to the capabilities of smart video surveillance

technologies. A set of requirements for a workplace surveillance system using smart video surveillance techniques is proposed.

1.4 Organisation of the thesis

This thesis has the following structure:

Chapter 1 gives an overview of the existing video surveillance systems, their generations, techniques and applications. The objectives, as well as the contribution to knowledge of this research work are stated. Finally the chapter concludes with a brief overview of the chapters to follow.

Chapter 2 identifies the requirements of smart video surveillance practical applications. Architectures and techniques of smart video surveillance systems are introduced. A 3D workplace surveillance model is then proposed and discussed. Requirements for a workplace surveillance system and its structure are then proposed and discussed.

Chapter 3 reviews the state-of-the-art of change detection. The change detection problem is first defined, different factors that can produce apparent image change are then discussed. The main computational steps involved in change detection are introduced. The various physics-based and non-physics-based image representations are reviewed. Background modelling approaches that exploit this information for change detection are then described. Methods for threshold selection, noise removal and shadow detection are briefly explained. Subsequently, principles and methods for evaluating and comparing the performance of change detection algorithms are discussed. Finally, the survey findings are concluded.

Chapter 4 introduces fundamental physics-based image formation models and their relation with different colour spaces.

Chapter 5 proposes two novel change detection methods. The first is based on reflectance ratio. The second is based on surface spectral reflectance.

Chapter 6 presents a performance evaluation of the proposed change detection algorithms and one of the state-of-the-art algorithms which derives the same functionality. Three evaluation methods are adopted, illustrative visual

comparison, objective, and computational complexity evaluation.

Chapter 7 presents conclusions and recommendations for future work. It discusses the contributions of this research work, the practical results obtained, and the future work required for improving the proposed change detection algorithms.

"A useful source of information and guidance is often overlooked by academics creating automated video surveillance systems, and that is the experience of real-world closed circuit television (CCTV) installations and operatives."

M. Hannah Dee and A. Sergio Velastin [5]

2

Workplace Video Surveillance

2.1 Introduction

The previous chapter identified smart video surveillance as the latest generation of video surveillance systems capable of interpreting the activities of objects and their interactions in a real environment. In order to understand the capabilities of such systems, this chapter starts by reviewing their applications' requirements and the techniques used to fulfil these requirements. In the literature, smart video surveillance applications are mainly tied to security and safety applications. The fact that video surveillance systems are been used in practice for other purposes as well as the reviewed capabilities of such systems motivate us to classify smart video surveillance systems in terms of their practical applications. This classification leads to the introduction of workplace surveillance, as an example of

efficiency improvement applications. Subsequently, workplace application as a promising application is discussed along with its limitations. The perspective of the potential users of workplace surveillance systems is discussed. The workplace surveillance needs and implications are then studied and the practicality of using video-based surveillance techniques to address workplace surveillance applications is discussed. The chapter proceeds by introducing a model which relates the requirements of workplace application, its implications with practical smart video surveillance capabilities. The scope of workplace surveillance application is then presented. Requirements for a workplace surveillance system and its structure are then proposed and discussed.

2.2 Smart video surveillance systems

Demand for smart video surveillance is increasing; making selection of the appropriate surveillance system a difficult task, depending on an application's requirements. There is a very broad spectrum of systems that claim to be useful for smart video surveillance. Unfortunately very few of them are up to the challenges of a real world operating environment where robustness, efficiency, and flexibility are essential [30]. Understanding the requirements of smart surveillance systems is a key to selecting the right solution(s) for specific application(s).

2.2.1 Requirements of smart video surveillance

The global aim for video surveillance systems is to extract targets and events in a real-time, interpret targets and activities, which might be considered interesting, present a report of such activities to the operator, provide an appropriate automatic response when predefined activities occur, supply efficient storage and rapid retrieval of video data and associated meta-data, for later analysis. Requirements for smart video surveillance systems to achieve these aims could be divided into four main categories, which are segmentation, indexing, content analysis and video retrieval.

- **Segmentation :** To enable efficient video search, the surveillance system has to employ motion segmentation techniques to segment foreground objects from

the scene background in each frame and then analyse the segmented video to create a symbolic representation of the foreground objects and their movement.

- **Indexing :** After extracting objects of interest from a series of consecutive video frames, video indexing is used to provide access to the video sequences by annotating the meta-data, the area, shape, contour and semantic visual template of each object in every video frame. The system then tracks each object through successive video frames, estimating the velocity at each frame and determining the trajectory of the object and its intersection with the trajectories of other objects. An index mark is then placed to identify the activities of interest such as appearance, disappearance, place, removal, entrance, exit, motion and rest of these objects. This meta-data, symbolic record of video content, is stored in the database to be used for later indexing and retrieval.

- **Retrieval :** In order to analyse the recorded video, it is required to retrieve the scenes of interest from the video sequence. The surveillance indexing sub-system stores the metadata in a database for video retrieval through the user interface. The interface allows the operator to retrieve a video sequence of interest, play it forward or backward and stop on individual frames. The commonly used retrieval methods for video data are text based query, metadata based query and content based query. Surveillance videos are associated with the date and time of creation as a text metadata; the text information may be derived from automatic analysis of video. A text based search can be used as a first filter for a video retrieval process. As an example of metadata based query the operator may specify queries on a video sequence based upon object-based and activity-based parameters. In content based query, the query is based on abstract features of images, i.e. colour and texture are extracted in the indexing stage and compared in the search stage to find similar images. This search scheme also supports various queries based on object motion trajectory information.

- **Content Analysis :** The types of queries practical applications are interested in typically require the recognition of certain activities or objects in the scene, smart systems move toward understanding the content.

2.2.2 Techniques for smart video surveillance systems

Due to the structural and functional similarities between smart video surveillance servers discussed in the literature, the general decomposition proposed in [37] is used here. A smart video surveillance server consists of change detection, blob localisation, blob tracking, object classification, object identification, object tracking, movement classification, activity classification, and scene visualisation modules.

In order to understand each module, interactions between different modules and the used techniques applied to solve each task, we have captured the required input and output of each module and we have identified various features used. Figure 2.1, presents a comprehensive visualisation of the smart video surveillance process. This information has been captured by studying the video surveillance system literature and capturing their main features.

- **Video Acquisition :** A smart video surveillance system starts by acquiring a video stream from imaging sensors. A number of colour components, such as intensity, luminance, chromaticity, hue and saturation are extracted from the video stream and a suitable image representation is to be selected, as shown in Figure 2.1.

- **Change Detection :** This module aims to segment video stream into background (static) and foreground (moving) components [17], to provide a focus of attention for later stages, making these later processes more efficient since only foreground pixels need be considered. The output of this module is a binary mask, a set of connected pixels which correspond to foreground (moving objects). Chapter 3 presents an overview of change detection algorithms.

- **Blob Localisation :** This module has the role of determining the position

of the detected target within the 2D image; the real-world 3D position could be extracted by combining 2D positions from more than one camera. This module uses mainly spatial features extracted from the detected moving blob. As Figure 2.1 shows, the 2D position (centroid) of a detected blob, its contour, corners, and area are examples of spatial features.

- **Blob Tracking :** Once the 2D positions of blobs have been detected; the next step is to track these detected positions. Figure 2.1 shows that temporal features, features that are linked with an entire track and cannot be obtained from a single frame, are extracted from detected blobs and are tracked to create a trajectory for each blob. Direction, speed or trajectory of a moving blob are examples of temporal features. Foreground blobs detected from each new frame are associated with previously tracked blobs. Spatio-temporal features are combined to identify features that are unique on the blob and in order to associate the right blob to its previous tracking history. Also, tracking can incorporate validity checking to remove false positives from the detection phase.

- **Object Classification :** While tracking moving blobs, it is important to know the nature of the object of interest. Object classification can be considered as a standard pattern recognition issue. This module classifies tracked objects into semantic categories using a model database. Detected objects are categorised as dynamic objects, semi-static objects and static objects, as shown in Figure 2.1. Dynamic objects include human, parts of human body, human-groups, vehicles, where semi-static objects include man-made resources. Object classification is to assign a label to each potential target pixel. Ideally, all pixels that belong to the same object will share the same label.

- **Object Tracking :** Once objects have been classified; the next step is to track these detected classes. Object tracking adds temporal consistency to sequence analysis; otherwise, objects may appear and disappear in consecutive frames due to detection failure. By tracking multiple objects, detection of occlusion

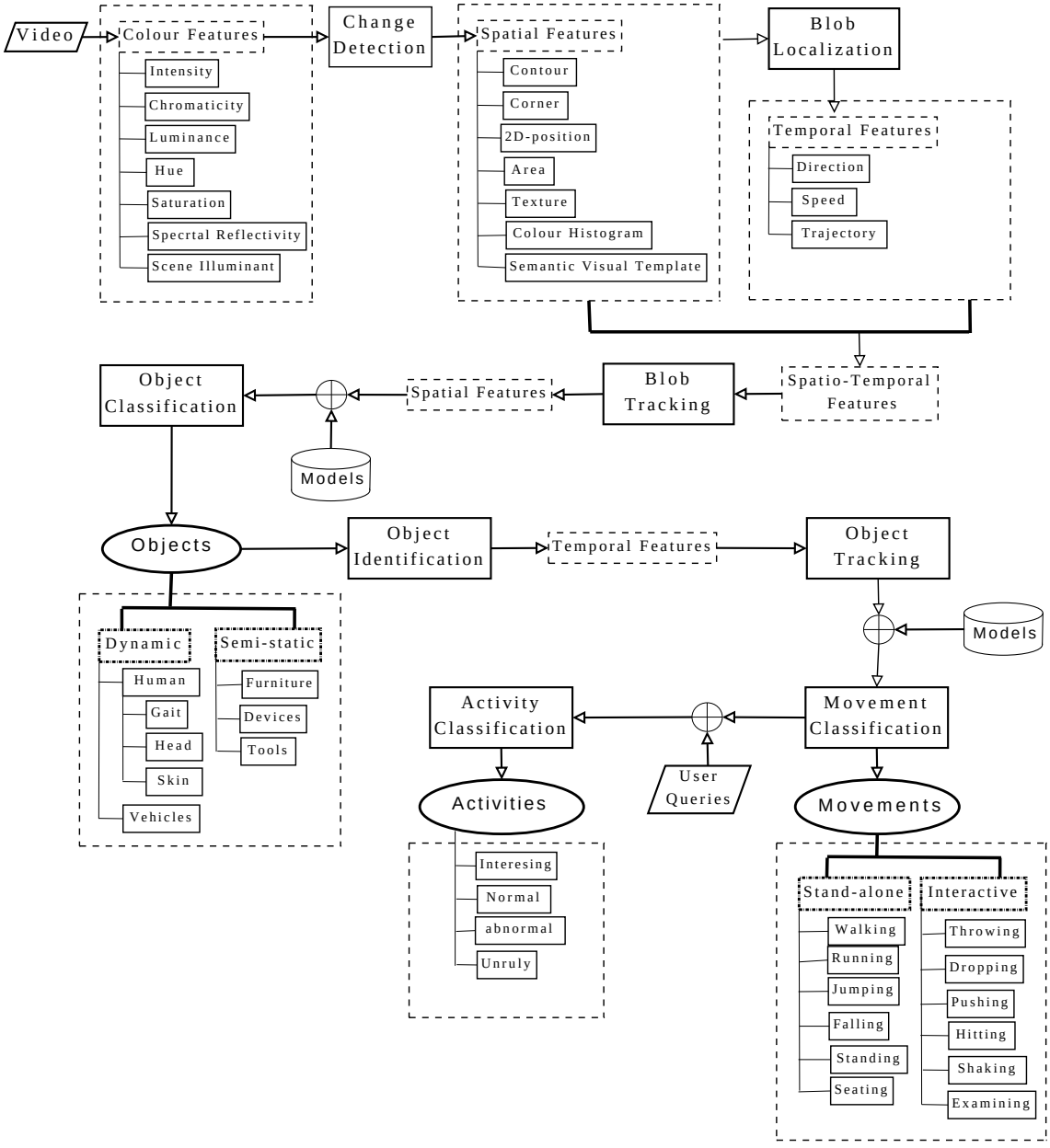


Figure 2.1: A generic model of a smart video surveillance server

is made easier, as occlusion is expected when two or more tracked object move past each other, giving the ability to track the exchange of objects between people or that a person is carrying a particular type of object.

- **Movement Classification :** This module attempts to understand the trajectories of tracked objects and the interactions between them. In this stage, the system may attempt to classify the consistent and predictable object motion [38]. In Figure 2.1 movements are classified into two categories, stand-alone or interactive, where stand-alone movements refer to the action of an individual object, while interactive movements refer to the interaction between two or more objects.

- **Activity Classification:** Statistical learning techniques are often utilised to classify between normal and abnormal activities, based on a priori information, and a user query, as shown in Figure 2.1. The overall aim is to produce a high-level, compact, natural language description of the scene activities.

- **Scene Visualisation :** This module describes the cooperative process of activity classification and the subsequent derivation of a representative visual presentation of the core of the content. Video visualisation uses video retrieval methods like text based query, content based query, and metadata based query.

2.3 Classification of smart video surveillance applications

The majority of the literature associates the video surveillance with security applications, the main reasons for this fact are:

- First generation surveillance systems (i.e. CCTV) were more suitable for security applications, where operator controlled CCTV systems and basic automated video surveillance systems are more suitable for security application requirements rather than other applications.

- The market was more interested in security and safety applications.

- Image processing techniques were not mature enough to fulfil other application requirements.
- The diversity of requirements in different applications.
- The missing link between applications needs and techniques, limitations and capabilities.

Different classification for video surveillance systems, have been proposed in the literature. It has been classified under a technological perspective as belonging to three successive generations [39], or according to the degree of awareness of observed people being monitored [13]. However, none of the published video surveillance systems reported classification of practical applications for smart video surveillance systems. In this thesis, smart video surveillance applications have been divided into four categories: security, safety, entertainment, and efficiency improvement applications.

- **Security Applications :** To alert security officers to a burglary in progress or to a suspicious individual loitering in the parking lot is an example of security improvement applications. Security is critical in public areas, and there is a growing need from the public for improved security in urban environments [1]. The purpose is to detect anomalous behaviours from a person or a group of people, in malls, lobbies, banks, shopping complexes, subways, highways, tunnels, parking areas, public areas and large facilities. The requirement of high speed reaction to dangerous situations and the need to implement a reliable system, explains why exiting commercial applications in the security domain are implemented using automated video surveillance rather than smart video surveillance. This is due to the fact that smart video surveillance techniques are still computationally expensive which limits their real-time capabilities.
- **Safety Applications :** Improved safety is one of the growing needs from the public, in-house, in transport and in public areas. Examples of safety improvement applications are detection of accidents on highways, and

provision of safe public areas. The requirements for safety applications are less than those for security applications, which gives smart video surveillance the ability to contribute more in the practical side in particular to help disabled people, the elderly or children, living in a domestic environment.

- **Entertainment Applications :** Tracking the positions of players and ball, during a match is used for improving entertainment experience, applied in augmenting digital TV, or low-bandwidth match play animations for Web or wireless display (i.e. mobile, PDA,..). Although, requirements of entertainment applications are higher than those of security applications, there are fewer life threatening situations and the market is mainly driven by communication and interactive games interfaces companies trying to use video surveillance to increase public use of their networks. The increasing use of mobile devices has opened a huge market for entertainment applications.

- **Efficiency Improvement Applications :**

  In addition to the previous applications, video surveillance technology has been proposed to improve efficiency of an activity or process. It has been used to measure traffic flow, monitor pedestrian congestion in public places, and compile consumer demographics in shopping malls and amusement parks, which are all considered as efficiency improvement applications. Performance improvement applications and resource utilisation improvement applications are the two main types of efficiency improvement applications. For example, human motion analysis is an important tool for improving the performance of players and patients in sports and medical applications. A system for tracking multiple people with multiple cameras in sports environment is used in efficiency improvement applications for the analysis of fitness and tactics of teams and players [40]. The resource utilisation improvement is concerned with managing different types of resources in the best way possible to attain the objectives of the business. Workplace surveillance is an example of efficiency improvement application. Video-based workplace surveillance has been chosen to be the proposed

application for the rest of this thesis.

2.4 Video-based workplace surveillance

A definition of workplace surveillance is given by [35] as

> "Any form of systematic management monitoring of individual staff members' job performance, with an eye to ensuring compliance with management expectation."

Workplace surveillance includes a wide range of issues: computer monitoring, such as Internet, e-mail and computer terminal keystrokes monitoring; video and audio surveillance; work site searches; access to medical records; information about personal lifestyle and off-duty conduct; drug testing; background checks, such as arrest and conviction records; insecure personnel records; polygraphs; and performance monitoring [32].

In this thesis workplace surveillance means, the use of a video-based data collection process to gather data about workers activities, job performance and resource utilisation, while the main objective is to manage different types of resources in the best possible way to attain the objectives of the business.

The application of smart video monitoring for workplace surveillance aims to improve the utilisation efficiency and effectiveness of resources. Smart video surveillance techniques can offer continued improvement and additional capabilities over time with little or no additional investment in infrastructure [13]. However, workplace surveillance has a number of implications, there are conflicting rights and interests. There is therefore a need for better understanding of the capabilities of smart video surveillance algorithms versus the requirements and implications of workplace surveillance applications.

It is found that monitoring has potential for control or feedback, and that the difference is in the way the system is introduced. Alison Cripps concludes in [41] that:

> "While advances in surveillance technologies are generally welcomed by employers, there is a growing sense of unease amongst employees,

> academics and interest groups as to the ethical boundaries that such technologies may cross. There is a concern that new technologies will leave employees open to abuse and discrimination and further, that workplace surveillance represents a threat to worker privacy and dignity."

For a workplace surveillance system to be effective, both people and technology must be taken into account during the development of the system [41]. A workplace surveillance system does not address the cultural issues associated with resource management directly. However, it facilitates the creation of an environment for managing resources, through a comprehensive video surveillance system that captures and stores accumulated workplace knowledge and make it more widely available [31].

There is therefore a need to have a better understanding of smart video surveillance algorithms capabilities versus requirements and implications of workplace surveillance applications.

2.4.1 Implications of workplace surveillance

The growth of new surveillance technology and practices has increased the potential for negative effects on the people subjected to it. Some of the negative feedback of video surveillance in the workplace can best be interpreted as acts of resistance [36], like: decreased work productivity, increased stress, lack of privacy and uncertainty. It is argued that workplace surveillance leads to perception that work quantity is more important than quality, which leads to increase performance in simple tasks and decrease performance in complex tasks [36]. These systems may also increase stress on employees in the workplace because of continuous monitoring of their behaviours, even if that is not objectively the case.

Employees may be concerned that the surveillance system is not giving their managers a complete picture about their performance, and that the employers might apply inappropriate values when judging this part of the picture.

Since, workplace surveillance is generally an employer initiative, it is important for the researcher to understand the factors that affect the monitoring activity: what

works and what does not work; and what is practically feasible by the available technology; and finally what is or can make an activity acceptable to employees.

Application specific requirements are a key to selecting the right solution(s) for the right application(s). In order to put the requirements of smart video-based workplace surveillance, it is important to select what to survey, what is feasible using the current technology, and most importantly, one has to select how to value what is found in surveillance data. It is also required to see what is accepted by employees and what is not accepted. In [36] it is concluded that "the way the system is used makes a big difference in employee attitudes towards computer monitoring". In the workplace, the acceptability of a system depends mainly on the manner in which surveillance technologies are employed, the information collected, the purpose for which it is used and the benefits which derive. If the employee is convinced that surveillance mechanisms are used incorrectly, intrusively or without justification, then this might affect the employee's trust and commitment to the business [42].

2.4.2 Workplace surveillance model

The quadrant model of workplace surveillance proposed in [31] consists of two dimensions as shown in Figure 2.2. The first dimension is the employers' opinion of surveillance effectiveness, depending upon whether a certain monitored activity can improve productivity. The second dimension is the acceptability of surveillance by employees.

As shown in Figure 2.2, quadrant 1 locates activities that work effectively, needed by employers, and are acceptable by employees. Some examples are by implementing features which improve employee's safety and security, or features which are used to establish outcomes (such as raises and promotions), or to be automatically identified to access specific locations. To use an overt surveillance system is effective as deterrent, and acceptable by employees. Quadrant 2 has activities which are socially acceptable but do not improve the work efficiency. Quadrant 3 contains activities that are not acceptable by employees and not required by employers, quadrant 2 and 3 will not be considered because there is no point in monitoring activities if productivity benefits are not likely to happen [31].

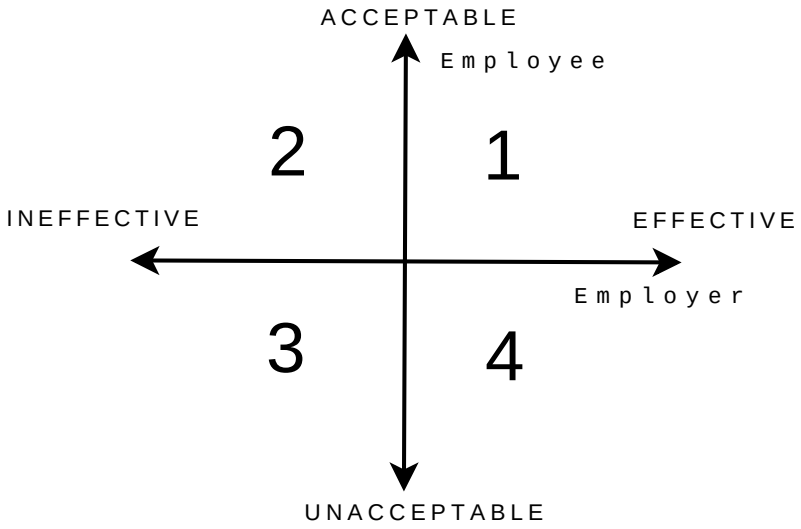


Figure 2.2: A quadrant model of workplace surveillance as proposed in [31]

The main conflict appears to be in quadrant 4, with activities that are important in increasing work efficiency, but considered not acceptable by employees. For example, it will not always be possible to convince employees to be automatically identified and tracked in order to evaluate their performance, or to accept being under covert surveillance. There are therefore contradictions between the employee's rights and employer's interests. Employers have legal interests in increasing work efficiency, while employees have rights to privacy. Solutions must be found to resolve these conflicts.

A broad examination of workplace surveillance must combine a number of different aspects, including the maturity of the used techniques. A theoretical analysis is required to differentiate between relevant employers requirements; implications for employees, trust, and capabilities of algorithms. An examination of the technology is required to clarify what can be done, how easily, and at what cost it can be done, and how invasive it will be.

Smart video surveillance system in the workplace should be considered as a tool of resource management, and the output of the system is just a factor which should be combined with a number of other factors, for the managers to get the complete picture.

This thesis proposes a 3D model workplace surveillance, which extends the 2D quadrant model proposed in [31] by adding a third dimension which is the

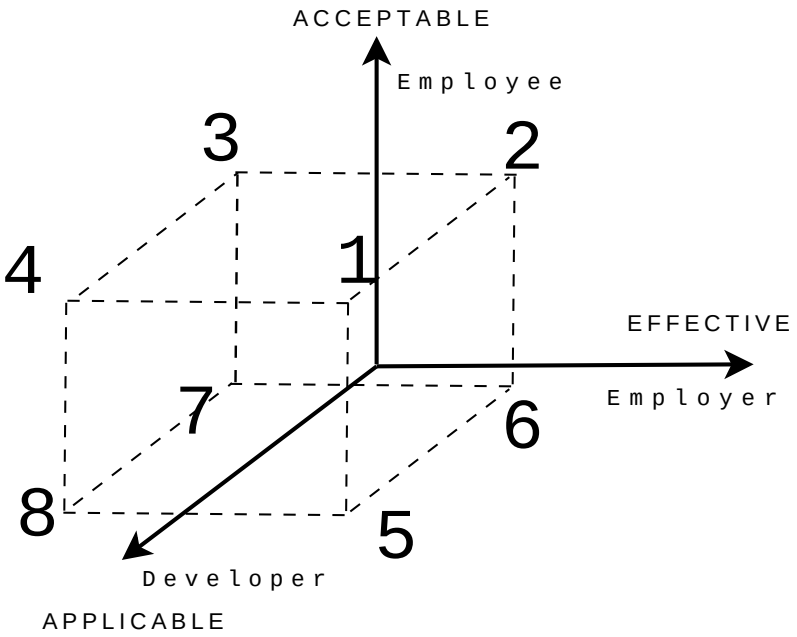


Figure 2.3: An octant model of workplace surveillance

applicability of surveillance using current smart video surveillance techniques.

As shown in Figure 2.3, octant 1 corresponds to activities that work, are socially acceptable and practically applicable. Examples of activities that belong to octant 1 are, tracking employees (without identification) and measuring the time, space and resource utilisation in the workplace, tracking and identifying valuable resources.

Octant 2 corresponds to activities which cause no offense, and contribute to increased productivity, but are practically not feasible. For example, measuring the amount of the work done is not feasible using the current technology for smart video surveillance.

Octant 3 and 4 are the areas where activities are acceptable by the employees, but they are ineffective in increasing productivity, for example, measuring how much time the employee is standing, or walking. In octant 4 where the activities are practically feasible, more efforts are needed to find some use of the information extracted from the system.

Some activities are practically feasible and effective for managing the work-flow but they are unacceptable to employees, like in octant 5. Hence, there should be a clear understanding and mutual agreement regarding the reasons for and objectives of monitoring these activities. Data collection, collation and utilisation should be

transparent at all times, the process must be perceived to be 'fair' with a legitimate purpose. Its design and development should include representation and input from those meant to be monitored. The design of the system must allow individuals to challenge any incorrect 'facts' and to have the appropriate corrections made.

Tracking identified employees is not feasible practically and not acceptable, while it is effective, this example of activities belongs to octant 6.

Finally octants 7 and 8 contain activities that are not favoured by either employers or employees, and there is no point in monitoring activities if productivity benefits are not likely to happen.

2.4.3 Scope of workplace surveillance

The main uses of smart video surveillance systems in workplace applications are, efficient utilisation of resources (e.g. human, time, natural, material,...), coordination of services, handling critical situations, communication and training. Those represent the ones discussed in Octant 1.

To achieve the above features and at the same time for the system to be acceptable by employees, this thesis proposes that a video-based workplace surveillance system must be an overt and smart system that can classify, identify and track all valuable material resources.

Due to the specific nature of workplace application, where the monitored scene contains valuable material resources owned by the employer, it is possible to tag all valuable resources which make the object classification, identification and tracking practical and feasible.

The scope of this work is on indoor workplaces, where the moving blob is classified directly as a moving person, a snapshot of the face is to be taken, using a skin detection algorithm, and stored securely, only authorised persons (e.g. manager, or director) can access these snapshots to be identified for communication, to handle a critical situation or for training purposes and after an agreement between the employer and the employee is established. Human' identification is done manually by the authorised person; otherwise employees are not identified, and so their privacy is not invaded.

The output of the proposed workplace surveillance system is the location of all valuable resources, in the workplace; tracking the movement of resources; detecting human activities in the workplace, for example walking, talking; detecting human-human interactions; and detecting the interaction between humans and other resources.

2.4.4 Requirements for workplace surveillance systems

Efficiency improvement in the workplace refers to the directing, organizing and controlling of different resource utilisation activities to ensure that:

- The material resources are utilised effectively.
- The schedules are followed to ensure accomplishment of activities on time.
- Individual employee work performance is monitored.
- Customer service is improved.
- Employees comply with legal obligations.
- Health and safety standards are enhanced.
- Staff training is assisted.
- Production processes are monitored.

A smart video surveillance system could be effectively applied to the following tasks

- Automatic detection, identification and tracking of pre-tagged resources.
- A snapshot of the face is to be taken, to be manually identified for communication, to handle a critical situation or for training purposes and after an agreement between the employer and the employee is established.
- Automatic people tracking and counting in each room to improve the utilisation of the space and for managing critical situations (e.g. fire).
- Automatic people counting in a waiting room to improve the comfort of customers or (guests).

- Automatic detection of persons in unauthorised areas.
- Automatic detection of not moving people for more than a certain period of time to improve safety.
- Automatic detection of human-object interaction to improve the utilisation of different resources.
- Access control for authorised people in certain areas
- Detecting human activities in the workplace, for example walking, standing; by attaching these behaviours to each person, identified by his/her tag, after his/her agreement, to be used in for communication or for training purposes.

The operational requirements for smart video-based workplace surveillance system depend to some extent on the specific context and functional requirements. A workplace surveillance system should be robust, able to operate for extended periods of time, possibly weeks or months, requiring little or no human intervention. The system should be extensible and adaptable.

Adaptability would enable accommodating changes in the workplace environment, to cope with changes in lighting, scene geometry and scene activity. The system should operate in real-time and provide continuous and accurate natural language annotations of scene activities.

Tagged objects should be detected, identified, tracked. Humans and group of humans should be detected and their behaviours should be recognised depending on the application being considered. The proposed workplace surveillance system should have the ability to take real-time decisions. The annotated video generated should be efficiently stored and allow a smart querying for later retrieval by the appropriate authorities.

2.4.5 Proposed workplace surveillance structure

Figure 2.4 presents the proposed smart video-based workplace surveillance structure, the system uses models of tags and keeps detecting the existence of any of these tags.

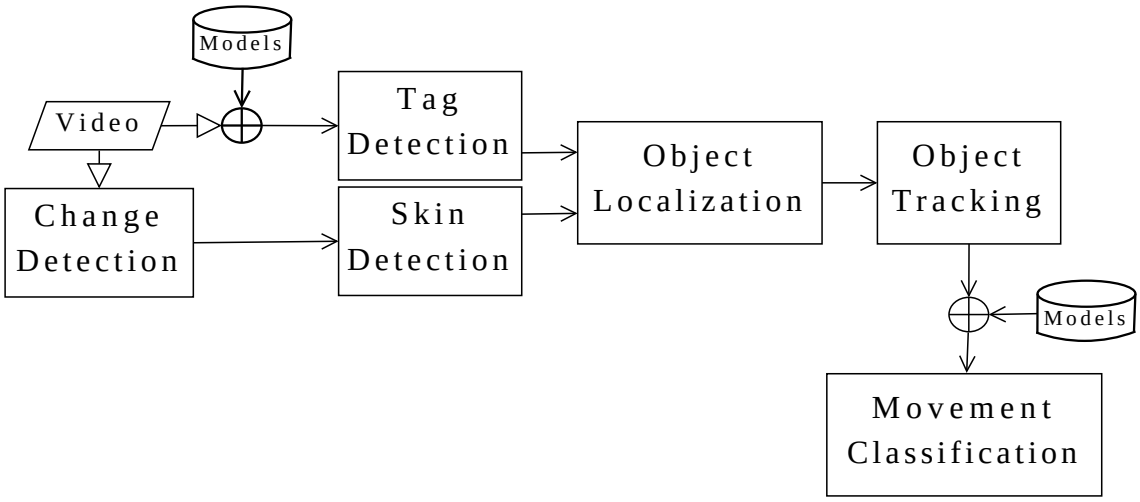


Figure 2.4: Smart video-based workplace surveillance proposed structure within the constraints limitations of current state of the technology

A complete identification of the resource is captured from its tag, dimension, colour, to whom it belongs and where should it be.

Moving objects are assumed to be human, a change detection module segments foreground in order to detect human existence. Skin detection module is required to detect skin and to capture a snapshot of human faces. The positions of the detected resource and moving objects are extracted from the 2D image using an object localisation module and then resources are tracked using an object tracking module. Where the requirement of workplace surveillance is to detect normal basic movements, a simple movement classification module is required.

The main challenge of this proposed structure is the development of a robust change detection module, which can deal effectively with different indoor workplace environments, where the surveillance task is done within confined areas, and the camera is near from moving objects. The main focus of this thesis is to develop the front end module of a video-based workplace surveillance system which is the change detection module. The change detection technique needs to address the practical requirements of the workplace surveillance system.

The main operational requirements for a video-based workplace surveillance system include real-time operation, flexibility in operation with minimum settings, adaptability to scene variations as well as being robust and reliable. The failure of such systems reflects weakness in their front end modules, mainly the change

detection module.

2.5 Summary

In this chapter, the requirements of commercial applications using smart video surveillance are identified as segmentation, indexing, content analysis and video retrieval. Smart video surveillance techniques including change detection, blob localisation, blob tracking, object classification, object identification, object tracking, movement classification, activity classification, and scene visualisation have been introduced. Different inputs and outputs of each smart video surveillance system component have been discussed and various features have been identified. An overall visualisation of the smart video surveillance process has been presented.

A new classification of smart video surveillance for commercial applications is then proposed. Workplace surveillance presents a very promising application, which can make use of the existing techniques in smart video surveillance to improve the efficiency of workplace. Importance and implication of workplace video surveillance have been discussed.

A new 3D workplace surveillance model has been introduced and discussed. This model links capabilities of smart video surveillance algorithms versus requirements and implications of workplace surveillance applications in order to conclude feasible, useful and acceptable requirements for smart video-based workplace surveillance. The scope of such application has been discussed and a set of requirements has been proposed. Finally, a video-based workplace surveillance structure has been proposed to fulfil the proposed requirements.

A workplace surveillance system should operate in real-time and should be robust, extensible and adaptable, able to accommodate changes in the workplace environment, to cope with changes in lighting, scene geometry and scene activity. The majority of the workplace scenarios cover indoor environments, where the surveillance area is confined, video sensors are assumed to be near from moving objects and therefore appearance and disappearance of objects affects the scene

illumination. The change detection module is responsible for eliminating illumination variation and camera noise from changes caused by meaningful objects, which shows the robustness of the system. Other modules rely on the effectiveness of this module. The thesis will focus on the development of a robust change detection technique to fulfil the operational requirements of the indoor workplace surveillance application.

"If the technology takes off it could put an end to a longstanding problem that has dogged CCTV almost from the beginning. It is simple: there are too many cameras and too few pairs of eyes to keep track of them. With more than a million CCTV cameras in the UK alone, they are becoming increasingly difficult to manage."

J. Hogan [43]

3

Change Detection-Review

3.1 Introduction

Robust change detection techniques with high detection probabilities and low-false alarm rates is subject of extensive research [23, 44]. Developments of such techniques have become important in the past two decades. This is due to the fact that it represents a key computer vision component which contributes in a wide spectrum of applications in diverse disciplines.

Change detection has been subject of many studies in a number of application areas including video analytics [23], medical diagnosis [45], industrial inspection [46], remote sensing [1], object-oriented video coding [47] and interactive gaming [48].

The task of change detection is to extract pixels which are changing due to motion of objects in image sequences of a dynamic scene and labelling such pixels

which belong to the moving objects (foreground) or regions from the rest of the scene (background).

The output of these techniques is a foreground mask, a set of connected pixels which correspond to moving objects (foreground). The mask provided by such techniques can then be considered as a valuable low-level visual cue to perform high-level object analysis tasks such as object classification, tracking, and activity classification [17, 38, 49–51].

The scope of the change detection problem discussed in this thesis falls to moving object segmentation in image sequences captured by a static camera or motion compensated camera and taken at different times. Where, in this thesis, 'change' refer to changes due to the motion of objects (either a new object or a background object which starts to move). Although change detection has been studied for several decades, it still remains a difficult problem to automatically and accurately segment moving objects under various types of illuminations from video sequences [17].

In relation to the main concern in this work, namely, the change detection for indoor workplace surveillance purposes, discussed in the previous chapter, it must be said that change detection is an essential and critical component of such video surveillance system. Besides, it is one of the most difficult tasks in image processing, it is very difficult to correct errors made at this abstraction level [52].

This chapter presents a review of current advances within the field of change detection. The chapter begins by formally defining the change detection problem, illustrating different factors that can produce apparent image change and discussing the main computational steps involved in change detection. In order to achieve an image representation which is useful for change detection, the various physics-based and non-physics-based image representations are reviewed. Background modelling approaches that exploit this information for change detection are then described. Methods for threshold selection, noise removal and shadow detection are briefly explained. Subsequently, principles and methods for evaluating and comparing the performance of change detection algorithms are discussed. Finally, the survey findings are concluded.

3.2 Change detection problem

Appearances of scenes depend on different factors such as sensor properties, illuminant spectrum, material properties and viewing geometry. Most often, these parameters combine non-linearly to yield an image [53]. An image is the product of the illumination, reflectance and shape in a scene. For foreground segmentation task, only a subset of this information is required, and this piece of information should be extracted in a manner that is invariant to variations in the remaining scene properties.

Apart from the motion of objects relative to the background or disappearance of background object, the foreground mask may result from a combination of different physical problems, being a consequence of the image formation, e.g. sensor noise, illumination variation, shadows, reflections, in addition to camouflage foreground, similar background-foreground coloured objects, which result in two types of errors, false alarms and missed detections.

1. False alarms: Pixels that actually belong to the background, but which are classified as foreground, because of sensor noise, illumination variation, shadows or reflections.

2. Missed detections: Pixels that belong to the foreground object, but which are classified as background, because they are occluded by background objects or their colour is very similar to the background at the same position.

In [54] change detection methods are classified into two groups. One group uses pixel-based methods and the other uses region-based methods. Region-based algorithms usually divide an image into blocks and calculate block-specific features; change detection is achieved via block matching. Pixel-based approaches have a lower computational cost than region-based methods because they compute the values for individual pixels at each instant. However, they are very sensitive to image noise, and cannot discriminate small changes in grey level, because of limited quantization levels. Region-based approaches work well against image noise, but they cannot show robust results when the illumination conditions change [55].

This review covers pixel-based approaches due to the special problem of indoor change detection and the importance of estimating or eliminating local illumination variations, the output is then post-processed to enforce smooth regions in the change mask.

In the literature, the problem of change detection is discussed, identifying three different kinds of approaches: optical flow [38, 85]; temporal differencing [56]; and background modelling, commonly known as background subtraction [51, 57–66] or a hybrid approach which combines a number of approaches.

For surveillance applications, approaches based on background modelling are most promising [50]. And as it is concluded in [52], most surveillance cameras are static or have small temporal motion that is corrected by using motion compensation, so a computationally complex optical flow approach is not necessary. Temporal differencing is computationally attractive, but it cannot handle homogeneously coloured objects.

Background modelling main idea is to automatically generate and maintain a representation of the background, from a number of static background frames that is then used to classify any new observation as background or foreground. A detailed modelling of all type of changes is required to segment relevant changes for a given application.

3.3 Architecture of change detection systems

The architecture of change detection approaches based on background modelling typically consists of two cascaded modules, background modelling and post-processing, a feedback module, background maintenance (see Figure 3.1).

The background modelling module is the core of the change detection algorithm, which aims to model different local illumination variation and background periodic dynamics such as waving trees or ocean waves. A post-processing stage is crucial to filter out irrelevant changes, such as camera noise, holes in foreground objects and shadows before making the change detection decision. A background maintenance stage is required, to update the background model with global illumination variation,

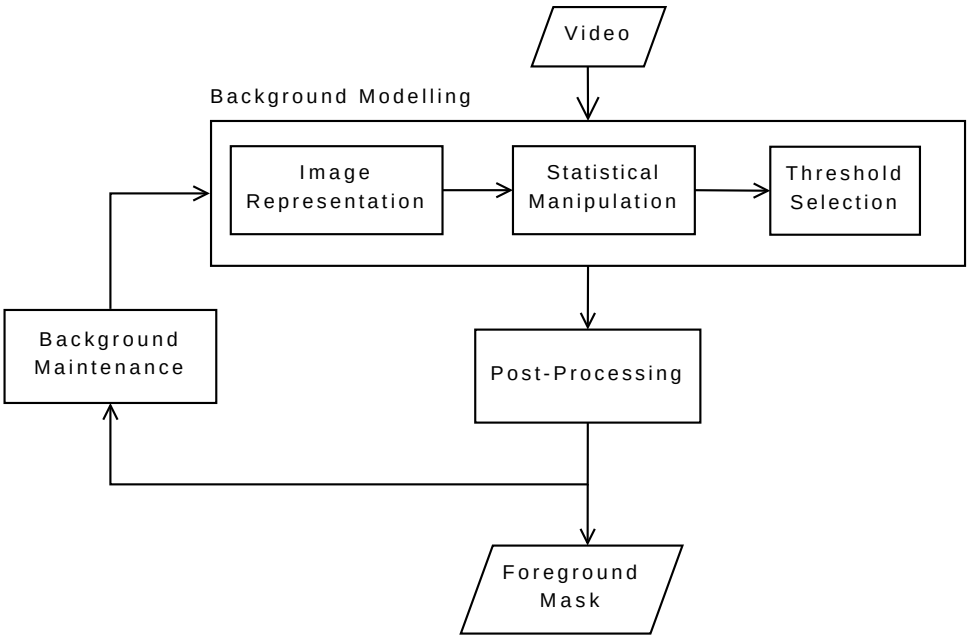


Figure 3.1: Change detection block diagram

a foreground object which stops and becomes part of the background or moving background object.

This survey decomposes background modelling module into three major steps, image representation, statistical manipulation and threshold selection, similar to the decomposition proposed in [48]. The first step comprises the choice of the image representation feature(s), where an ideal representation is the one able to convert the camera output into information which describes the material and is invariant to the illumination. The second step aims to statistically manipulate the chosen feature(s) to discount for background periodic motion and illumination variation. The third step aims to define the decision rule which will be implemented to segment foreground objects from the background.

In this chapter, existing methods for change detection based on background modelling are classified as either non-physics-based or physics-based depending on the chosen type of image representation. Non-physics based methods use one of the known colour spaces as a cue to model the scene. While the word physics refers to the extraction of intrinsic features about the materials contained in the scene based on an understanding of the underlying physics which govern the image formation. This process is achieved by applying image formation models which attempt to

estimate or eliminate the illumination and/or the geometric parameters in order to extract information about the surface surface spectral reflectance.

The rest of the chapter is completely dedicated to the review of most important algorithms for pixel-based background modelling change detection. It is nearly impossible to inform here about all the literature in this area. A number of surveys have been conducted by other researchers in this area [17, 38, 50, 51], which have been of great help.

3.4 Image representation

An important factor in change detection and foreground-background segmentation is the choice of the transformation that is applied to the raw data in order to obtain the information that are relevant to the specific application domain.

The basic idea behind image representation is how to select the photometric invariant features which are the best invariance to the background changes while maintaining a high detection rate for the foreground objects. For example, in indoor scenes, the classification must be invariant to a shadow, reflection and change of illumination that might occur due to light from a nearby light source. Similarly, in outdoor scenes, periodic background dynamics such as waving trees or ocean waves, in-variance to such periodic motion is critical.

Several image representations have been proposed, which can be classified into three types: spectral, spatial, and temporal features. Spectral features could be associated with greyscale (intensity, brightnes, luminance) [11, 61, 67, 68], raw colour [58, 69–71], normalised colour [71–75], colour information (Saturation, Hue, chromaticity)[73, 76–78, 111, 112], intrinsic image [79] and reflectance ratio [80, 81]. Some spatial features are also exploited such as spatial gradients [66, 71, 74, 82, 83], texture [65, 84] and optical flow [85], and temporal features could be associated with temporal gradients at the pixel [68, 86], interframe changes.

A summary of different features used in state-of-the-art change detection techniques, their advantages and disadvantages are presented in Table 3.1.

Table 3.1: A summary of different features used in state-of-the-art change detection techniques

Features	**References**	**Advantages**	**Disadvantages**
Spectral features	[11, 61, 68] [58, 69, 75] [70, 71, 73]	Crucial to discriminate a foreground object from its surrounding background.	Sensitive to camera noise. Cannot be computed accurately in regions that have low intensities.
Spatial features	[66, 71, 74] [65, 83, 84]	Can obtain invariance to local illumination and camera noise.	Not very suitable for a region that has low spatial gradients.
Temporal features	[11, 80, 85] [68, 86]	Useful in handling dynamic scenes. Can eliminate transient environmental noise such as rain and snow.	Cannot be computed accurately in regions that have low texture.

Spectral features are crucial to discriminate a foreground object from its surrounding background; however such features are sensitive to camera noise and cannot be computed accurately in regions that have low intensities. Spatial features can obtain invariance to local illumination and camera noise but not very suitable for a region that has low spatial gradients. General robustness to lighting changes can be achieved using edge-based or block-correlation methods. However, the foreground segmentation achieved by these methods is not precise, as the foreground is either made of blocks or of edges and not of accurate regions [87]. The temporal features suffers from the same problem, where it can not be calculated accurately in regions that have low textures, however it is useful in handling dynamic scenes and camera motion. Spatial and temporal smoothing are often used to reduce camera noise and to eliminate transient environmental noise such as rain and snow [50].

In most background modelling algorithms, the features are chosen arbitrarily and the same features are used globally over the whole scene. Recently, a number of authors start to propose frameworks for selecting suitable features for different parts of the scene [62, 64].

A Bayesian framework that incorporates spectral, spatial, and temporal features to characterize the background appearance is proposed by Li et al. in [62]. Principal feature representations for both the static and dynamic background pixels are investigated. Under this framework, the background is represented by

the most significant and frequent features, i.e. the principal features, at each pixel. A Bayes decision rule is derived for background and foreground classification based on the statistics of principal features.

Another framework for feature selection for background modelling has been introduced recently by Parag et al. [64]. The use of a boosting algorithm, namely RealBoost, to choose the best combination of features at each pixel is proposed. Given the probability estimates from a pool of features calculated by Kernel Density Estimate (KDE) over a certain time period, the most useful features to discriminate foreground objects from the scene background are selected. In their framework they conclude that while temporal gradients may be useful in handling dynamic scenes, however it cannot be computed accurately in regions that have low texture and is thus not very useful in such regions. Similarly, normalized colour, spatial gradients or texture features may be considered for obtaining invariance to illumination. Nevertheless, spatial gradients or texture is not very suitable for a region that has low spatial gradients since such feature will be unable to detect any object that has a low gradient itself.

Among these cues, proposed in the literature, spectral features offer the greatest generality with respect to the content of the scenes considered in video surveillance applications. The motion of moving objects results in intensity changes in magnitude such that intensity changes are important cues for locating moving objects in time and space, however, their relationship is not unique [88].

Colour-based vision applications face the challenge that colours are variant to illumination. Each image representation has its strength and weakness and is particularly applicable for handling a certain type of variation [64].

Photometric invariant features are functions describing the colour configuration of each image pixel to compute a function of the input images that is invariant to confounding scene properties but is discriminative with respect to desired scene information [89]. There has been considerable work in developing photometric invariants [53, 90]. One photometric invariant practical approach is discounting local illumination variations, such as shadows and reflections.

Examples of photometric invariant features for Lambertian or matte surfaces

Table 3.2: A summary of different spectral features used in state-of-the-art change detection techniques

References	Image representation	Advantages	Disadvantages
[11, 61, 68]	Grey-scale (Intensity, Luminance, Brightness)	A simple way of combining colour information. Computationally inexpensive.	Suffer from the effect of camouflage foreground .
[58, 69, 75] [70, 71]	RGB	Computationally inexpensive	The high correlation among the R, G, and B components if the intensity changes, all the three components will change accordingly. The measurement of a colour in RGB space does not represent colour differences in a uniform scale.
[73, 76]	YUV $YCbCr$	The colour distribution in UV and CbCr spaces is comparatively stable. Partly gets rid of the correlation of RGB	Correlation still exists due to the linear transformation, though not as high as RGB.
[71, 72, 74, 75] [73]	rgb	It is relatively robust to fast illumination changes. It is invariant to illumination variation and viewing geometry. More effective than RGB coordinates for suppressing unwanted changes due to shadows.	It suffers from a problem inherent to the normalisation: Very noisy (unstable chromatic components) at low intensities this is due to the nonlinear transformation from the RGB.
[58, 77, 91]	HSV	Hue and Saturation are insensitive to surface orientation changes, illumination direction changes and illumination intensity changes. Hue depends only on the spectral reflectance of the surface and the spectral power distribution (SPD) of the illuminant. Hue can be used for segmentation in 1-D space if the saturation is not low.	It faces the problem of unstable hue values at low saturation. If the colour lies close to white or black, Hue and Saturation play little role in distinguishing colours.
[92]	$CIELab$	It is standardised. It separates chrominance and luminance information. Can control colour and intensity information independently. Provide efficient measuring of small colour difference.	The coordinates of the white point are needed. Has the same singularity problem as other nonlinear transformations .
[81, 93, 94] [80, 99]	Reflectance ratio	It is based on shading model. It is invariant to either local geometry or spectral reflectance.	Valid only for Lambertian surfaces
[81, 95, 96]	Intrinsic images	It is invariant to illumination variation.	Computationally expensive.

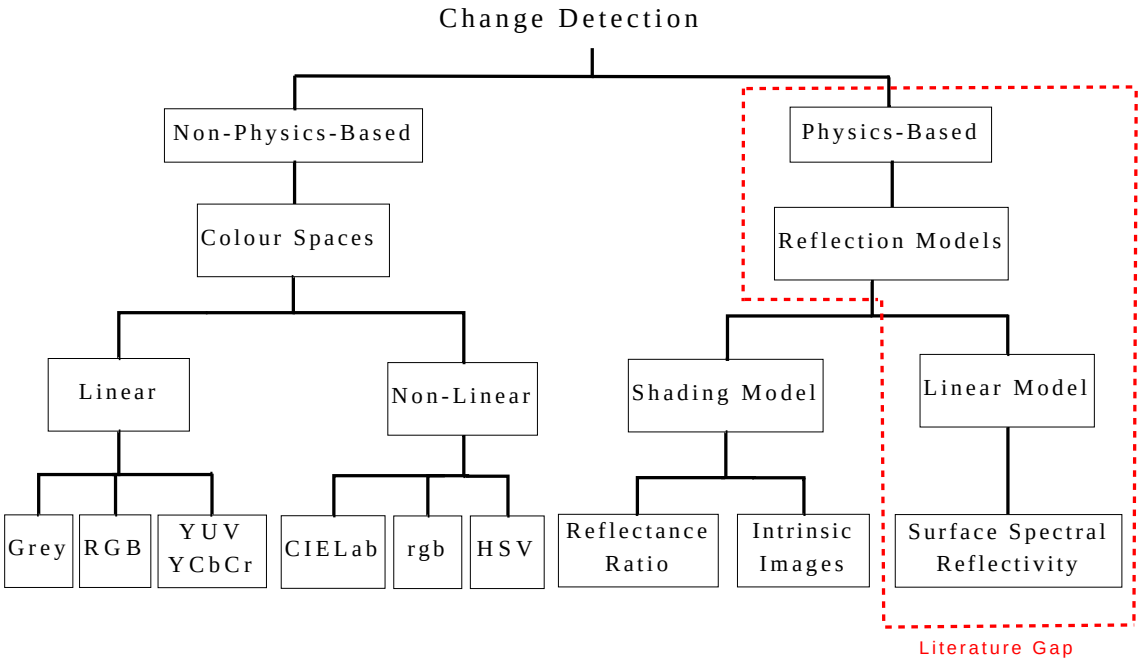


Figure 3.2: Classification of change detection approaches

are the normalised-RGB colour space, UV plane which represents chromaticity in YUV colour space, are insensitive to the luminance change [97] as well as hue and saturation in the HSV colour space [77]. Section 4.4.2 shows that those features depend only on the spectral reflectance of the surface and the spectral power distribution of the illuminant, and are invariant to the geometric relation between a surface normal and the illumination direction.

Other photometric invariant features are the reflectance ratio and intrinsic images. The reflectance ratio is introduced by [98] and used in [80, 81, 93, 94, 99, 100] which is invariant to either spectral reflectance or local geometry as Section 4.3.1 shows. Reflectance ratios can be computed when multiple images of a scene are available. Intrinsic images and the Log-Chromaticity Domain (LCD) are developed by [101], they are invariant to the illuminant spectral power distribution.

Additional photometric invariant is the surface spectral reflectance derived from the Shafer's dichromatic model [102]. The reflectance can be estimated by an analytic Bi-directional Reflectance Distribution Function (BRDF) model [103], or using a linear basis of reference objects [104–107].

Table 3.2 gives a summary of different spectral features used in state-of-the-art change detection techniques. The gap appears to be in the use of physics-based linear models and features such as surface spectral reflectance to represent the image in change detection algorithms.

Figure 3.2 classifies change detection approaches in terms of the image representation used to either physics-based or non-physics-based. We describe it as physics-based because it does not only contain colour images but also data on colour image formation such as spectral reflectance measurements, illuminant spectral power distribution, and camera spectral response. The dotted part represents the gap in the literature.

Physics-based representations have proven success in solving real-world computer vision problems, i.e. shadow detection, where limited number of assumptions about the scene are considered. However, few solutions have been proposed in the literature with regard to the use of physical models of image formation to solve the change detection problem. Only approximated models, e.g. shading model, have been used to derive physics-based photometric invariants, namely reflectance ratio and intrinsic images.

Researchers in the field of colour constancy [19–22], have introduced physics-based colour features, giving more care to the elements which govern the camera output. These features have not been fully investigated in the context of change detection in video surveillance. A limited amount of work has been done towards the use of physics-based image representations; only simple approximate models, discussed in 4.3.1, have been investigated and implemented [17].

Although the linear models for image formation have been applied and proven useful for a number of applications involving human skin [108] and pigment-based analysis and synthesis [109], nevertheless this model has not been applied yet in the field of change detection. The reason appears to be the computational complexity of such models, and hence unfeasibility of real-time implementation. This thesis is the first attempt to investigate the use of such models in change detection.

3.4.1 Non-physics-based image representations

Early change detection approaches were described for the intensity case (grey-scale images) [11, 61, 67, 68], the reason is that processing colour images require computational times considerably larger than those needed for grey-scale images. Nevertheless, this is no longer a major problem due to the decreasing costs of computation, increasing speed and relatively inexpensive colour cameras are easily available nowadays.

However, the main problem, which still resides, is to determine how to combine the information from colour channels and to determine which colour space is appropriate [55]. Because grey-scale images are the 1D projection of the three channels of colour images, so grey-scale images are regarded as a simple way of combining colour information.

An interesting approach is proposed by Fuentes and Velastin in [97], where a modified version of the luminance contrast is used, comparing the luminance coordinate in the YUV colour space 'Y' of a pixel in both the current and background images.

A recent algorithm, [63] uses grey-scale image representation, and introduces a background subtraction scheme using Independent Component Analysis (ICA) and, particularly, aims at indoor surveillance for possible applications in home-care and health-care monitoring, where moving and motionless persons must be reliably detected. The results show that parts of the foreground objects were missing due to the camouflage foreground in grey-scale images; therefore Tsai et al. [63] suggest the extension of their ICA model to separate vector-valued signals in RGB colour images, in order to eliminate the effect of camouflage, where colour may provide more information to discriminate a foreground object from its surrounding background.

There are several techniques that attempt to pre-compensate for illumination variations between frames caused by changes in the strength or position of light sources in the scene. In some of the earliest attempts to deal with illumination changes intensity normalization is used [110], i.e. normalizing the pixel intensity values to have the same mean and variance as those in the estimated background.

Alternatively, both, the current frame and the background can be normalized

to have Zero mean and unit variance. This allows the use of decision thresholds that are independent of the original intensity values. Instead of using global statistics, the frames can be divided into corresponding disjoint blocks, and the normalization independently performed using the local statistics of each block. This can achieve better local performance at the expense of introducing blocking artefacts [17]. However, algorithms which generally employ normalized colours typically work poorly in dark areas of the image.

As colour images can provide richer information than grey ones to measure the difference between two pixels, colour image is becoming more popular in the change detection literature [58, 70, 71, 74–76]. Generally, these are sort of dimensional extensions of techniques originally designed for grey-scale images.

The colour distribution in RGB space has a wide range of evolution because of the illumination variation at different parts of the field, Lu et al. [112] show that the colour distribution in UV space is comparatively stable. They conclude from their initial empirical data that a colour classifier defined on UV plane is capable of accurately classifying colours despite the mild illumination variation on the field under a stable lighting condition.

Developers have used different colour spaces, e.g. RGB, rgb, HSV, YUV and $CIELAB$. However normalised rgb and HSV are the most common colour spaces used. It has been shown that these colour spaces are more tolerant to minor variations in the illumination [113].

The rgb colour space, is one of the photometric invariants widely used in change detection [71, 73–75] but a rationale is usually not given [143]. Section 4.4.2 shows that rgb representation is invariant to illuminant intensity, it depends on illumination colour and surface reflectance but not on its orientation which makes it useful for material-based segmentation [89, 90].

The rgb colour space is widely used for its fast computation, and it is more effective than RGB coordinates for suppressing unwanted changes due to shadows of objects [72], however the rgb space, suffers from a problem inherent to the normalisation, where low intensities result in unstable chromatic components [78].

Like rgb, many existing invariants seek to isolate information about the

material properties in a scene and are therefore designed to be invariant to local illumination and viewing geometry [89, 90]. The hue H and saturation S of an HSV colour space are used in [58, 91] to obtain a limited level of intensity invariance in an indoor scene, however the HSV colour space faces the problem of unstable Hue values at low $Saturation$. Blauensteiner et a. [78] deal with this problem by modelling the hue-saturation relationship using saturation-weighted hue statistics. The method builds upon the improved Hue, Luminance, and Saturation space (IHLS), where they represent chrominance information for background modelling and shadow suppression.

Colour space selection

In order to utilize colour images, the selection of a colour space is a crucial issue. Applying different colour spaces can significantly change the detection results. Some authors [92, 113, 114] have compared the performance of several colour spaces, the detection results show different error ratios according to which colour space is used.

A colour space comparison has been carried out by Hwang et al. [92]. Their aim was to determine an appropriate colour space for their proposed change detection method. They have compared RGB, $CIELAB$, and $CIELCH$ colour spaces. They, then propose a criterion, by means of statistical formulation, for measuring the expected number of error pixels in order to determine which specific colour space that has best change detection result corresponding to current image sets. They conclude that each colour space's ability to detect changes differs according to the camera used, the captured scene, and the illumination. Their results show that the $CIELAB$ and $CIELCH$ colour spaces give the best performance for indoor scenes, while the $CIELAB$ and RGB colour spaces give the best performance for outdoor scenes.

An empirical study has been carried out by Kumar et al. in [113] with the motivation of determining which colour space is best for foreground segmentation. They compared different colour spaces, which includes RGB, XYZ, $YCbCr$, HSV and the rgb, for foreground and shadow detection, their results show that the $YCbCr$ is the best colour space for optimal foreground and shadow detection and they used

this in their proposed framework for real-time behaviour interpretation from traffic video in [115]. They conclude that different colour spaces show different efficiency in detection of foreground objects.

Most of the recent change detection techniques [58, 70, 71, 74–76] use well-known colour spaces to represent the image, trying to model the variation in illumination and camera noise using either statistical manipulation techniques or post-processing adjustments. These algorithms still suffer from illumination change and camera noise [17].

3.4.2 Physics-based image representations

The colour is not a physical property of an object as stated by Maloney in [104], rather it is a perceptual phenomenon and therefore a subjective human concept. The perceptual ability that permits humans to discount spectral variation in the illumination and assign stable colours to objects is called colour constancy.

The camera output is dependent upon: the camera characteristics, the incident illuminant and the scene content. For Indoor surveillance application using static camera, the scene and camera never change, then the aim of colour constancy is to identify and eliminate the effects of the illuminant.

The physical approach, used in colour constancy [20, 102, 104], studies colour image formation in order to estimate scene illuminant spectral power distribution in order to extract the full objects surface spectral reflectance taking into consideration camera sensors sensitivities. Physics-based representations of objects using images under a large number of lighting conditions have been proposed for illumination invariant change detection [100, 116, 117] and shadow removal [80, 118].

Change detection based on physical models would permit image representation via the study of the process of image formation. Algorithmically, the basics of these approaches are often very similar to those of other change detection methods, and only differ from them in the fact that these algorithms explicitly use the reflectance models of surfaces to represent a colour image.

Various reflection models are used that describe the light reflected by a surface as a weighted combination of a diffuse and a specular components, most of them are

developed in the field of computer graphics [102, 119, 120]. They differ in the way these two components are modelled and weighted when combined.

Some models, such as the Phong model [119] empirically approximate some of the underlying rules of optics and thermal radiation, however, it does not have a physical basis. This can represent a limitation if the model is used to predict the colour appearance of a surface. Other models, such as the Torrance-Sparrow model [120] are thoroughly derived but they are impractical for computer vision applications. Approximate models, such as the Dichromatic Reflection Model [102], are still derived from physics-based reflectance models, but they are modified so as to emphasize the desired aspects of the models as well as to ignore their other unnecessary aspects. The Dichromatic Reflection Model is used for these reasons in this thesis.

The dichromatic reflection model in [102] is a usual choice for those algorithms employing a physical model to represent colour images. This model postulates that the light reflected from an infinitesimal piece of surface of an inhomogeneous dielectric material is the result of the addition of two components, namely, the surface (specular) reflection and the body (diffuse) reflection. Each of these parts is further divided into two elements, one accounting for the geometry and another purely spectral. The main limitation of this technique is that it can be applied only to inhomogeneous dielectrics, Lambertian surfaces. Many common materials can be described this way, including paints, varnishes, paper, ceramics, and plastics. Metals, glass, and crystals are excluded as they are optically homogeneous. Only opaque surfaces are considered in the model.

By applying the dichromatic reflection model for lambertian surfaces, three main photometric invariants which represent the intrinsic surface reflectance can be extracted, reflectance ratio, intrinsic image and surface spectral reflectance.

Reflectance ratio

Reflectance ratio is a photometric invariant developed by Nayar et al. [98, 121]. Originally, it was developed for grey-scale images and used to obtain a surface segmentation that is invariant to illumination and surface geometry. Bunyak et

al. [99] extend the reflectance ratio in [98, 121] to three colour components in order to classify each pixel locally by exploiting this invariant. Their algorithm exploits spatial and spectral information; no a priori knowledge about camera, illumination or object/scene characteristics are required. A method for estimating the body colour of surfaces based on a physical model is introduced in [80], using the spatio-temporal reflectance ratio test, derived from dichromatic reflection model. Lu et al. [112] present a colour classification algorithm for their Robocup system that is adaptive to continuous variable lighting, motivated by the dichromatic reflection model.

A change detection method based on Phong's model is proposed in [100]. The ratio of the intensities of two frames of an image sequence is used to detect changes in the time-varying illumination case. Durucan et al. extend the former method by combining it with optical flow method in [94] and by proposing the Linear Dependence Detector (LDD) in [93].

Intrinsic image

Assuming that the scenes contain Lambertian surfaces, intrinsic image which contains mainly physical object information can be extracted. In [81], a homomorphic filter is applied to the input intensities in order to extract the reflectance component. The reflectance component is then provided as input to the decision rule step of a moving object detection process.

Weiss [79] uses multiple input images of a fixed scene, assuming constant reflectance and changing illumination, and proposes a Maximum Likelihood (ML) estimation framework to estimate a single reflectance image, intrinsic image, and multiple illumination images. Matsushita et al. [96] extend Weiss's algorithm and introduce an eigenspace that captures the illumination variations to eliminate shadows from surveillance images. Tappen et al. [122] propose an algorithm that uses multiple cues to recover shading and reflectance intrinsic images from a single image. Using both colour information and a classifier trained to recognize grey-scale patterns, each image derivative is classified as being caused by shading or a change in the surfaces reflectance. Porikli [95] proposes a method that

decomposes a scene into time-varying background and foreground intrinsic images and proves the effectiveness of the intrinsic background/foreground decomposition even under sudden and severe illumination changes.

Finlayson et al. [101, 123] prove that the log-chromaticity domain (LCD), $\ln(\frac{R}{G})$ and $\ln(\frac{B}{G})$, are independent of light intensity and there even exists a weighted combination of LCD which is independent of both light intensity and light colour. Finlayson et al. [101] recently devised a method for the recovery of an illumination invariant image, which is similar to the intrinsic image, from a single colour image. They assume the input image contains both non-shadowed surfaces and shadows cast on those surfaces. They calculate an angle for an"invariant direction" in a log-chromaticity space by minimizing the entropy of colour distribution.

Surface spectral reflectance

In much of the previous literature, computational models are used to model the surface reflectance [102, 124] and illuminant [19, 21] spectra by a finite weighted sum of basis functions [125, 126]. Given a linear model for image formation [105–107], there exists a linear relationship between the sensor absorptions under two differing illuminants [127].

Negahdaripour [128] uses Principal Component Analysis (PCA) to extract a set of basis images that represent the views of a scene under all possible lighting conditions. Pizarro and Singh [129] model the illumination component as a piecewise polynomial function.

According to Radke et al. [17], these sophisticated models of illumination compensation are not commonly used in the context of foreground detection. However, these linear models have been applied successfully in a number of real-time computer vision problems [108, 109, 130].

3.5 Statistical manipulation

Change detection methods typically build a statistical model in order to define a decision rule which discriminates between the pixels of the moving object and those of background pixels. Background modelling approaches build an object-free model of the background; this model is learned over time using a series of salient background frames. The new frame is compared to the background model and the differences are marked as foreground objects [61].

If the camera is fixed and the background can be expected to stay relatively constant, the background can be modelled as a single static image that may be easily estimated and segmented. The required measurement in this case is the intensity in case of grey-scale images, the colour components in case of colour images or the reflectivity components in case of a physics-based representation. If the background is not actually constant, then modelling both the mean intensity at a pixel and its variance gives an adaptive tolerance for some variation in the background.

Haritaoglu et al. [11], in their W^4 surveillance system, use a grey-scale image representation, and the background is modelled by determining the minimum and maximum background intensity values as well as the largest inter-frame distance between consecutive frames. If the observed pixel is more than certain threshold levels away from either the minimum or the maximum, it is considered foreground. These values are estimated over several frames and are periodically updated for background regions.

Various approaches of increasing complexity have been considered in the literature. If a scene contains motion that should be considered part of the background, more tolerant models are required. Statistical models used in existing change detection approaches are divided into parametric and non-parametric approaches. Parametric approaches use a specific functional form with adjustable parameters chosen to fit the model to the data set; examples are: Gaussian models such as mixtures of Gaussians, and elliptic boundary model. Examples of non-parametric approaches are: normalised lookup table, Bayes classifier and self organizing map.

3.5.1 Parametric methods

Several researchers [58, 76] have described adaptive background subtraction techniques in which a single Gaussian density is used to model the background. Pixels belonging to the foreground are determined as if they lay a number n of standard deviations from the mean of the background model, and then they are clustered into objects. In [58, 76], a single Gaussian was considered to model the statistical distribution of the background at each pixel. The parameters of this model are the mean and covariance matrix.

Horprasert [73] uses the Ycbcr colour space or normalised rgb and assumes that there is an expected chromaticity line on which the pixel value should be kept. The distortion from this line is given as both chromaticity and brightness distortion being generated by standard deviation. Shadows and highlights are also detected. The limitation of this algorithm is that shadows on very dark backgrounds or several shadows added together will not be detected effectively.

Francois and Medioni et al. [58] use the HSV colour space, and build a background model as a Gaussian distribution. The current image is subtracted from the mean value model and the resulted difference values of each pixel give the information of classifying to either foreground or background by comparing it to the standard deviation model. With this assumption a background model is generated by considering the mean value and standard deviation for each pixel.

Similar to the previous assumption McKenna [71] models the background with mean value and also standard deviation. However, his system considers the normalized rgb colour space and the edge. For each channel the models are generated and a combination of both classification results gives the final segmentation mask.

Jabri [74] uses both information the colour and the edge similar to the one of McKenna [71]. The background model is trained in both mentioned parts by calculating the mean and standard deviation for each pixel of any colour channel. With subtraction of the incoming current image on each channel, confidence maps are generated for both information colour and edge. After that a combination of the two maps are utilized by taking its maximum values.

Hong [75] uses well-known RGB and normalized rgb colour to model the background. As mentioned in previous methods the mean and standard deviation are used again and these are calculated over each colour components. Each colour space has its own classification part in which the current image is converted first in each colour space. Within each colour space the pixel can be classified in four categories, background, foreground, highlight, or shadow. This method is efficient only in less dynamic scenes but has difficulties with periodic backgrounds motion (e.g. swaying trees), background elements moving, and fast illumination changes (flood of sunlight, shadows or lights switched on).

The previous model does not work well in the case of dynamic natural environments since they include periodic motions like waving trees or ocean waves. By using more than one Gaussian distribution per pixel, it is possible to handle such backgrounds. When a single Gaussian is insufficient to model the distribution of pixel values, a finite Mixture of Gaussians (MoG) may be used instead.

The MoG method is quite popular and is the basis for a large number of related techniques, many authors have proposed improvements and extensions to this algorithm [69, 115, 118, 131]. MoG maintains a density function for each pixel. Thus, it is capable of handling multimodal background distributions. In MoG, the model parameters can be adaptively updated without keeping a large buffer of video frames.

For example, Raja et a. [91] use mixture of Gaussians with the HSV colour space to model a multi-coloured foreground object, in which each Gaussian models one colour in the object. Stauffer and Grimson [69] use multiple Gaussians to model the scene and develop a fast approximate method for updating the parameters of the model incrementally. Such an approach is capable of dealing with multiple hypotheses for the background and can be useful in scenes such as waving trees, beaches, escalators, rain or snow.

Tian et al. [131] use both grey-scale images and colour images, to extend and improve MoG background subtraction method developed by Stauffer and Grimson [69, 132] by integrating texture and intensity information to remove shadows and to enable the algorithm working for quick lighting changes. They discriminate

between static foreground objects, and abandoned or removed objects by analyzing the change in the amount of edge energy associated with the boundaries of the static foreground regions, without using any tracking or motion information.

In [115] three Gaussians are used to model the histograms of each channel of the three colour components to model the background for traffic surveillance application. The Expectation-Maximization (EM) algorithm is used which gives very good model-fitting but is computationally expensive [64]. This idea was extended but instead of using the exact EM algorithm, an online K-means approximation was used.

Liu et al. [118], assume Phong reflection model, extract surface reflectance component using homomorphic filtering. They assume that reflectance component fits Gaussian distribution, and then use a MoG to model it in order to remove shadow. They show that their method is not sensitive to the change of illumination.

The Mahalanobis distance is used with a correlated image representation to measure the distance between background pixels and a foreground pixel. Background pixels are associated with the least Mahalanobis distance. The *log* likelihood value $l_{(x,y)}$ is computed and used to classify pixels.

Wren et al. [76], in his Pfinder system, uses the YUV colour space, and includes a background estimation module, by associating each pixel in the background with its mean colour value and a covariance matrix. The colour for each pixel is described by Gaussian distribution. Pixels are classified into background or moving objects by measuring the deviations from the background model. Deviations are measured in terms of Mahalanobis distance in YUV colour space. If the distance is sufficient then the process of building a blob model is initiated.

Hidden Markov Models (HMM) can also be used to represent the pixel process where its states can represent different states that might occur in the pixel process, such as background, foreground, shadows, day and night illumination. It can also be used to handle the sudden changes in illumination where the change from a status to another, such as the change from dark to light, day to night, indoor to outdoor, can be represented as the transition from state to state in the HMM. Three-states

HMM are used in [70] to represent background, shadows and foreground, where both background and shadows are modelled as single Gaussian distribution.

However, parametric approaches are generally computationally intensive and are sensitive to the tuning of their parameters. These approaches are very sensitive to sudden changes in global illumination. For the case of a static scene, the variances of the background components may become very small. A sudden change in global illumination can then turn the entire frame into foreground. This method can deal with periodic background motion, however it cannot follow fast illumination changes, which cause spurious "foregrounds" and may miss targets in such cases. To deal with the limitations of parametric methods, when the Gaussian assumption for the pixel intensity distribution does not hold a nonparametric approach to background modelling is used.

3.5.2 Non-parametric methods

When the density function is more complex and cannot be modelled parametrically, a non-parametric approach able to handle arbitrary densities is more suitable. This approach utilizes a general nonparametric kernel density estimation technique for building a statistical representation of the scene background. Each pixel is modelled as a random variable in a feature space with an associated probability density function estimated directly from the data without any assumptions about the underlying distributions. It avoids having to choose a model and estimate its distribution parameters.

A computational framework for efficient density estimation is introduced by Elgammal et al. in [72]. They propose the use of Fast Gauss Transform (FGT) for efficient computation of colour densities, allowing the summation of a mixture of M Gaussians at N evaluation points. They use the median of the absolute differences between successive frames as the width of the kernel. Thus, the complexity of building the model is the same as median filtering.

Hwang et al. [55] show that each colour band's noise does not follow the well-known zero means Gaussian distribution. To accurately model each channel's noise, they propose a Generalized Exponential Model (GEM), which estimates the

noise distribution on the Euclidean distance and corresponds to unchanged regions. Subtracting the estimated noise distribution from the whole distribution provides the distribution of unchanged regions and changes. The detection is then done by a simple pattern classification step.

Other approaches use fuzzy classification where a difference image is generated for each RGB colour space component. For every channel's result a corresponding threshold is determined by use of unimodal thresholding method for considering the fuzzy set of moving pixels. Then these thresholds avail to generate fuzzy images which at least are combined to one final fuzzy image. Subsequently, a preliminary mask is achieved by thresholding which describes all detected moving pixels in all appearances.

For example, Shen [133] uses the well-known RGB colour space and the system can be represented in two sections. One of them is the block for generation of fuzzy classification and the other one is the block for elimination of falsely detected segmentation regions. The fuzzy classification is applied to take into account the mobility of pixels precisely instead of the so called binary classification.

The background can also be represented by a group of clusters which are ordered according to the likelihood that they model the background and are adapted to deal with background and lighting variations. Incoming pixels are matched against the corresponding cluster group and are classified according to whether the matching cluster is considered part of the background. Butler et al. [134] model each pixel by a group of K clusters where each cluster consists of a weight w_k and an average pixel value or centroid c_k.

Kim et al. [135] quantize sample background values at each pixel into codebooks which represent a compressed form of background model for a long image sequence. They adopt a quantization/clustering technique, their method can handle scenes containing moving backgrounds or illumination variations, and it achieves robust detection for different types of videos. Mixed backgrounds can be modelled by multiple codewords.

Unlike MoG, Kim et al. do not assume that backgrounds are multimode Gaussians. Also, in contrast to Kernel, Kim does not store raw samples to

maintain the background model.

A background subtraction scheme which uses Independent Component Analysis (ICA) is proposed recently by Tsai et al. in [63]. They use an ICA model to measure the statistical independency based on the estimations of joint and marginal probability density functions from relative frequency distributions of the background. Their proposed ICA model can separate two highly-correlated images. Then, their trained de-mixing vector is used to separate the foreground in a scene image with respect to the reference background image.

The creation of a small number of eigenbackgrounds to capture the dominant variability of the background rather than implicitly modelling the background dynamics, is introduced in [136]. This approach looks at global statistics rather than the local constraints used in the previously described approaches. A threshold on the difference between the original image and the part of the image that can be generated by the eigen-backgrounds differentiates the foreground objects from the background, with the assumption that the remaining is due to foreground objects.

3.6 Threshold selection

The simplest method for detecting changes is by differencing the intensities of corresponding pixels. If the difference in grey-scale exceeds a preset threshold, the pixel is regarded as changed. Therefore, the threshold chosen is critical: if the threshold is too high, it will suppress significant changes, if too low, it will flood the difference map with false changes [55]. By changing the decision threshold, in order to reduce one type of error, the other type of error increases; it is not possible to minimize both errors at the same time.

The threshold in the decision rule determines if a pixel in the test image corresponds to a moving object or not. Due to camera noise and illumination variations, a certain number of pixels are classified as a moving object even if they do not belong to it: a high value of the threshold allows maintaining a low number of false detections. On the other hand, objects presenting a low contrast with respect to the moving object risk to be eliminated if the threshold is excessively

high.

Several pixel-based methods for determining the threshold of change detection have been proposed. Rosin and Ioannidis [137] carry a comprehensive survey and comparison of four different threshold selection methods: the noise intensity model, the signal intensity model, the noise spatial model, and the signal spatial model. Smits and Annomi [138] present a method of threshold selection for specific application requirements (in terms of missed detection and false alarm rates) using Receiver Operating Characteristics (ROC) curve. Wren et al. [76] propose a method for selecting thresholds at each pixel adaptively, based on the grey-level distributions of the background points.

The use of ROC curves to select the optimal threshold for a set of parameters is proposed in [139]. This work is extended in [140] which introduces a minimax threshold selection method, a probabilistic criterion designed to minimise the maximum possible Bayesian risk when no knowledge about source probabilities is available. The minimax equation and the related line can be represented on the ROC curve. Oberti et al. [140] show that the interaction between the minimax equation and an ROC curve gives the best operating point. The threshold can be computed to produce a desired false alarm rate or missed detection rate.

3.7 Post-processing adjustments

The outputs of foreground region detection algorithms explained in previous section generally contain noise and therefore are not appropriate for further processing without special post-processing. A post-processing step can be divided into two stages, noise removal and shadow detection.

3.7.1 Noise removal

A noise removal step helps in improving the quality of the segmented mask using some morphological operations. Small holes in objects are most likely segmentation errors and the segmentation can be improved by closing such holes in objects. Small regions in the mask are likely to be either camera noise errors or small changes

which are irrelevant for the application in hand. Hence, a filter is applied to remove regions if their size is below a small threshold.

Morphological operations, erosion and dilation [11], can be applied to the foreground pixel map in order to remove noise that is caused by items listed above. The aim in applying these operations is to remove noisy foreground pixels that do not correspond to actual foreground regions and to remove the noisy background pixels near and inside object regions that are actually foreground pixels. Erosion, as its name implies, erodes one-unit thick boundary pixels of foreground regions. Dilation is the reverse of erosion and expands the foreground region boundaries with one-unit thick pixels. The order of these operations affects the quality and the number of units used affects both the quality and the computational complexity of noise removal [17].

3.7.2 Shadow detection

Although shadow is not part of the foreground, the drop in luminance of the shadow pixels is usually interpreted as a significant change in the pixel's colour and hence, misclassified as foreground object [141]. Shadows increase false alarms rate in change detection and lead to severe changes in the shapes of objects or merging of multiple objects.

Shadow detection and suppression from the set of foreground points aim to prevent moving shadows being misclassified as moving objects, thus improving object detection [142, 143].

If the brightness reduction is not very significant, shadow detection algorithms can classify foreground pixels as shadow pixels if the change in hue is small and the brightness is reduced slightly [144]. Many algorithms have been proposed to handle the shadow problem. Wren et al. [76] use the normalised components, U/Y and V/Y, of a YUV colour space to remove shadows in a relatively static indoor scene. In [145] a non-parametric approach is proposed that use colour, global and dynamic features for shadow removal to enhance object detection. The detection is based on the observation that shadows change significantly the brightness of an area without significantly modifying the colour information. On the basis of the

same observation, a statistical background subtraction algorithm which exploits a computational colour model that separates the brightness from the chromaticity components of a pixel is presented in [73].

Physics-based shadow removal

Marchant et al. [146] propose a physics-based method for shadow compensation in scenes illuminated by daylight. The illumination is represented as a black body and the colour RGB camera filters are assumed to be of infinitely narrow bandwidth. The ratio $(\frac{R}{B})/(\frac{G}{B}^A)$ is used to detect shadows where A is found only depending on surface reflection as the illumination changes and can be pre-calculated from the daylight model and for the specific camera.

Nadimi and Bhani [80] introduce a physics-based approach to separate moving cast shadows from the moving objects in an outdoor environment. Their approach does not rely on any geometrical assumptions such as camera location and ground surface/object geometry. The approach is based on a new spatio-temporal albedo test and dichromatic reflection model and accounts for both the sun and the sky illuminations.

In [99], Bunyak et al. extend the reflectance ratio in [98, 121] to detect moving cast shadows to improve the performance of moving object detection, in which they use the MoG approach.

3.8 Background maintenance

The role of the background modelling process is to estimate a background model by observing the scene during a training period; this model could be updated continuously. Continuous updating of the model can make the foreground extraction more robust to illumination change or objects that become part of the background.

Haritaoglu et al. [11], in his W^4 surveillance system, detect when its background model is invalid by detecting when 80% of the image appears as foreground. Limitations are the existence of many people, occlusion, shadows and

slow moving objects.

Another class of background modelling methods try to model the short-term dynamical characteristics of the input signal. Several authors [61, 147] have used a Kalman-filter [148] based approach for modelling the dynamics of the state at a particular pixel. Ridder et al. [147] propose a linear prediction method which model each pixel value using a Kalman Filter to compensate for illumination variation and adaptively estimate the background and detect the foreground. Monnet et al. [61] present an autoregressive form to predict the frame to be observed. Two different techniques are studied to maintain the model, one that update the states in an incremental manner and one that replaces the modes of variation using the latest observation map.

A simpler version of the Kalman filter called Weiner filter was considered in [68] that operates directly on the data. Toyama et al. [68] use a one step Wiener filter which is linear predictor of the intensity at a pixel, based upon the time history of intensity at that particular pixel. This can account for periodic variations of pixel intensity. The measurement includes two parts, the intensity at the current frame, and the recent time history of intensity values at a given pixel. Toyama et al. [68] proposes the Wallflower algorithm in which background maintenance and background subtraction are carried out at three levels: the pixel level, the region level, and the frame level. Toyama computes the current background estimate by applying a linear predictive filter on the pixels in the buffer using a Wiener filter to predict a pixel's current value from a linear combination of its N previous values. Pixels whose prediction error is several times worse than the expected error are classified as foreground pixels. The filter coefficients are estimated at each frame time based on the sample covariance, making this technique difficult to apply in real-time. Kahl et. al. [149] enhance the computational efficiency of the previous approach by using an incremental linear PCA model in combination with local, spatial transformations. They propose a new scheme for novelty detection by classifying new observations from previous samples, as either novel or belonging to the background. This approach is capable of handling non-stationary background scenarios, such as waving trees, rain and snow.

Koller et al. [83] handle the change of lighting condition using a moving-window average method. An obvious problem with this technique is that all information coming from both background and foreground is used to update the background model. If some objects move slowly, these algorithms will fail. The solution to this problem is that only those pixels not identified as moving objects are used to update the background model.

KaewTraKulPong and Bowden [150] present a new update algorithm for learning mixture models, for their proposed method for detecting moving shadows. Zivkovic [152] extends the former method by updating the number of components of the mixture which is constantly adapted for each pixel.

Elgammal et al. [72] present a way to combine the results of two background models (a long term and a short term) in such a way to achieve better update decisions. Short-term model is a very recent model of the scene. It adapts to changes quickly to allow very sensitive detection. The sample is updated using a selective-update mechanism, where the update decision is based on the final result of combining the two models. Long-term model captures a more stable representation of the scene background and adapts to changes slowly. The sample is updated using a blind-update mechanism. Computing the intersection of the two detection results eliminate the persistence false positives from the short term model and will eliminate as well extra false positives that occur in the long term model results. Grimson and Stauffer [69] use an online estimation method to update background images in order to adapt to illumination variance and disturbance in backgrounds.

3.9 Evaluation methods

Change detection approaches can be evaluated using analytical or empirical methods. The analytical method considers the principles, requirements and complexity of algorithms, while, empirical methods measure the quality of the segmentation results [151].

Empirical evaluation methods for change detection approaches can be classified into two classes: discrepancy (reference) methods and goodness (stand-alone)

methods. The former class needs some reference images to arbitrate the quality of segmentation while the latter class can perform the evaluation without the help of reference images. Discrepancy methods use objective criteria which indicate the difference between the segmented images and reference images [153]. Goodness methods use subjective criteria which reflect some desirable properties of segmented images.

Erdem et al. [154] present three performance evaluation metrics that do not require segmented ground truth. They propose spatial differences of colour and motion and the boundary of the segmented video image and the temporal difference between the colour histogram of the object in the current frame and previous video frames. The authors show that under certain assumptions, the time-consuming annotation of ground truth is not necessary. However, when more than segmentation only is required, ground truth will have to be generated anyway.

In most cases ground truth is essential for performing a quantitative analysis of an algorithm's results. There are three main approaches to generating ground truth. The first uses synthetic data; this method enables ground truth to be easily provided, the problem is that the synthetic data will probably not faithfully represent the full range of real data. Alternatively, real image data can be manually annotated, or to specify an ideal image segmentation. Now we have the opposite problem: the image data is good, but the ground truth is dubious. Since manual mark-up is tedious and time consuming large volumes of ground truthed data are likely to have errors. In addition, the process has become subjective, and different annotators often give different ground truth. A third approach avoids explicitly determining a ground truth dataset, and relies instead on evaluating the algorithms's outputs by a human panel. Two disadvantages are the time consuming nature of the exercise (more images need to be viewed), and the difficulty in incorporating additional algorithms into the evaluation results at a later date (unless the same panel is reconvened).

A framework for performance evaluation of change detection in traffic and surveillance applications is introduced by Desurmont et al. in [155]. They propose the creation of ground-truth data; available evaluation of data sets; performance metrics; and presentation of the evaluation results.

Metrics for evaluation of image segmentation methods are proposed in [156] with the goal to create objective measures corresponding to evaluation by a human observer.

The use of Receiver Operating Characteristics (ROC) curves and pixel-based metrics for evaluation is proposed in [139], which includes the probability of a False Alarm (FA) against the probability of a Miss-Detection (MD), to extract useful information about the system performance. A Perturbation Detection Rate (PDR) analysis is proposed by Chalidabhongse et al. in [157] that has some advantage over ROC analysis. No ground truth is needed for the evaluation method, but the method does not consider detection rates through the video frame or over time.

A pixel-based evaluation framework for testing thresholding algorithms in the content of surveillance applications is introduced by Rosin et al. in [137].

Zhang [153] concludes that the results obtained in the field of segmentation evaluation are still far from satisfactory. A number of factors still limit the advancements of segmentation evaluation and the performance improvements of segmentation algorithms:

1. There is no common mathematical model or general strategy for evaluation.
2. It is difficult to define wide-ranging performance metrics and statistics.
3. The testing data used in evaluation are often not representative enough for actual application.
4. Appropriate ground truths are hard to determine objectively.
5. Often large costs (both time and effort) are involved in performing comprehensive valuations.

3.10 Summary

The main problem in change detection algorithms is how to separate changes due to sensor noise, global and local illumination variations, background periodic motion, static foreground, moving background, and similar background-foreground coloured objects from those caused by moving objects.

Although the change detection problem has been increasingly studied in recent years there exists no generally accepted method to detect moving objects in image sequences [17, 38, 50, 51]. The conclusion is that different approaches to change detection should be taken when addressing different kind of scenes.

For surveillance applications, approaches based on background modelling are widely used, which may be decomposed into three major steps, image representation, statistical manipulation and threshold selection. Table 3.3 summaries other steps involved in change detection and how each step contributes toward the solution of one or more problem. For example, statistical modelling contributes mainly to the background periodic motion problem using both spectral and temporal features. The background maintenance step is crucial to overcome static foreground objects, moving background as well as global illumination variation. Camera noise causes mainly false alarms, which could be solved using spatial features either as part of the image representation or in the post-processing stage. Other post processing techniques such as shadow removal are important to eliminate the variation in local illumination. The main source of missed detection is the camouflaged foreground, which could be solved mainly by the use of an appropriate image representation.

Various spectral, spatial and temporal features have been proposed in the literature as cues to inform the presence of a moving object in a scene. Among these cues, spectral features offer the greatest generality with respect to the content of the scenes considered in video surveillance applications. Change detection approaches may be classified depending on these spectral features as either physics-based or non-physics-based.

The analysis of the state-of-the-art outlined the fact that most of the reviewed approaches are non-physics-based that use well-known colour spaces to represent the image. Various statistical approaches have been proposed for modelling a given scene background. In the majority of these approaches, the image representations are chosen arbitrarily and the same representations are used for different scenes, without empirical studies or even justifications to the rationale behind the chosen representation. Nevertheless, the selection of an image representation is a crucial issue; each image representation's ability to detect changes differs according to the

Table 3.3: Summary of change detection problems

Problem	**Consequence(s)**	**Solution**	**Features**
Camera Noise	False Alarm Missed Detection	Post-Processing	Spatial
Global Illumination Variation	False Alarm	Illumination Estimation Background Maintenance	Spectral Temporal
Local Illumination Variation	False Alarm Missed Detection	Post-Processing Image Representation	Spectral
Camouflaged Foreground	Missed Detection	Image Representation	Spectral
Periodic Background Motion	False Alarm Missed Detection	Statistical Modelling	Spectral Temporal
Static Foreground	False Alarm Missed Detection	Background Maintenance	Spectral Temporal
Moving Background	False Alarm Missed Detection	Background Maintenance	Spectral Temporal

camera used, the captured scene, and the illumination.

These algorithms still suffer from illumination change and camera noise and the gap appears to be in the image representation used which does not give any consideration to the elements which govern the camera output. Using camera output directly without considering the illumination type, camera sensor characteristics, and the physical meaning behind this output, makes it more difficult to develop robust algorithms.

Few solutions have been proposed in the literature with regard to the use of physical models of image formation to solve the change detection problem. Only approximated models, e.g. shading model, have been used to derive physics-based photometric invariants, namely reflectance ratio and intrinsic images. In the next chapter, this issue is investigated and the image formation models that are used in the proposed change detection approaches are discussed.

The gap appears to be in the use of physics-based linear models and features such as surface spectral reflectance to represent the image in change detection algorithms.

These models have not been applied yet in the field of change detection. The reason appears to be the computational complexity of such models, and hence unfeasibility of real-time implementation. This thesis is the first attempt to investigate the use of such models in change detection.

The criteria used in the evaluation method for judging the performance of the segmentation represent an essential element and critical factor in change detection evaluation. However, the review of the principles and methods for evaluating and comparing the performance of change detection approaches shows that there is no widely accepted evaluation test framework. Moreover, Most of the goodness metrics proposed gives a measure to the segmentation quality not to the robustness of the algorithm.

"Understanding the physics of light, reflection and image formation is a requisit for effective low-level computer vision."

A. Cavallaro [158]

4

Image Formation Models

4.1 Introduction

The majority of change detection approaches, reviewed in Chapter 3 use one of the colour spaces as its input directly, without relying on any physical aspects of image formation. This chapter presents an overview of the image formation process through a review of each of its elements and their interaction. The purpose of the review is to introduce the fundamental models and features that will be of use in the proceedings of this thesis for the foreground segmentation and change detection. The theory of image formation is discussed along with the introduction of a number of physical features, namely, reflectance ratio, intrinsic image and surface spectral reflectance. This is followed by a review of a number of colour spaces as well as the relation between each colour space and image formation models.

4.2 Image formation

Appearances of scenes depend on four fundamental elements: an illuminant, a medium, a material and a vision system. The illuminant represents the source of visible electromagnetic energy and is characterised by its SPD . The medium is the medium in which electromagnetic waves travel [1]. The surface of the material modulates the incident electromagnetic energy and is represented by the surface spectral reflectance, the fraction of incident radiation reflected by this surface. The vision system is identified by the spectral sensitivities of its photosensitive sensors which represent the response of such elements to the received electromagnetic waves.

Apart from the aforementioned spectral features which characterise the elements of the scene appearance another important parameter is the geometrical features, which represent the scene structure, the illuminant orientation, the surface roughness and the viewing geometry. These features combine non-linearly to yield a digital image. Recovering these features from images is an important problem in image processing; however, this recovery is generally hard with the limited amount of information provided by standard commercial imaging devices.

4.2.1 Illuminants

A light source illuminates the material surface by emitting electromagnetic waves composed by a mixture of energy at different wavelengths. The emitted light denotes the visible part of the electromagnetic energy that covers wavelengths from approximately 400 nm to 700 nm. The visible spectrum represents only a small portion of the complete electromagnetic spectrum [2]. The power emitted at each wavelength gives the SPD of the source.

Black-body (Planckian) radiator is a special type of theoretical light that emits radiation, due only to thermal excitation, at a maximum rate for its given temperature (T^oK) and absorbs all the radiations that strikes it.

[1] where in this thesis the medium is assumed be air

[2] Infra Red illuminants spans the range of IR-A 700 nm to 1400 nm, IR-B 1400 nm to 3000 nm and IR-C 3000 nm to 1m, are used for applications such as military, surveillance, night vision, etc..

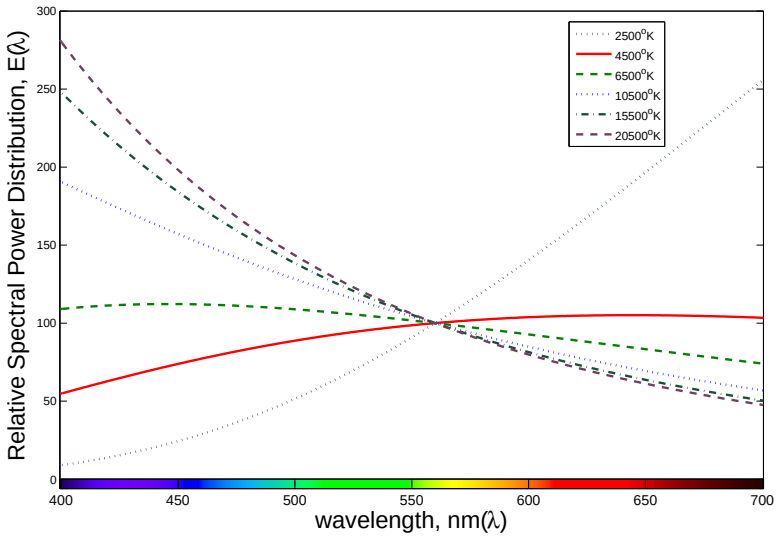


Figure 4.1: Relative spectral power distribution of six Planckian radiators, the relative spectral power distribution is normalised such that it has a value of 100 at a wavelength of $560nm$

The SPD of a black-body may be described as a function of its absolute temperature and wavelength by Planck's formula:

$$E(\lambda) = c_1 \lambda^{-5} \left(e^{\frac{c_2}{\lambda T}} - 1 \right)^{-1} \approx c_1 \lambda^{-5} e^{\frac{-c_2}{\lambda T}} \tag{4.1}$$

where $E(\lambda)$ is the light SPD in $Watts/m^2$; $c_1 = 3.74183 \cdot 10^{-16}\ Watts \cdot m^2$; $c_2 = 1.4388 \cdot 10^{-2}\ m \cdot {}^oK$; T is the illuminant colour temperature in ${}^oK$; λ is the wavelength in m. Given the intensity of the radiation e (depends on the power of the illuminant but is independent of surface and light colour); Plank's law gives the spectral power of the lighting source as:

$$E(\lambda) \approx e \cdot c_1 \lambda^{-5} e^{\frac{-c_2}{\lambda T}} \tag{4.2}$$

Figure 4.1 shows the relative spectral power distribution of six Plankian radiators, the relative spectral power distribution is normalised such that it has a value of 100 at a wavelength of $560nm$. The temperature of the Planckian radiator and the wavelength together determine the relative amount of radiation being emitted (colour of the light source). The temperature of a Planckian radiator is called colour temperature which uniquely specifies the colour of the source.

An important measure that can be used to describe the colour properties of a

light source is its Correlated Colour Temperature (CCT) [159]. The CCT of a source represents the colour temperature of a black-body radiator that has nearly the same colour as the source.

Correlated colour temperature

CCT is a measure of the shade of whiteness of a light source, by comparison with a blackbody. For example, the CCT of a typical incandescent lighting is 2700^oK which is yellowish-white. CCT for Halogen lighting is 3000^oK. Fluorescent lamps are manufactured to a chosen CCT by altering the mixture of phosphors inside the tube. Warm-white fluorescents have CCT of 2700^oK and are popular for residential lighting. Neutral-white fluorescents have a CCT of 3000^oK or 3500^oK. Cool-white fluorescents have a CCT of 4100^oK and are popular for office lighting. Daylight fluorescents have a CCT of 5000^oK to 6500^oK, which is bluish-white.

Another way of describing the colour properties of a typical light source (i.e. candle, lamp or sun) may be done by means of standard illuminants, such as the Commission Internationale de l'Eclairage (CIE) illuminants.

CIE illuminants:

CIE has codified the SPDs of different types of white light sources and called them CIE illuminants [160]. There is a wide range of illuminants available, which represent the general types of white illumination found indoors and outdoors, some examples of CIE illuminants are:

- **A** represents incandescent or tungsten light (2856^oK) found in the home.
- **B and C** simulate average daylight, where **B** represents non sunlight of 4874^oK and **C** represents average sky daylight of 6774^oK.
- **D** is the standardized representation of natural daylight, illuminant **D65** represents mid-day sun of 6504^oK.
- **F** represents the spectral quality of the most common fluorescent lamp found in office environment. 12 CIE fluorescent illuminants are in active use today (F1, F2,, F12), e.g. **F2** represents cool white fluorescent 4100^oK, **F7** represents a

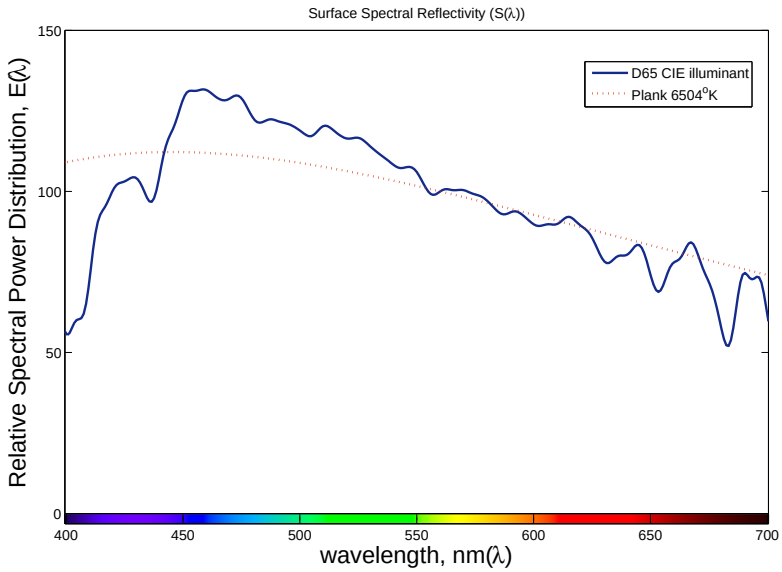


Figure 4.2: Relative spectral power distribution of CIE illuminant D65

broadband florescent lamp which simulates CIE illuminant D65 (6504^oK) and **F11** represents a narrow tri-band of 4000^oK.

Figure 4.2 shows the relative SPD of CIE illuminant $D65$ and the equivalent blackbody radiator of 6504^oK represented by Plank's formula. The CIE standard illuminants are used in the calculation of colour measurements in colourimetric software.

4.2.2 Materials

Once a material surface is hit by a bundle of light emitted by an illuminant, the electromagnetic waves may be transmitted, absorbed, or reflected back into the air. The quantities of transmitted, absorbed and reflected energy sum to the incident energy at each wavelength. Those quantities are typically measured in relative terms as fraction of the incident energy. Some materials may emit light, or fluorescence effect may occur, where the material absorbs light at specific wavelengths and then reflect light at different wavelengths. The surface's absorptance, transmittance and reflectance are obtained.

In this thesis, materials are assumed to be opaque so light transmission through the material is not considered. The materials are assumed not to be fluorescence

and so the emission is not considered to focus on reflection. The radiance, reflected light, regardless of the material types as well as the shape of the material can be described as:

$$Radiance(\lambda) = E(\lambda)S(\lambda) \tag{4.3}$$

Where λ is the wavelength, $Radiance(\lambda)$ is the SPD of the reflected light, $E(\lambda)$ is the SPD of the incident light, $S(\lambda)$ is the surface spectral reflectance of the material.

The surface spectral reflectance of a material refers to the ability of the material to reflect different spectral distributions when some kind of light shine on it. A reflectance model is the function which describes the relationship between incident illumination SPD and reflected light at a given point on a surface and at each wavelength. $S(\lambda)$ is defined as the ratio between the reflected SPD to the incident SPD. This model depends on the material composition, surface's structure and viewing geometry. For human perception, this reflectance model is a crucial factor in recognizing an object optically; since different materials will have different spectral reflectance. The Bidirectional Reflectance Distribution Function (BRDF) is the general model of reflectance, however the measurement of such model is difficult and computationally expensive, simplified models are proposed to model the reflectance for computer vision problems.

Different materials have different mechanisms of reflection, optically; most materials can be divided into two categories:

1. Homogeneous materials
2. Inhomogeneous materials

Homogeneous materials

Homogeneous materials have a uniform refractive index throughout their surface and bodies, so if a bundle of light enters its surface, it will be totally reflected and a specular only reflection is produced. Most homogeneous objects, have constant SPD over the visible wavelength, making the reflected light, $Radiance(\lambda)$, have similar SPD to the incident light $E(\lambda)$. Examples of homogeneous materials are mirrors

and polished metal.

A second important class of surfaces are the glossy or mirror-like surfaces, often referred to as specular surfaces. Radiation arriving at a specular surface along a particular direction can leave only along the specular direction, obtained by reflecting the direction of incoming illumination about the surface normal.

Relatively few surfaces are either ideal diffuse or perfectly specular. The BRDF of many surfaces can be approximated as a combination of a Lambertian component and a specular component. A model that models the interaction between light and surfaces and that takes the two components into account is discussed in the following section.

Inhomogeneous materials

On the other hand, inhomogeneous materials have varying refractive index throughout their surfaces and bodies. If a light hits its surface, part of the light reflects (specular reflection), while the other part enters the object and then reflect back to the air causing diffuse reflection. Examples of such surfaces include cloth, many carpets, matte paper and matte paints. Such material's surfaces are known as diffuse (matte) surfaces or Lambertian surfaces. The radiance leaving a lambertian surface is independent of illumination incidence angle, and such surface looks equally bright from any direction. For Lambertian surfaces, the diffuse reflectance is often called albedo.

Specularities or highlights are the bright spots on the surface of inhomogeneous materials. Relatively few materials are either ideal diffuse or perfectly specular. The reflection models discussed in Section 4.3 model reflection of opaque inhomogeneous objects as a linear combination of diffuse and specular reflections.

4.2.3 Vision system

The light emitted by sources of illumination and modulated by surfaces in the scene arrives at the capturing sensors of the colour vision system that is observing the scene. The vision system senses the captured electromagnetic signal and then transforms the information carried by light into a colour image of the physical

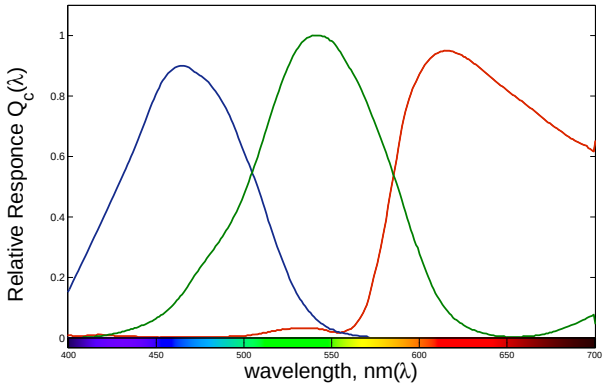


Figure 4.3: Spectral sensitivity characteristics of the Sony ICX098BQ camera [161]

world.

Colour Charge Coupled Device (CCD) camera, which is an example of a typical vision system, contains a set of sensors which convert electromagnetic energy into electric signals, which are, then sampled and quantized. Conventional CCD cameras insert colour filters, with different spectral sensitivity to the various wavelengths, over each sensory element, typically red, green, and blue filters in order to obtain colour information. The sensor Irradiance is a measure of signal that reaches the camera's sensor,

$$Irradiance(\lambda) = E(\lambda)S(\lambda)Q(\lambda) \tag{4.4}$$

where $Q(\lambda)$ is the spectral sensitivity of each of the three colour filters(Q_R, Q_G and Q_B). These sensors transform the continuous colour signal into three components by performing a spectral integration transformation between the space of spectral colours and the sensors response colour space. The colour spectral space, Ω, which has infinite dimension, is then represented by a finite-dimensional colour space.

$$I_o = \int_{\Omega} E(\lambda)S(\lambda)Q(\lambda)d\lambda \tag{4.5}$$

where I_o is the image intensity received by a given sensor of the camera, and represented in a vector of three colour components R_o, G_o, B_o.

A 3-CCD camera has three CCD sensors for each pixel element, however a single-CCD camera, captures one colour component with each CCD sensor and the two missing colour components are estimated from the adjacent existing sensors using

Bayer filters.

Figure 4.3 shows the spectral sensitivities of the Sony ICX098BQ [161] CCD sensors, excluding lens characteristics and light source characteristics, as an example of a typical surveillance camera.

Apart from the spectral sensitivity of the colour filters, the formation of digital image colour values includes other factors, such as lens characteristics, and the electronics of the camera.

Camera model

There are several noise sources in a CCD imaging systems that prevent the measurement of the actual scene intensity. Healey et al. [162] examine the effect of these noise sources and describe the following model for a single pixel recorded using a CCD camera:

$$I = g(I_o + \mu_{DC} + N_S + N_R) + N_Q \tag{4.6}$$

where I is the measured image intensity for a given sensor, g is the camera gain, I_o is image intensity received by a given sensor of the camera, μ_{DC} is an offset which represents the dark current, N_S is the shot noise, N_R is the readout noise and N_Q is the quantisation noise. The dark current μ_{DC} is constant over time. The shot noise N_S has a Poisson distribution with $\mu_S = 0$ and σ_S depends on I_o. The readout noise N_R has a Gaussian distribution with $\mu_R = 0$ and constant σ_R. The quantisation noise N_Q has a uniform distribution $U(-\frac{q}{2}, \frac{q}{2})$ where q is the smallest step in pixel value.

Gamma correction

Cathode Ray Tube (CRT) monitors have a non-linear relationship between the input voltages and the rendered intensities. To reproduce an image accurately, the image is gamma-corrected in such a way that the monitor displays the desired intensities [160]. The ITU-R Recommendation BT.709 transfer function defines gamma-compensated values $\grave{R}$, $\grave{G}$ and $\grave{B}$ from R, G and B as :

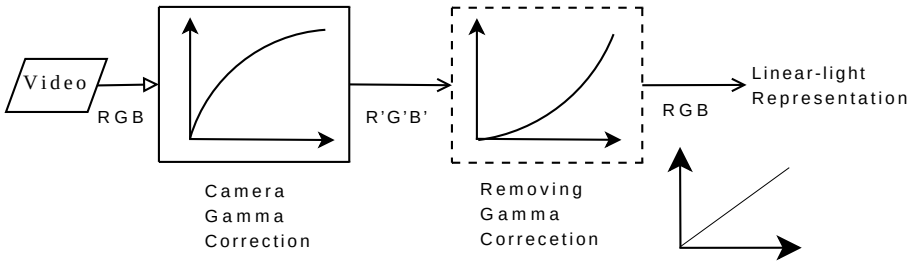


Figure 4.4: Gamma correction diagram

$$\grave{I} = \begin{cases} 4.5I & \text{if } I \leq 0.018 \\ 1.099I^{0.45} - 0.99 & \text{if } 0.018 < I \end{cases} \tag{4.7}$$

where $\grave{I}$ represents a vector of $\grave{R}\,\grave{G}$ and $\grave{B}$. Since gamma-correction is already applied by the camera as standard practice, most digital image data should be interpreted as $\grave{R}\grave{G}\grave{B}$ and not RGB. The prime symbols in this equation, and in these to follow, denote non-linear components.

In order to simulate the physical world of image formation, linear-light coding is necessary, non-linear gamma correction that is imposed by the camera has to be removed, to convert the image data back into its linear-light representation, as shown in Figure 4.4.

$$I = \begin{cases} \frac{\grave{I}}{4.5} & \text{if } \grave{I} \leq 0.081 \\ \left(\frac{\grave{I}+0.099}{1.099}\right)^{\frac{1}{0.45}} & \text{if } 0.081 < \grave{I} \end{cases} \tag{4.8}$$

Withagen et al. [163] analyze a range of different types CCD cameras for their use in measurements, and they extend the camera model in Equation 4.6 by adding a parameter h with a value in the range of $[0, 1]$ to model the effect of the iris and shutter control and another parameter γ to represent the gamma adjustment.

$$I = g^{\gamma}(h(I_o + \mu_{DC}) + N_S + N_R)^{\gamma} + N_Q \tag{4.9}$$

Using experiments, Withagen et al. [163] show that for several cameras, except a typical consumer webcam, the amount of additive noise is exceeded by the amount of

multiplicative noise at intensity values larger than $10\% - 30\%$ of the intensity range. They propose a simplification of the model under normal operating conditions, where they neglect the dark current and the additive noise.

$$\grave{I} = g^{\gamma}(h \cdot I_o + N_S)^{\gamma} \tag{4.10}$$

The image pixel colour value on which image processing tools operate represent a measure of signal that reaches the camera's sensor (Irradiance) from a point on a surface mixed with camera calibration parameters and camera noises. For image analysis, the signal that is radiated from the material surface (Radiance) is more important. Therefore, information about the radiance needs to be extracted from the image pixel value.

By combining the measured colour modeled by Equation 4.5 and the camera model in Equation 4.9, the relation between camera irradiance and material radiance may be than presented by means of Equation 4.11. In this case, the signal I_o that reaches the camera's sensor from a point on a surface is proportional to the surface reflactance.

$$I_c = g(h(\int_{\Omega} E(\lambda)S(\lambda)Q_c(\lambda)d\lambda + \mu_{DC}) + N_S \tag{4.11}$$

where I_c is represented in a vector of three Gamma-corrected camera output colour components R, G, B. Unlike Equation 4.10, the assumption used in [163] is not considered and the effect of the dark current is considered.

Equation 4.11 concludes that image pixel colour values may be interpreted as a function of physical phenomena. To this end, the first objective of the chapter, that is defining different image formation elements and linking the physical scene properties to a digital colour image, is achieved. Different image formation models and colour representation systems have been proposed in order to process colour information, those are reviewed in the next sections.

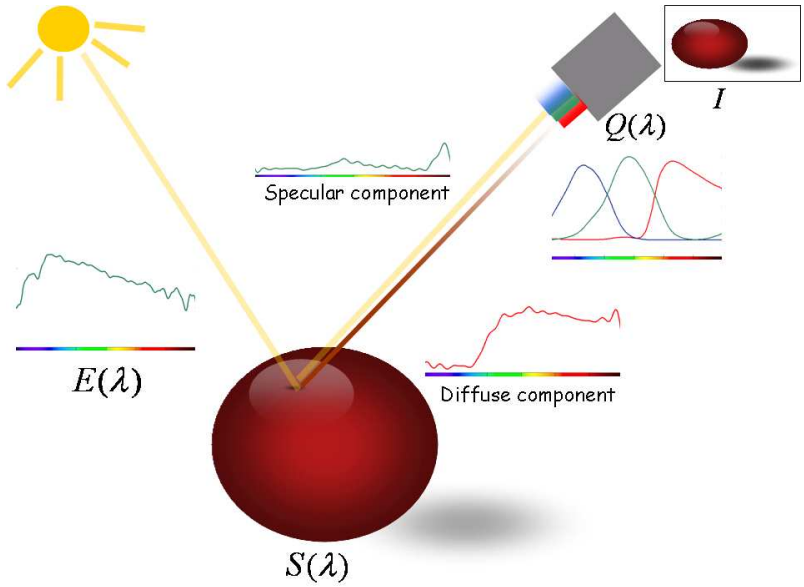


Figure 4.5: Schematic diagram of image formation

4.3 The dichromatic reflection model

There is a great number of reflection models, most of them developed in the field of computer graphics [102, 119, 120]. Among these methods, the dichromatic reflection model in [102] is a usual choice for those algorithms employing a physical model to represent colour images.

The dichromatic model represents the reflected light of an inhomogeneous dielectric material as a linear combination of diffuse and specular reflections. Each of these parts is further divided into two elements, one accounting for the geometry and another purely spectral.

$$I(\lambda, \bar{x}) = w_d(\bar{x})E(\lambda, \bar{x})S_d(\lambda, \bar{x}) + w_s(\bar{x})E(\lambda, \bar{x})S_s(\lambda, \bar{x}) \tag{4.12}$$

Where λ is the wave length, $\bar{x}$ is the position of a surface point in a 3D world coordinate system, $I(\lambda)$ is the reflected light spectral power distribution, w_d is a geometrical parameter for diffuse reflection, w_s is a geometrical parameter for specular reflection, $E(\lambda)$ is the spectral power distribution function of the illumination, $S_d(\lambda)$ is the diffuse surface spectral reflectance of the object, $S_s(\lambda)$ is the specular surface spectral reflectance of the object.

The neutral interface reflection assumption [164] states that the spectral

distribution of the specular component, $E(\lambda, \bar{x})S_s(\lambda, \bar{x})$ is similar to the spectral power distribution of the incident light $E(\lambda, \bar{x})$. So $S_s(\lambda, \bar{x})$ is assumed to be a constant scalar with respect to the wavelength $S_s(\lambda, \bar{x}) \approx K_s(\bar{x})$, so assuming $\widetilde{w}_s(\bar{x}) \approx w_s K_s(\bar{x})$, then the dichromatic model becomes:

$$I(\lambda, \bar{x}) = w_d(\bar{x})E(\lambda, \bar{x})S_d(\lambda, \bar{x}) + \widetilde{w}_s(\bar{x})E(\lambda, \bar{x}) \tag{4.13}$$

Starting with both dichromatic reflection model and camera model, the camera output (R, G and B) depends on three factors:

1. Illuminant (light source) (E)
2. Camera (sensors) spectral sensitivity characteristics (Q)
3. Spectral reflectance of the object (S)

$$\begin{aligned} R =& w_d \int_\Omega E(\lambda)S(\lambda)Q_R(\lambda)d\lambda + \tilde{w}_s \int_\Omega E(\lambda)Q_R(\lambda)d\lambda \\ G =& w_d \int_\Omega E(\lambda)S(\lambda)Q_G(\lambda)d\lambda + \tilde{w}_s \int_\Omega E(\lambda)Q_G(\lambda)d\lambda \\ B =& w_d \int_\Omega E(\lambda)S(\lambda)Q_B(\lambda)d\lambda + \tilde{w}_s \int_\Omega E(\lambda)Q_B(\lambda)d\lambda \end{aligned} \tag{4.14}$$

where Ω is the visible range from $400nm$ to $700nm$, $I(\lambda)$ is the spectral power distribution of the reflected light, w_d is a geometrical parameter for diffuse reflection, w_s is a geometrical parameter for specular reflection, $E(\lambda)$ is the spectral power distribution function of the illumination, $S(\lambda)$ is the surface spectral reflectance of the object, $Q_R(\lambda)$, $Q_G(\lambda)$ and $Q_B(\lambda)$ are the Red, Green and Blue camera sensor spectral sensitivities characteristics respectively. Figure 4.5 shows a schematic diagram of image formation mdel.

The assumptions made by the dichromatic reflection model are:

1. There is a single light source, that can be a point source or an area source;
2. The illumination has a constant SPD across the scene;
3. The amount of illumination can vary across the scene.

For what concerns the surface properties, the model assumes that:

1. The surface is opaque;
2. The surface is not optically active (no fluorescence);
3. The colorant is uniformly distributed

The assumption of illumination being due to only one source of illumination is limiting the application of such models to scenes where there is a dominant illuminant. The assumptions about the surface are typical for reflection models and not too unrealistic.

4.3.1 Shading model

Shading model is an approximated model derived from the dichromatic reflection model. For images of scenes which contain Lambertian surfaces; shading reflection model adopt two approximations to the dichromatic model: diffuse only reflection and narrowband sensor sensitivity approximation.

Diffuse only reflection approximation

Assuming Lambertian surfaces, theoretically, an image taken by a digital colour camera (for diffuse only reflection) can be described as:

$$I_c = w_d \int_{\Omega} E(\lambda)S(\lambda)Q_c(\lambda)d\lambda \tag{4.15}$$

Narrowband sensor sensitivity approximation

The second approximation that may be taken is ideal camera sensors, combined with the first approximation (diffuse only reflection). For narrowband sensor sensitivity, camera sensors are assumed ideal and are expressed using the Dirac delta function as:

$$Q_c = \delta(\lambda - \lambda_c) \tag{4.16}$$

the observed image intensity at a pixel x can then be modelled as the product of two components: the illumination $E(\lambda)$ from the light source(s) in the scene and

the reflectance $S(\lambda)$ of the object surface. Thus each colour channel becomes:

$$I_c = E(\lambda_c) \;\; S(\lambda_c) = E_c \;\; S_c \tag{4.17}$$

Intrinsic image

Only the reflectance component S contains information about the objects visible in the scene. A type of illumination-invariant feature, intrinsic image, can be derived by first filtering out the illumination component E from the image. If the illumination has lower spatial frequency content than the reflectance component S, a homomorphic filter can be used to separate the two components of the intensity signal. That is, logarithms are taken on both sides of the shading model to obtain:

$$ln(I_{c(x,y)}) = ln(E_{c(x,y)}) + ln(S_{c(x,y)}), \tag{4.18}$$

Since the lower frequency component is now additive, it can be eliminated using a high-pass filter. The intrinsic image can thus be estimated as:

$$S_{c(x,y)} = e^{F(ln(E_{c(x,y)})+ln(S_{c(x,y)}))}, \tag{4.19}$$

where $F(\cdot)$ is a high-pass filter. The intrinsic image can be provided as input to the decision rule step of a change detection process.

4.3.2 Linear models

Several researchers [105–107] show that both illumination and surface spectral reflectance are relatively smooth functions of the wavelength of light in the visual spectrum and that they can be expressed using finite-dimensional linear models. Judd et al. [126] showed that the daylight could be accurately described by a linear mixture of three basis functions.

$$E(\lambda) \approx \sum_{i=1}^{3} e_i E_i(\lambda) \tag{4.20}$$

The surface reflectances of a great variety of materials have been studied.

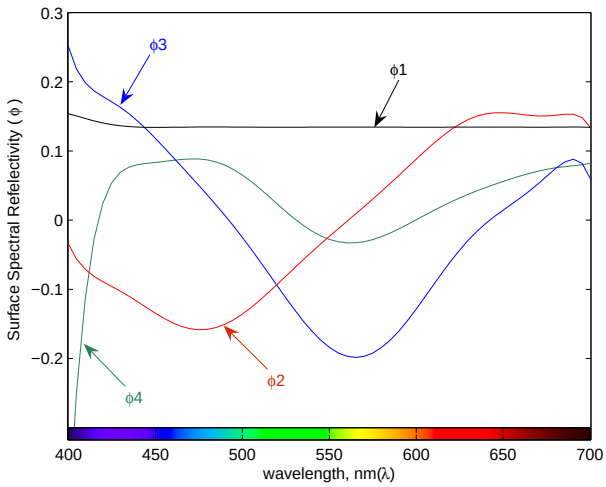


Figure 4.6: Parkkinen's first 4 basis functions [165]

Parkkinnen et al. [125], Maloney [106] studied the reflectance properties of the Munsell chips, which is a database of experimentally measured surface spectral reflectance characteristics. Parkkinen concludes that 8 basis functions can cover almost all existing data in Munsell chips database [166]. However, it has been shown in the literature [20] that the spectral reflectance calculated using the first three basis functions has average error 0.0055 and 0.01. Figure 4.6 shows the first 4 Parkkinen basis functions. As it can be seen from the figure, a significant characteristic of Parkkinen basis functions is the flat distribution of his first basis function.

$$S(\lambda) \approx \sum_{i=1}^{n} w_i \phi_i(\lambda) \tag{4.21}$$

where $\phi_i(\lambda)$ is the i^{th} reflectance linear basis function, and w_i is its corresponding weight, n is the number of basis functions used.

Using finite-dimensional linear models to represent the surface spectral reflectance provides a compact description of data, with few basis functions we can represent surface spectral reflectance for general materials. The linear models have been extensively used in some colour constancy algorithms where the main aim was to recover either the illumination or the reflectance functions, or both of them. Main works in such topics are those of Tominaga and Wandell [116], Marimont and Wandell [107].

Illuminant estimation

The aim of illuminant estimation is to imitate the human ability to separate the illumination PSD from the surface reflectance. Several approaches have been proposed, this includes methods relying on linear models [19], neural networks [21], reliance on highlights and mutual reflection [20], and Bayesian and probabilistic approaches [22].

From a mathematical point of view the problem of estimating the illuminant is under-determined, using a camera which gets only three spectral samples. If there are enough surfaces in the scene, then, in theory, the illuminant can be reliably estimated, and therefore, in turn, can surface reflectance. For example, the scene can be assumed to be colourimetrically unbiased, a uniform background that was everywhere set to the average, adaptation reflectance, that is, no particular colour predominates (equivalent background hypothesis or the grey 'grey-world' assumption). The spatial average colour then coincides with the illuminant colour [167].

The Grey-world assumption is one of the earliest developed, and is based upon the assumption that the spatial average of surface reflectances in a scene is achromatic. Since the light reflected from an achromatic surface is changed equally at all wavelengths, it follows that the average of the light leaving the scene is the colour of the unknown iluminant [127]. The implied diagonal transform is simply the ratio of the average grey of the image illuminated under the canonical to that of the unknown.

The Retinex model uses similar assumptions to those of the Grey-world algorithm. This method assumes the presence of a white patch, from which the chromaticity of the illuminant is perfectly preserved. The maximum of the channel responses is assumed to arise from the reflectance of a white surface, and is subsequently used in the computation of the diagonal transform. The method is susceptible to specularities, since they can produce a maximum reflectance easily greater than that of a pure white surface. However, priori knowledge of the presence of specularities is beneficial as they preserve more of the illuminant chromaticity.

Gamut-mapping is an alternative technique used to determine an unknown illuminant by recovering the transform which best projects the measured gamut into that of the canonical. The diagonal transforms are determined by computing the convex hull of both the canonical and unknown illuminant reflectances, computing the set of mappings which take individual hull points from the unknown hull onto those of the canonical hull, and choosing the best transform from the intersection of all the transform sets. Each variant of the gamut mapping technique can be differentiated by their process of selecting the diagonal transform from the feasible set. The feasible set is further constrained by restricting the amount in which the illumination can vary [127].

McCamy et a. [159] derived a simple equation to compute correlated colour temperature from CIE 1931 chromaticity coordinates x and y, which is useful in designing sources to simulate CIE illuminants.

$$n = \frac{x - x_e}{y_e - y} \tag{4.22}$$

where $x_e = 0.3320$ and $y_e = 0.1858$ Then

$$CCT = 449n^2 + 3525n^3 + 6823.3n + 5520.33 \tag{4.23}$$

This equation proves useful for implementation in real-time applications. It was derived from the assumption that CCT may be represented by a third-order polynomial function of the reciprocal of the slope of the line from specular pixels to the chromaticity of the light. The method is based on the fact that the isotemperature lines for CCTs of principal interest nearly converge towards a point on the chromaticity diagram. McCamy's method has a maximum absolute error of less than 2^oK for colour temperatures ranging from 2856^oK to 6500^oK (corresponding to CIE illuminants A through D65) [168].

4.4 Colour spaces

A colour space is a mathematical representation of continuous spectral colours in a three-dimensional vector space, using three primary colours, which allows colour analysis and manipulation. Different colour spaces can be derived by defining

different primary colours.

Colour spaces are developed according to output devices, physiological or physical properties. A number of standard colour space specifications are well developed. However, different definitions can often be found for the same colour space. For more information about colour spaces, the reader is referred to [169] for a detailed review.

In this section, the main colour spaces that have been proposed for change detection problem, reviewed in Chapter 3, are discussed. Colour spaces can be classified by their device dependency, e.g. device independent colour components are the same on all output devices. On the other hand, device dependant colour spaces will have different components for different output devices. Colour spaces can be also classified as user-oriented models which try to build a bridge between the user and the hardware used to manipulate colour. However in this section a classification of colour spaces depending on the transformation from the camera output is used. This presentation follows a classification of colour spaces in two groups:

1. Linear colour spaces

2. Nonlinear colour spaces

4.4.1 Linear colour spaces

A linear colour space can be defined by a linear transformation of the colour matching functions of another. Such a transformation provides a corresponding invertible linear mapping between the tristimulus component vectors in the two spaces.

RGB colour space

The RGB camera output maps spectral power densities to a triple of numerical components that are the mathematical coordinates of R, G and B. It is device-dependent colour space used for capture and display devices. It is an additive colour system. RGB space may be modelled as a cube with the three axes corresponding to red, green and blue. The bottom corner, when red = green = blue = 0 is black,

while the opposite top corner, where red = green = blue = 255 (for an 8-bit per channel display system), is white. The RGB colour space suffers from the high correlation among the R, G and B components; if the intensity changes, all the three components will change accordingly. Therefore, the measurement of a colour in RGB space does not represent colour differences in a uniform scale; it is difficult to evaluate the similarity of two colours from their distance in RGB space. RGB space is used in some change detection algorithms since no transform is required.

Grey-scale

grey-scale represents the 1D projection of the three channels of colour images that is regarded as a simple way of combining colour information. In Rec. 709 [169] standard, the weights to compute true CIE luminance from linear red, green and blue (indicated without prime symbols) are:

$$Y_{709} = 0.2125R + 0.7154G + 0.0721B \tag{4.24}$$

To compute non-linear video luma ($\grave{Y}_{601}$) BT. 601-4 (formerly CCIR Rec. 601) from non-linear red, green and blue:

$$Y_{601} = 0.299\grave{R} + 0.587\grave{G} + 0.114\grave{B} \tag{4.25}$$

YUV colour space

The YUV colour space separates RGB into luminance and chrominance information [169], by subtracting *Luma* ($\grave{Y}$) from non-linear blue to form ($\grave{B} - \grave{Y}$) and by subtracting luma ($\grave{Y}$) from non-linear red (to form ($\grave{R} - \grave{Y}$)), *Chroma*.

$$\begin{bmatrix} \grave{Y} \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.299 & -0.587 & 0.886 \\ 0.701 & -0.587 & -0.114 \end{bmatrix} \begin{bmatrix} \grave{R} \\ \grave{G} \\ \grave{B} \end{bmatrix} \tag{4.26}$$

with $\grave{Y}$ in the range of $[0, 1]$ and U, V, in the range of $[-0.5, +0.5]$

YCbCr colour space

The $YCbCr$ colour space separates RGB into luminance and chrominance information [169]. Luminance information is stored as a single component ($\grave{Y}$), and chrominance information is stored as two colour-difference components (Cb and Cr). Cb represents the difference between the blue component and a reference value. Cr represents the difference between the red component and a reference value. Digital video widely uses this colour space. Given R, G and B in the range of $[0, 1]$,

$$\begin{bmatrix} \grave{Y} \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 65.841 & 128.553 & 24.966 \\ -37.797 & -0.587 & 112 \\ 112 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} \grave{R} \\ \grave{G} \\ \grave{B} \end{bmatrix} \tag{4.27}$$

with $\grave{Y}$ in the range of $[16, 235]$ and Cb, Cr, in the range of $[16, 240]$.

4.4.2 Non-linear colour spaces

HSV colour Space

The HSV is a non-linear transformation from the RGB camera output, which is more intuitive to human vision. It separates colour information of an image from its intensity information. Colour information is represented by Hue and $Saturation$ values, while $Intensity$ describes the brightness of an image pixel. Hue represents basic colours, and is determined by the dominant wavelength in the spectral distribution of light wavelengths. It is the location of the peak in the spectral distribution. The $Saturation$ is a measure of the purity of the colour, and signifies the amount of white light mixed with the Hue. It is the height of the peak relative to the entire spectral distribution.

$$(H, I, S) = (arctan(\frac{\sqrt{3}(R-B)}{(2R-G-B)}), \frac{R+G+B}{3}, 1 - \frac{min(R,G,B)}{I} \tag{4.28}$$

$$H = arctan(\frac{\sqrt{3}(\int_\Omega E(\lambda)S(\lambda)(Q_R(\lambda) - Q_B(\lambda))d\lambda)}{\int_\Omega E(\lambda)S(\lambda)(2Q_R(\lambda) - Q_G(\lambda) - Q_B(\lambda))d\lambda)}) \tag{4.29}$$

$$S = 1 - \frac{3 \cdot min(\int_\Omega E(\lambda)S(\lambda)Q_R(\lambda)d\lambda, \int_\Omega E(\lambda)S(\lambda)Q_G(\lambda)d\lambda, \int_\Omega E(\lambda)S(\lambda)Q_B(\lambda)d\lambda)}{\int_\Omega E(\lambda)S(\lambda)(Q_R(\lambda) + Q_G(\lambda) + Q_B(\lambda))d\lambda} \tag{4.30}$$

The *Hue* and *Saturation* are dependant on the absolute illuminant power distribution, the sensors sensitivities and the surface spectral reflectance. However, they are insensitive to surface orientation changes, illumination direction changes and illumination intensity changes. Hue is invariant to certain types of highlights, shading, and shadows, therefore it can be used for segmentation in 1D space if the saturation is not low.

The main drawback of the hue that it has a nonremovable singularity near the axis of the colour cylinder, where a slight change of input R, G and B values can cause a large jump in the transformed values. Hue and Saturation play little role in distinguishing colours if the intensity of the colour lies close to white or black. The singularities may create discontinuities in the representation of colours. The hue values near the singularity are numerically unstable. That is why pixels having low saturation are left unassigned to any regions in many segmentation algorithms.

There are some variants of HSI colour spaces, such as HSB *Hue Saturation Brightness*, HSL *Hue Saturation Lightness*, and HSV *Hue Saturation Value*.

rgb colour space

The rgb colour space attempts to represent the real colour information of an image pixel and to be invariant to the brightness of the image. This colour space normalises the variations of intensities uniformly across the spectral distribution in order to make colours invariant to the change in lighting intensities. It is a non-linear transformation from the RGB camera output. The rgb colour space is formulated as:

$$(r, g, b) = (\frac{R}{R+G+B}, \frac{G}{R+G+B}, \frac{B}{R+G+B}) \tag{4.31}$$

This transformation projects a colour vector in the RGB cube into a point on the unit plane. Two of the rgb components suffice to define the coordinates of the colour point in this plane. Since rgb is redundant ($b = 1-r-g$), the generalised normalised colour is given by

$$(I, r, g) = (\alpha R + \beta G + \gamma B, \frac{R}{R+G+B}, \frac{G}{R+G+B}) \tag{4.32}$$

Where α, β and γ are chosen constant such that $\alpha + \beta + \gamma = 1$. I represents the intensity component, while r and g are chromatic components of an image pixel. By substituting Equation 4.15 in 4.31,

$$\begin{aligned} r =& \frac{w_d \int_\Omega E(\lambda)S(\lambda)Q_R(\lambda)d\lambda}{w_d \int_\Omega E(\lambda)S(\lambda)(Q_R(\lambda)+Q_G(\lambda)+Q_B(\lambda))d\lambda}) \\ g =& \frac{w_d \int_\Omega E(\lambda)S(\lambda)Q_G(\lambda)d\lambda}{w_d \int_\Omega E(\lambda)S(\lambda)(Q_R(\lambda)+Q_G(\lambda)+Q_B(\lambda))d\lambda}) \\ b =& \frac{w_d \int_\Omega E(\lambda)S(\lambda)Q_B(\lambda)d\lambda}{w_d \int_\Omega E(\lambda)S(\lambda)(Q_R(\lambda)+Q_G(\lambda)+Q_B(\lambda))d\lambda}) \end{aligned} \tag{4.33}$$

Equation 4.33 shows that the three components, r,g, and b are dependant on the absolute illuminant power distribution, the sensors sensitivities and the surface spectral reflectance, and are insensitive to surface orientation (viewing geometry) changes, illumination direction (local illumination) changes and illumination intensity changes. The rgb colour space is relatively robust to the change of the illumination; however the normalization reduces the sensitivity of the distribution to the colour variability. It is more effective than RGB coordinates for suppressing unwanted changes due to shadows. But an obvious shortcoming of normalised RGB is that the normalised colours are very noisy if they are under low intensities [78]. This is due to the nonlinear transformation.

4.5 Summary

The transformation from real scenes to digital images depends on four main elements: an illuminant, a medium, a material and a vision system. Colour values are the integration of the product of incident illumination spectral characteristics,

intrinsic object surface spectral reflectance and camera sensors sensitivities. Geometrical features represent the scene structure, the illuminant orientation, the surface roughness and the viewing geometry. Spectral and geometrical features which characterise the elements of the scene appearance combine non-linearly to yield digital image.

This chapter shows that among the various reflection models, which describe the image formation process, proposed in the literature; the dichromatic reflection model represents a practical choice for image processing and computer vision approaches. The dichromatic model represents the reflected light of an inhomogeneous dielectric material as a linear combination of diffuse and specular reflections. Each of these parts is further divided into two elements, one accounting for the geometry and another purely spectral. The main limitation of this technique is that it can be applied only to inhomogeneous dielectrics.

Colour constancy is the ability to separate the illumination power spectral distribution from the surface spectral reflectance. Extracting information about surface reflectance is a necessary condition for colour constancy. Because a surface reflectance spectrum cannot be recovered uniquely from three colour components, perfect extraction of a surface spectral reflectance is physically impossible. The colour constancy techniques are intorduced with the intention of using them in the context of change detection and foreground segmentation.

By applying the dichromatic reflection model for Lambertian surfaces, three main photometric invariants which are sensitive to changes in surface reflectance can be extracted, reflectance ratio, intrinsic image and surface spectral reflectance (using a linear model). Reflectance ratio and intrinsic image are derived from the shading model, which represents an approximated form of the dichromatic model, assuming sharp sensors.

This chapter shows that understanding the definition of different colour spaces with respect to its physical basis; is important to identify the most suitable photometric invariant feature for image representation.

"In our endeavor we came to learn that good laboratory technology should be supported by deep knowledge of the business, market, and user realities to become a success story. Actually, we can now corroborate that in certain cases technology transfer can be as challenging as the basic research that precede it."

C. Regazzoni, V. Ramesh, and G. Foresti [1]

5

Proposed Physics-Based Change Detection Algorithms

5.1 Introduction

The problem of change detection and foreground segmentation has been investigated within several research domains. The accurate segmentation of moving objects in video sequences represents a key process for smart and automated video surveillance systems. Consequently, the development of robust and flexible methods for the automatic segmentation of moving objects is of primary interest.

The workplace surveillance framework proposed in Chapter 2 is one of the applications where important objects are the moving objects. The automatic segmentation of such moving objects is possible through change detection. An

accurate segmentation is especially required where the objects and their movements are analysed to obtain information from the scene.

In Chapter 3, the image representation for change detection was reviewed. The review shows that different colour spaces have been proposed to represent the image in the majority of change detection algorithms. Most of the work has been the application to the three-dimensional colour space of algorithms originally developed and used for analysing grey-scale images. However several noise components limits the capabilities of such approaches. The main sources of noise are camera noise, illumination variation and texture similarity between foreground objects and background. Section 4.2.3 shows that the camera model suffers from various additive and multiplicative sources of noise, which can be represented by different distributions. The complexity of the model along with the diversity of such distributions makes it difficult for both parametric and non-parametric statistical modelling techniques, reviewed in Chapter 3, to discount for all these types of noise. Statistical modelling techniques are more suitable to model the periodic background motions in outdoor applications.

The physical approach used in the colour constancy field studies colour image formation while taking into consideration the characteristics of camera sensors, in order to estimate scene illumination and to extract the surface spectral reflectance of objects. Image formation models which aim to capture the full surface reflectance spectra are reviewed in 4.3.2. A gap, in the literature of change detection, is that the use of such physical models have not been fully investigated.

Linear models for image formation has been applied and proven useful for a number of applications [108, 109], nevertheless these models have not been applied yet in the field of change detection. The reason appears to be the computational complexity of such models, and hence unfeasibility of real-time implementation. However, the linear models developed in [105–107] to estimate the surface spectral reflectance give a good starting point toward the solution for such a problem. This thesis is the first attempt to investigate the use of such models in change detection.

Change detection based on physical models would permit image representation via the study of the process of image formation. Algorithmically, the basics of these

approaches are often very similar to those of other change detection methods, and only differ from them in the fact that these algorithms explicitly use the reflectance models of surfaces to represent a colour image.

This chapter starts by investigating the challenges of using models of image formation as a basis for change detection algorithms which target smart video-based workplace surveillance application. Two methods for change detection and the segmentation of foreground objects for indoor environments are proposed. The adopted strategy exploits physics-based features and namely relies on surface reflectance based properties of objects and is designed to be able to work automatically when illumination and scene's characteristics are unknown.

This chapter is devoted to the description of the proposed methodologies and of the adopted algorithmic solutions. An evaluation of the performance of the proposed algorithms to a number of test sequences and through the comparison with one of the state-of-the-art algorithms will be provided in Chapter 6.

5.2 Challenges

There are many challenges which face the use of physics-based image formation models as a basis for change detection algorithms in order to target indoor workplace surveillance applications. This can be divided into two different types. Firstly, challenges imposed by the operational requirements of the workplace application. Secondly, challenges related to the mathematical complexity of the reflection models.

The first type of challenges is related to the development of robust change detection techniques which can deal effectively with different indoor environments, where the surveillance task is done within confined areas, and the camera is near from moving objects. Robustness means the algorithm ability to operate for extended periods of time, with no or minimum parameter adjustments. The algorithm should adapt itself automatically with changes in the workplace environment, to cope with changes in lighting, scene geometry and scene activity without missing a significant number of events. The algorithm needs to operate in real-time.

The second type of challenges is related to the application of reflection models and colour constancy techniques in a real-time video surveillance application. There are three primary considerations:

1. The choice of a suitable reflection model.
2. The variability of surface spectral reflectance.
3. How to represent a reference of the scene surface reflectance under varying illumination conditions.
4. The choice of suitable colour constancy techniques.

These considerations are the main focus of this chapter. There are different reflection models proposed in the literature, as discussed in Section 4.3, however the choice of the suitable reflection model for such real-time change detection algorithm is not easy. There is a need to consider the assumptions of different models and to adopt the model with the set of assumptions which do not contradict with the real-world operational conditions.

The other consideration is the variability of surface spectral reflectance for different materials. Ideally, the algorithm should be able to model all types of materials contained in a workplace environment. However the available models are trained for only a limited number of materials.

One of the main problems is the illumination variation and how to represent a reference of the scene under varying illumination conditions. This leads to the challenge of estimating the illumination in real-time, and the choice of the suitable colour constancy technique which can ensure a rapid and convenient estimation of the illumination and extraction of the surface spectral reflectance.

5.3 Change detection algorithm based on reflectance ratio (Algorithm 1)

5.3.1 Background theory

This section describes a new change detection technique which models illumination variation using the ratio between foreground pixels and background pixels, based on reflectance ratio, assuming stationary camera. This method is based on shading model. The assumptions which has been made are that the scene contains only inhomogeneous materials, diffuse only reflections are considered and camera sensitivities are sharp. The Mahalanobis distance is chosen as a measure to segment foreground and shadows from static background. This method is an extension to the method proposed by Wren et. al. [76], but instead of using the YUV as an image representation, the reflectance ratio is proposed.

This method uses the RGB colour space and calculates the ratio (for each colour component) between a representative frame (i.e. background median) from a set of static background frames, and each new video frame. Statistical means, variances and covariances between ratio triples are calculated to build a background model; Mahalanobis distance is then calculated and several thresholds are determined to classify the pixel to one of the following types: foreground, background, and shadow.

5.3.2 The algorithm

Figure 5.1 shows a block diagram of Algorithm 1, the algorithm consists of three phases, the training phase, the testing phase and the post-processing phase.

Training phase

In the training phase, the background model and the covariance matrix for the reflected-intensity ratios from a pool of background frames are built (per pixel).

Starting with the shading model, discussed in Section 4.3.1, the intensities of the background model $R_{background}$, $G_{background}$ and $B_{background}$ and the new frame under

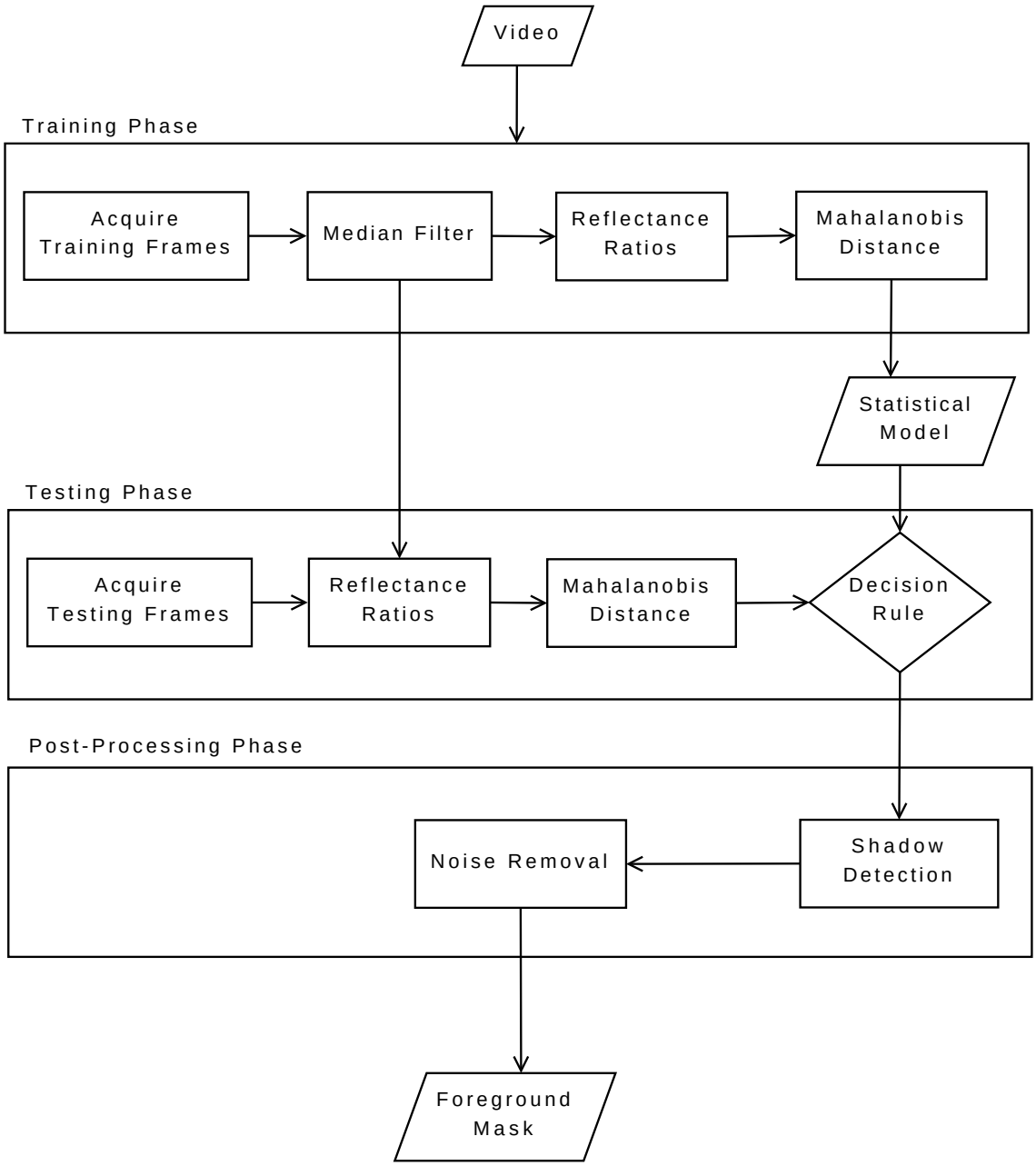


Figure 5.1: Algorithm 1 block diagram

investigation R_{new}, G_{new} and B_{new}, could be expressed as follows:

$$I^i_{background} = E^i_{background} \cdot S^i_{background} \tag{5.1}$$

$$I^i_{new} = E^i_{new} \cdot S^i_{new} \tag{5.2}$$

where I represents the pixel value at location (x, y), but the subscript (x, y) is removed for clarity, i refers to one of the colour components Red, $Green$, or $Blue$.

The main idea of this method is that the surface reflectance information could be eliminated by taking the ratio of the shading model of each new frame and a background representative frame (background model), I^i_{median}, where:

$$I^i_{median} = median(I^i_{background}) \tag{5.3}$$

If both the background model and the new video frame correspond to pixels which have got the same intrinsic colour information, then the reflectance ratios RR are:

$$RR^i = \frac{I^i_{new}}{I^i_{median}} = \frac{E^i_{new} \cdot S^i_{new}}{E^i_{median} \cdot S^i_{median}} \tag{5.4}$$

If the new pixel belongs to the background, the illumination might vary but the surface reflectance remains constant, then $S^i_{new} = S^i_{median}$ for all i, so the RR^i are:

$$RR^i = \frac{I^i_{new}}{I^i_{median}} = \frac{E^i_{new}}{E_{i_{median}}} \tag{5.5}$$

These three variables, RR^R, RR^G and RR^B, are statistically correlated, so the Mahalanobis distance D has been chosen to be the statistical metric. A pixel with small Mahalanobis distance is classified as background, otherwise the pixel is classified to either foreground or shadow.

$$D^2 = (RR - \mu_{RR})^T \sigma^{-1}_{RR} (RR - \mu_{RR}) \tag{5.6}$$

where μ_{RR} and σ_{RR} represent the mean and the covariance matrix of the reflectance ratios, respectively.

Testing phase

In the testing phase, for each new video frame, the reflectance ratios are calculated followed by Mahalanobis distance (per pixel). If the new pixel belongs to the foreground (a new object), then $S^i_{new} \neq S^i_{median}$ for all i, then the RR^i are :

$$RR^i = \frac{I^i_{new}}{I^i_{median}} = \frac{E^i_{new} \cdot S^i_{new}}{E^i_{median} \cdot S^i_{median}} \tag{5.7}$$

The ratio now depends on both illumination and surface reflectance and does not belong anymore to the statistical model of the background; higher Mahalanobis distance. Then an initial FG_{Mask} is calculated:

$$FG_{Mask} = \begin{cases} 1 & \text{if } D > \tau_{fg} \\ 0 & \text{otherwise} \end{cases} \tag{5.8}$$

where τ_{fg} is chosen using a threshold selection criteria discussed in Chapter 6.

Post-processing phase

Shadow detection

After the initial foreground mask FG_{Mask} has been created from the new frame by the testing stage, a shadow detection stage is required to differentiate between foreground pixels and background pixels which are in the shadow. The shadow detection approach adopts both the conventional condition of the drop in the pixel's colour in addition to an implicit range of Mahalanobis distance. A set of thresholds τ_{sh}, τ_1 and τ_2 are applied which depend on the intensity of the pixel and the Mahalanobis distance. A pixel is then classified into one of two categories background (BG_{Mask}) or shadow (SH_{Mask}) by the following decision procedure:

$$SH_{Mask} = \begin{cases} 1 & \text{if } D > \tau_{sh} \text{ AND } \tau_1 < (RR^i - \mu^i_{RR}) < \tau_2 \\ 0 & \text{otherwise} \end{cases} \tag{5.9}$$

$$BG_{Mask} = \text{NOT } (\ FG_{Mask} \text{ OR } SH_{Mask}\ ) \tag{5.10}$$

where τ_{sh}, τ_1 and τ_2 are chosen threshold empirically.

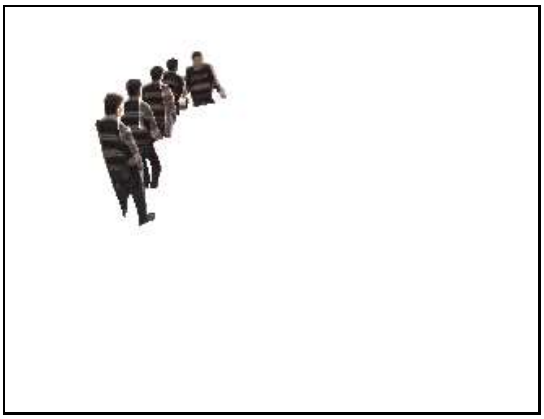
Noise removal

A noise removal stage then follows the shadow detection stage and applies a dilation step which performs a dilation operation thereon to expand or thicken the initial foreground mask FG_{Mask}. The dilation operation is the union of all translations of the mask by the structuring element SE. Then an erosion step performs an erosion operation on the resultant mask, which reduces the size of an object by eliminating area around the object edges. After the erosion operation has been completed; the detection stage performs a connected component blob analysis on the foreground mask FG_{Mask}. The erosion operation removes noise and eliminates foreground mask details smaller than the structuring element, where the size of the structuring element for both dilation and erosion is: $SE = 2$ pixels.

Sample Results for the algorithm are shown in Figure 5.2. Chapter 6 will discuss the results, and a comparison between this technique and one of the state-of-the-art techniques will be carried out.

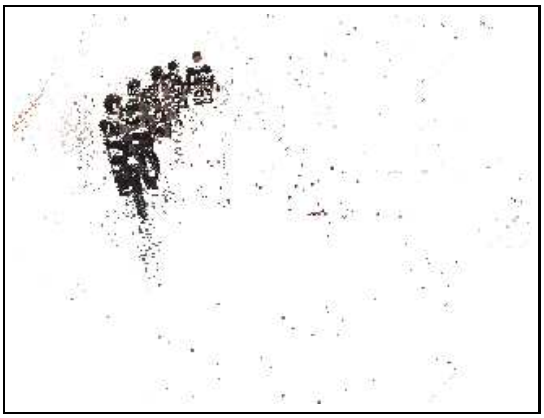
Algorithm 1 Change detection based on reflectance ratio

Input: N-background frames
Output: Foreground mask
{Modelling Phase}
Step 1: Initialization. Calculate the median of n-background frames.
Step 2: Division. Divide each frame from the N-background frames by the median.
Step 3: Background Modelling. Calculate the covariance matrix for the reflectance ratios calculated in Step 2 above.
{Testing Phase}
Step 4: Initialization. Retrieve the covariance matrix and the median frame produced in the background modelling phase.
Step 5: Division. Divide each new frame by the median frame (per pixel).
Step 6: Statistical Manipulation. Calculate Mahalanobis distance for these three ratios, using Equation 5.6.
Step 7: Thresholding. Compare the Mahalanobis distance to the threshold τ_{fg}, using Equation 5.8
Step 8: Shadow Detection. Compare both Mahalanobis distance to the threshold τ_{sh} and the intensities to the thresholds τ_1 and τ_2 to obtain the shadow, using Equation 5.9.

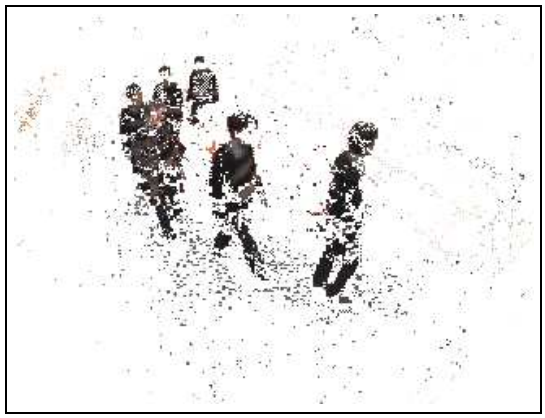


(a) Ground truth (part1)

(b) Ground truth (part2)



(c) Output without noise removal (Part1)



(d) Output without noise removal (Part2)

(e) Output with noise removal (Part1)

(f) Output with noise removal (Part2)

Figure 5.2: Samples of Algorithm 1 results for 'Intelligent Room' sequence

5.4 Change detection algorithm based on surface spectral reflectance (Algorithm 2)

5.4.1 Background theory

This section introduces a novel change detection technique, based on image formation models. Motivated by the dichromatic colour reflectance model, this method presents a new approach to change detection using explicit hypotheses about the physics that create images. The assumptions which have been made are that diffuse-only-reflection is applicable, and the existence of a dominant illuminant.

This approach is different from all the previous work, in that it relies on models, which can represent wide classes of surface materials. It makes use of the pre-trained linear Surface Spectral Reflectance (SSR) models, i.e. Parkkinnen basis functions, to represent the SSR of the objects in the scene. The method proposes a change detection that tries to identify image pixels exhibiting similar colour, but possibly varying intensity.

The rationale behind this approach is that the segmentation between foreground and background objects can be done through the matching between the SSR over the visible wavelengths of a reference background frame and each new frame. A possible implementation could use the value of correlation as the metric of the similarity between the pixels in the new frame with corresponding pixels in the background. A computational physical model and methods used to calculate the model is discussed first.

Surface spectral reflectance calculations

Figure 5.3 shows the block diagram of the SSR recovery module. this module is responsible for the calculations of the SSR. In order to build a computational physical model, discussed in Section 4.3.2, we start with the dichromatic model:

$$I_c = w_d \int_\Omega E(\lambda) S(\lambda) Q_c(\lambda) d\lambda + \widetilde{w}_s \int_\Omega E(\lambda) Q_c(\lambda) d\lambda \tag{5.11}$$

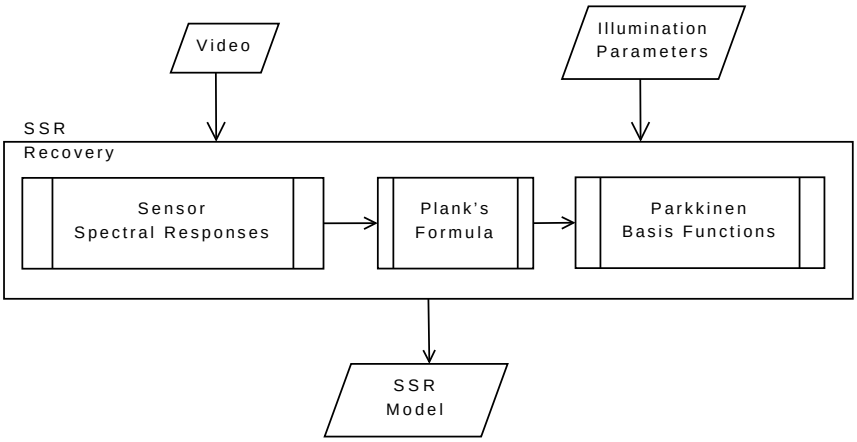


Figure 5.3: Surface spectral reflectance recovery block diagram

The first 3 basis functions of Parkkinen ($n = 3$) are used, shown in Figure 4.6, and the SSR is represented with :

$$S(\lambda) \approx \sum_{i=1}^{3} w_i \phi_i(\lambda) \tag{5.12}$$

The characteristic of the camera sensors shown in Figure 4.3 are used to represent the camera sensitivities. The SSR recovery module aim is to calculate the weights of Parkkinen basis functions, to obtain the SSR of the object, the model could be rewritten now as:

$$I_c = w_d \int_\Omega E(\lambda) \left(\sum_{i=1}^{3} w_i \phi_i(\lambda) \right) Q_c(\lambda) d\lambda + \widetilde{w}_s \int_\Omega E(\lambda) Q_c(\lambda) d\lambda \tag{5.13}$$

or

$$I_c = \int_\Omega E(\lambda) \left(\sum_{i=1}^{3} \widetilde{w}_i \phi_i(\lambda) \right) Q_c(\lambda) d\lambda + \widetilde{w}_s \int_\Omega E(\lambda) Q_c(\lambda) d\lambda \tag{5.14}$$

where

$$\widetilde{w}_i = w_d w_i, i = 1, 2, 3$$

Knowing that the first basis function of Parkkinen is constant so $\phi_1(\lambda) = K_\phi$, which

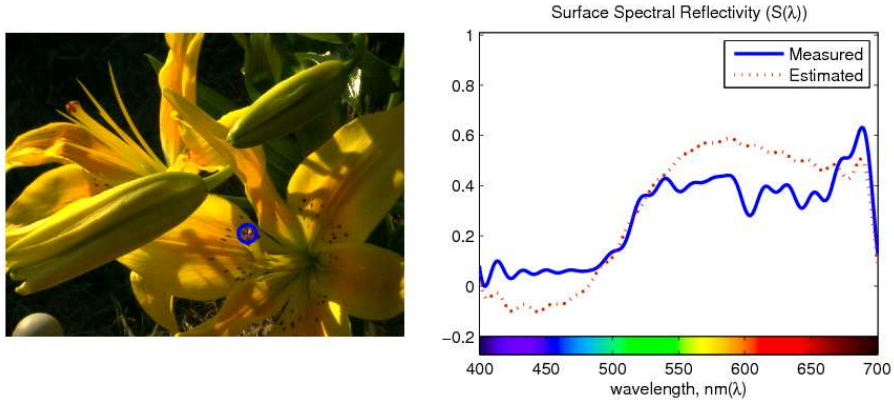


Figure 5.4: Reconstructed spectral reflectance for one pixel (Foster data set 1)

then could be merged with the specular component to give:

$$I_c = (\widetilde{w}_1 k_\phi + \widetilde{w}_s) \int_\Omega E(\lambda) Q_c(\lambda) d\lambda + \int_\Omega E(\lambda) \left(\sum_{i=2}^{3} \widetilde{w}_i \phi_i(\lambda) \right) Q_c(\lambda) d\lambda \tag{5.15}$$

by taking

$$X_{ic} = \int_\Omega E(\lambda) \phi_i(\lambda) Q_c(\lambda) d\lambda, i = 1, 2, 3 \tag{5.16}$$

These integrations are calculated to obtain the transformation matrix:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} X_{1R} & X_{2R} & X_{3R} \\ X_{1G} & X_{2G} & X_{3G} \\ X_{1B} & X_{2B} & X_{3B} \end{bmatrix} \begin{bmatrix} \widetilde{w}_1 \\ \widetilde{w}_2 \\ \widetilde{w}_3 \end{bmatrix} \tag{5.17}$$

Now the weights of the basis functions can be obtained from RGB values by:

$$\begin{bmatrix} \widetilde{w}_1 \\ \widetilde{w}_2 \\ \widetilde{w}_3 \end{bmatrix} = \begin{bmatrix} X_{1R} & X_{2R} & X_{3R} \\ X_{1G} & X_{2G} & X_{3G} \\ X_{1B} & X_{2B} & X_{3B} \end{bmatrix}^{-1} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{5.18}$$

In order to validate the estimated surface spectral reflectance, a set of measured SSR provided by Foster [167] are used. Figure 5.4.1 presents both measured and estimated surface spectral reflectance of one pixel. The result shows that the estimation is close to the measured one. Appendix A shows more results (see Figures A.1, A.2, A.3 and A.4).

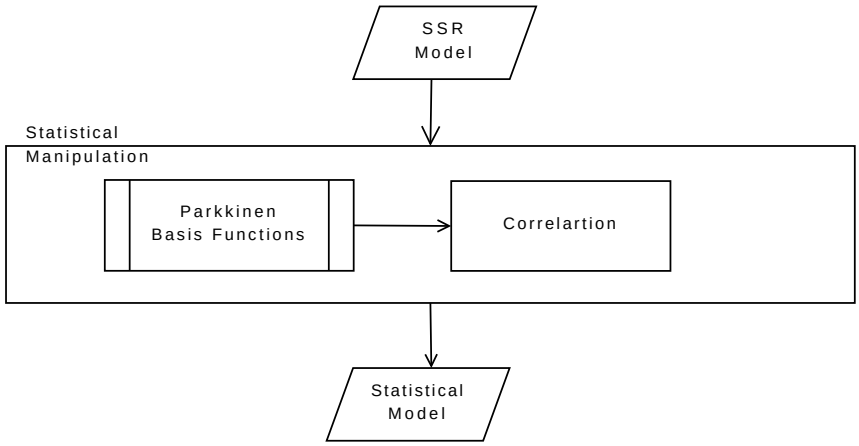


Figure 5.5: Statistical manipulation block diagram

Figure 5.5 shows the block diagram of the statistical manipulation module. This module role is to measure the deviation of SSR of two objects, whether they have the same material and colour or not, the statistical manipulation (matching) between their spectral reflectance functions can be used. Hence, The second transformation is the correlation between all three Parkkinen basis functions.

$$\begin{bmatrix} K_{11} \\ K_{22} \\ K_{33} \\ K_{12} \\ K_{13} \\ K_{23} \end{bmatrix} = \begin{bmatrix} Corr(\phi_1, \phi_1) \\ Corr(\phi_2, \phi_2) \\ Corr(\phi_3, \phi_3) \\ Corr(\phi_1, \phi_2) \\ Corr(\phi_1, \phi_3) \\ Corr(\phi_2, \phi_3) \end{bmatrix} \tag{5.19}$$

If a first surface has weights $\widetilde{w}_1$, $\widetilde{w}_2$ and $\widetilde{w}_3$, while another second surface has weights $\widetilde{w}_{1_{new}}$, $\widetilde{w}_{2_{new}}$ and $\widetilde{w}_{3_{new}}$, then the correlation between two surfaces becomes:

$$\begin{aligned} C = \widetilde{w}_1\widetilde{w}_{1_{new}}K_{11} + \widetilde{w}_2\widetilde{w}_{2_{new}}K_{22} + \widetilde{w}_1\widetilde{w}_{1_{new}}K_{33} \\ + \left(\widetilde{w}_1\widetilde{w}_{2_{new}}\right)K_{12} + \left(\widetilde{w}_1\widetilde{w}_{3_{new}}\right)K_{13} + \left(\widetilde{w}_2\widetilde{w}_{3_{new}}\right)K_{23} \end{aligned} \tag{5.20}$$

Figure 5.6 shows a block diagram for the illumination estimation phase. This phase aims to calculate the correlated colour temperature as an illumination parameter of the illuminant. McCamy's method is used in this approach, which is

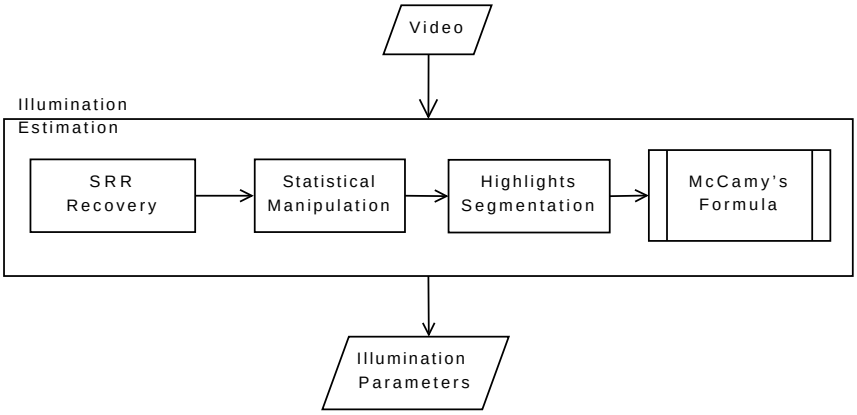


Figure 5.6: Illumination estimation block diagram

able to estimate the CCT from CIE 1931 chromaticity coordinates x and y.

The illumination estimation phase starts by segmenting areas in the image which represent high specularities (highlights); where a value for the illuminant CCT is initially set depending on the available knowledge of the environment. For indoor applications, as the one considered in this work, a value around 4000^oK is reasonable to cover the range of illumination used, this value is used to initialise the algorithm. McCamy's formula is then applied and the CCT is calculated. The illumination SPD is then calculated using Plank's formula (Equation4.2).

McCamy's method has a maximum absolute error of less than 2^oK for colour temperatures ranging from $2,856$ to $6,500^oK$ (corresponding to CIE illuminants A through D65) [168]. These errors are negligible over the range of interest in this technique. McCamy's method proves useful for implementation in real-time applications. It is derived from the assumption that CCT may be represented by a third-order polynomial function of the reciprocal of the slope of the line from specular pixels to the chromaticity of the light.

Highlights segmentation

In order to segment the areas with high specularities ($Highlight_{Mask}$), one of the acquired background frames BG is used and converted using the above SSR recovery module, then the autocorrelation for all pixels is performed using the statistical

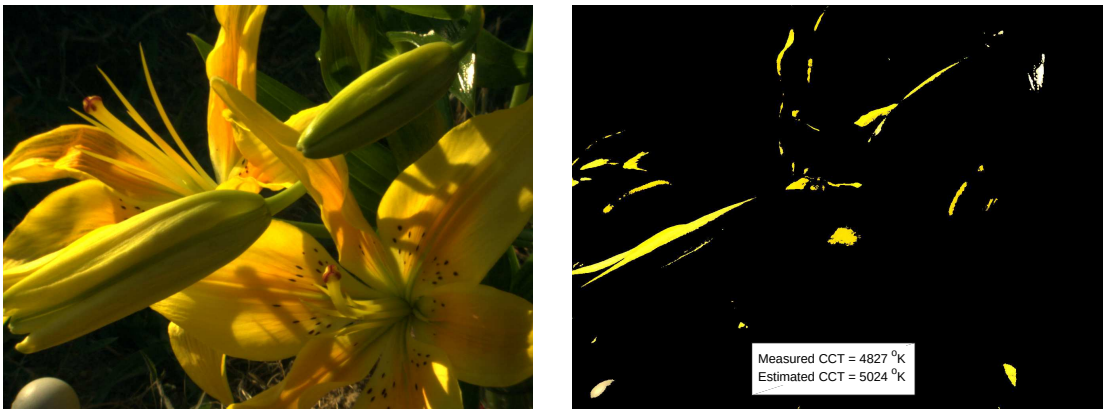


Figure 5.7: Segmented highlights from Foster data set 1

manipulation module. The highlights are then segmented as follows:

$$HL_{(x,y)} = \begin{cases} 0 & \text{if } Corr(BG, BG) < \tau_{HL} \\ 1 & \text{otherwise} \end{cases} \tag{5.21}$$

where τ_{HL} is chosen empirically, the value of 0.5 was found to be suitable.

Illumination estimation

The segmented pixels with high specularities are then converted to CIExyz and used to estimate the actual illuminant's correlated colour temperature, using McCamy's method according to the following two equations:

$$n = \frac{x - x_e}{y_e - y} \tag{5.22}$$

where $x_e = 0.3320$ and $y_e = 0.1858$ Then

$$CCT = 449n^2 + 3525n^3 + 6823.3n + 5520.33 \tag{5.23}$$

Samples of the segmented highlights are shown in Figure 5.7. Foster et al [167] recorded this scene in the Gualtar campus of University of Minho, Portugal, on 31 July 2002 at 17:40 under direct sunlight and blue sky. Their Telespectroradiometer reading produced CCT value of 4827^oK. The estimated CCT from the illumination estimation module is 5024^oK. More examples are given in Figures B.1 (Appendix B) .

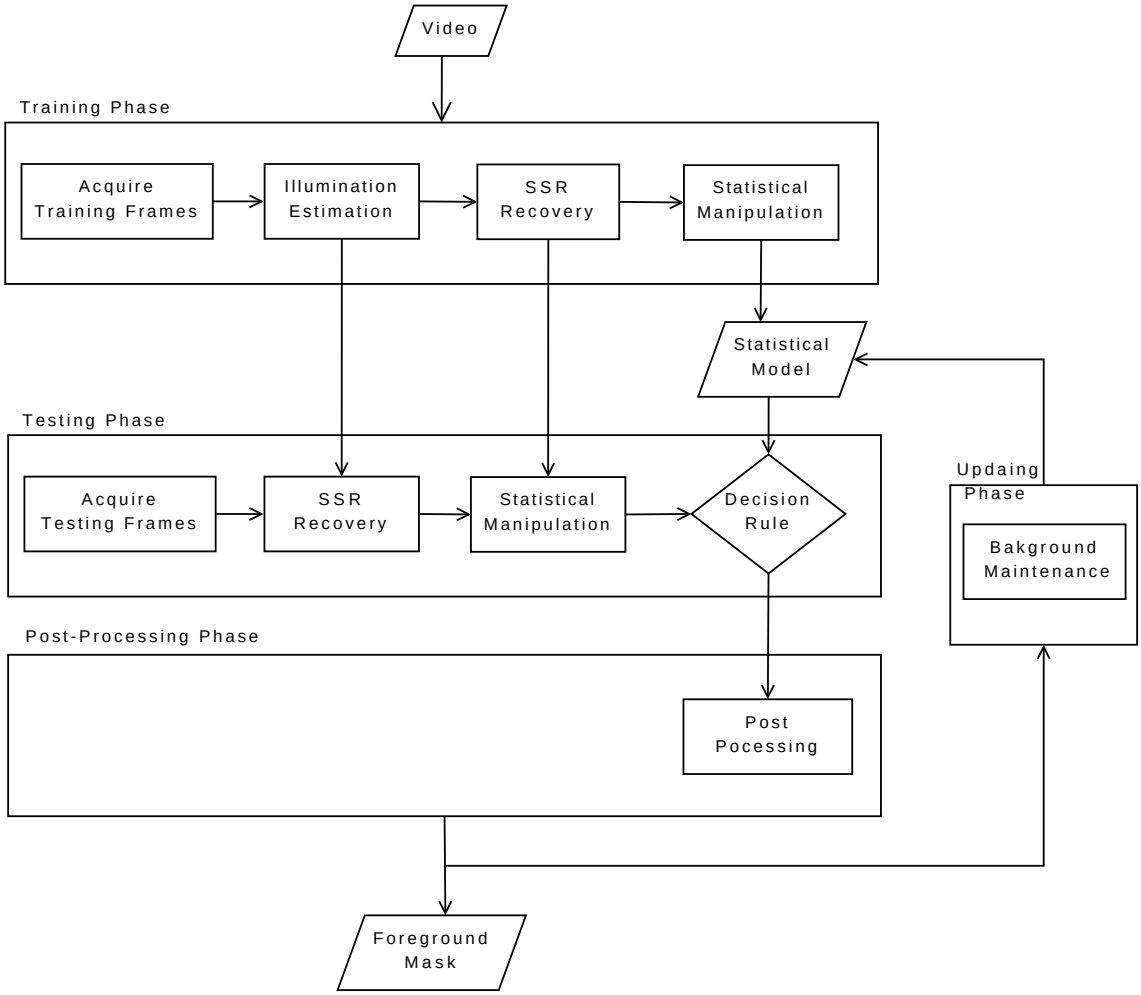


Figure 5.8: Algorithm 2 block diagram

5.4.2 The algorithm

Figure 5.8 shows the block diagram of Algorithm 2, the algorithm consists of four phases, the training phase, the testing phase, the post-processing phase and the background update phase.

Training phase

In order to build a model for the background, the training phase starts by acquiring a sequence of N input frames which represent relatively static background frames, followed by the illumination estimation module which estimates the CCT. Then the SSR recovery module converts the acquired background frames into SSR weights BGW using the estimated CCT. The mean of the background training frames BGW_{mean} is then calculated as a reference SSR ,

$$BGW_{mean} = mean(BGW) \tag{5.24}$$

The statistical manipulation module then passes through all background training frames BGW_i and calculates the correlation CR_{BG} between the SSR of each frame and the mean SSR of background frames using Equation 5.20,

$$CR_{BG} = Corr(BGW_i, BGW_{mean}) \tag{5.25}$$

where i represents the background frame number from 1 to N, where N is the number of background frames used to estimate the SSR of the background.

Finally, the statistical manipulation module calculates the maximum CR_{max} and minimum CR_{min} of CR_{BG} as a statistical model of the background,

$$CR_{min} = min(CR_{BG}) \tag{5.26}$$

$$CR_{max} = max(CR_{BG}) \tag{5.27}$$

Testing phase

To perform the foreground segmentation, the testing phase starts by capturing a new frame, and then the SSR recovery module converts the new frame NF to basis function weights NFW representing its SSR, using the illumination parameter

calculated in the illumination calculation stage of the training phase.

Figures 5.10(a) and 5.9(a) represent the reconstructed spectral reflectance for one pixel which has not been changed in the background frame and the new frame in 'Hall Monitor' and 'Intelligent Room' Sequences respectively. The figures show that the pixel maintain its spectral characteristics. Figure 5.10(b) and 5.9 (b) represent one pixel which has been changed in the background frame and the new frame in 'Hall Monitor' and 'Intelligent Room' Sequences respectively. The figures show that the pixel changes it spectral characteristics where new object covers the background.

The statistical manipulation module then calculates the correlation CR between the SSR of the new frame NFW and the mean SSR of background frames BG_{mean},

$$CR = Corr(NFW, BGW_{mean}) \tag{5.28}$$

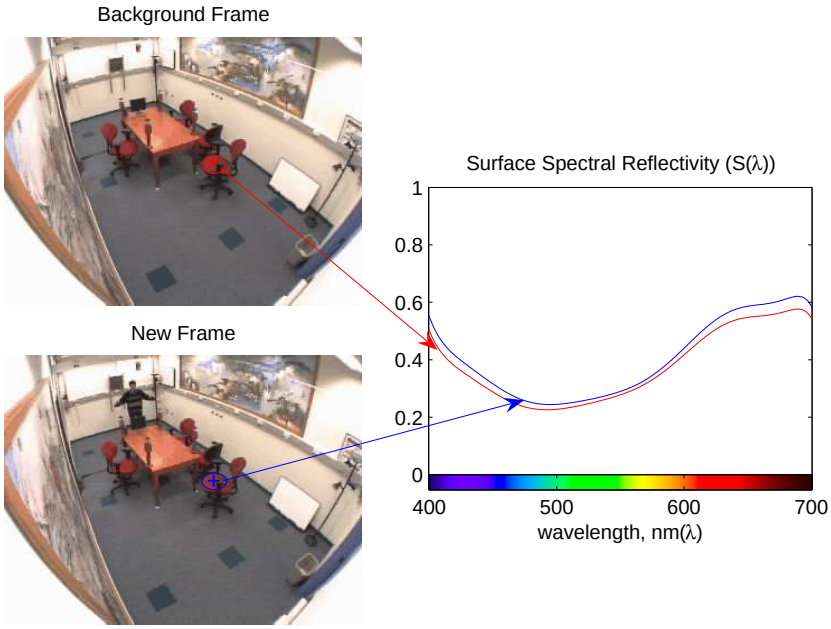
The threshold operation creates an initial binary foreground mask defined as FG_{Mask}, using the maximum CR_{max} and minimum CR_{min} values of correlation ratio CR_{BG} measured from the pool of background frames.

$$FG_{Mask} = \begin{cases} 0 & \text{if } (1-\tau_c)CR_{min} < CR < (1+\tau_c)CR_{max} \\ 1 & \text{otherwise} \end{cases} \tag{5.29}$$

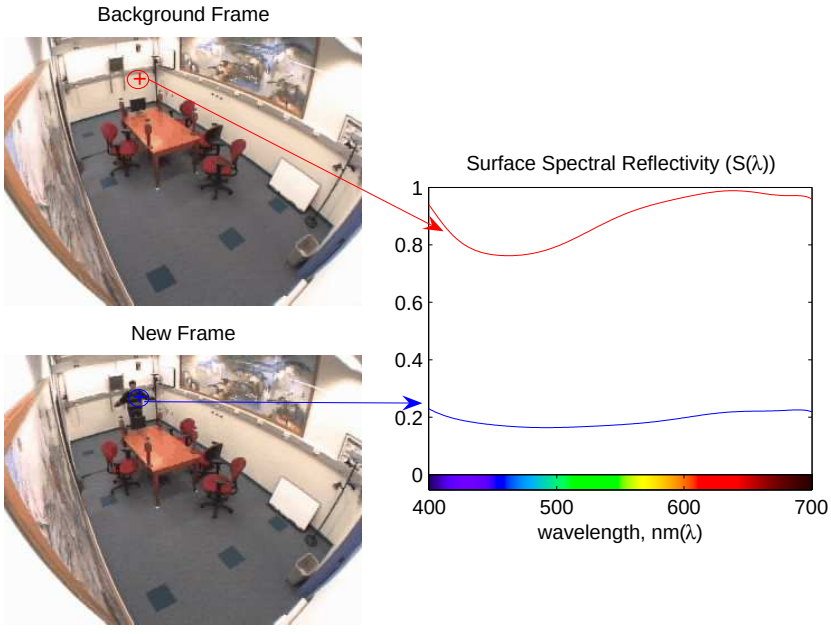
The threshold τ_c is calculated using a threshold selection criterion, discussed in Chapter 6.

Post-processing phase

After the initial foreground mask FG_{MAsk} has been created from the new frame by the testing phase, a post-processing phase which consists of a noise removal module applies a dilation step which performs a dilation operation thereon to expand or thicken the initial foreground mask FG_{MAsk}. The dilation operation is the union of all translations of the mask by the structuring element SE. Then an erosion step performs an erosion operation on the resultant mask, which reduces the size of an object by eliminating area around the object edges. After the erosion operation has been completed; the detection stage performs a connected component blob analysis on the foreground mask FG_{MAsk}. The erosion operation

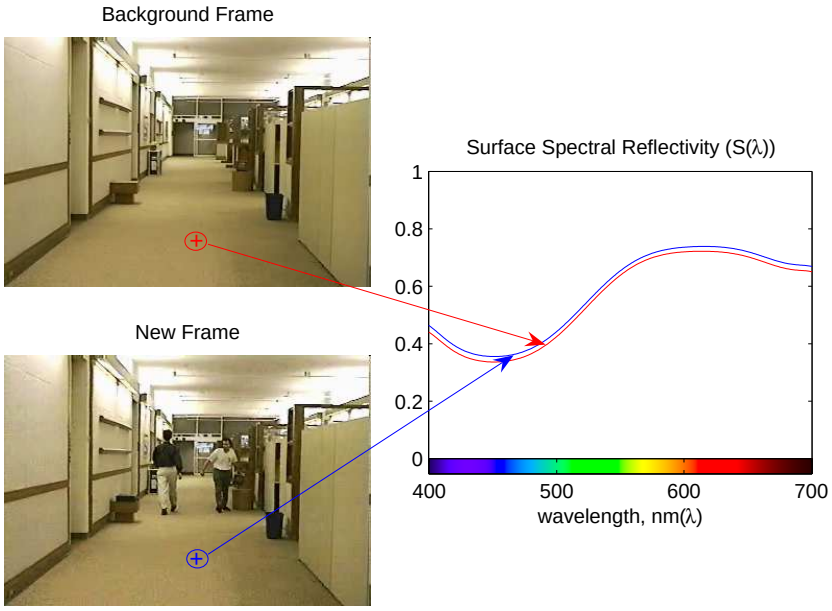


(a) The pixel has not been changed

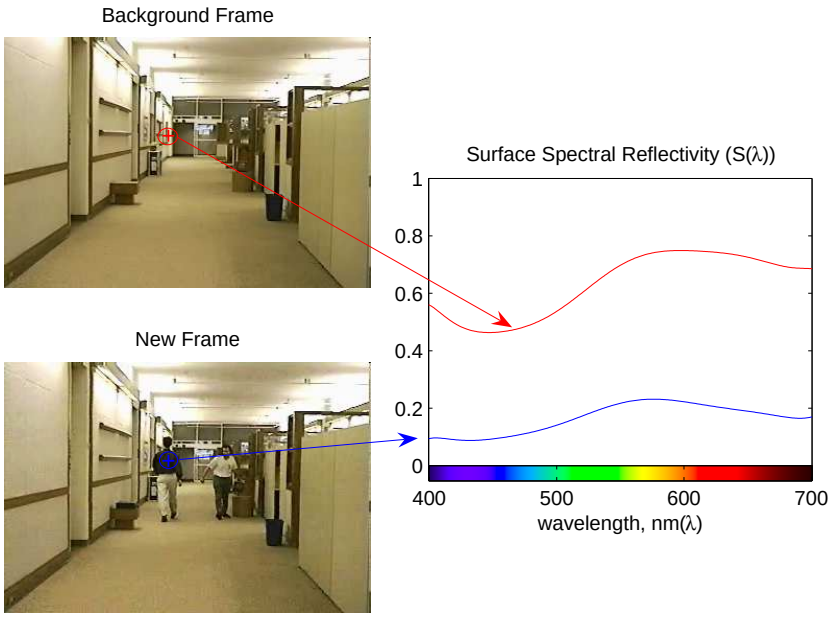


(b) The pixel has been changed

Figure 5.9: Reconstructed spectral reflectance for one pixel ('Intelligent Room' Sequence)



(a) The pixel has not been changed



(b) The pixel has been changed

Figure 5.10: Reconstructed spectral reflectance for one pixel ('Hall Monitor' Sequence)

removes noise and eliminates foreground mask details smaller than the structuring element, where the size of the structuring element for both dilation and erosion is: $SE = 2$ pixels.

Updating phase

The statistical background model is derived by observing the scene during a training period; this model may be updated continuously, especially where illumination changes occurs, or where objects may enter the scene and remain static for substantial time periods.

The background model is updated through a model maintenance phase, using an update rate α, according to the following formulas: if $FG = 0$ then update BGW_{mean} as:

$$BGW_{mean} = (1 - \alpha) \cdot BGW_{mean} + \alpha \cdot NF$$

The algorithm stores the value of the CCT recorded during the background modelling phase CCT_{BG}. For each new frame the algorithm calculates the new CCT value CCT_n. The algorithm then check if $abs(CCT_n - CCT_{BG}) > \tau_{CCT}$, where by experiments $\tau_{CCT} = 100^oK$ is found to be a reasonable change in the illumination. If the illumination changes the background model is then updated as follows:

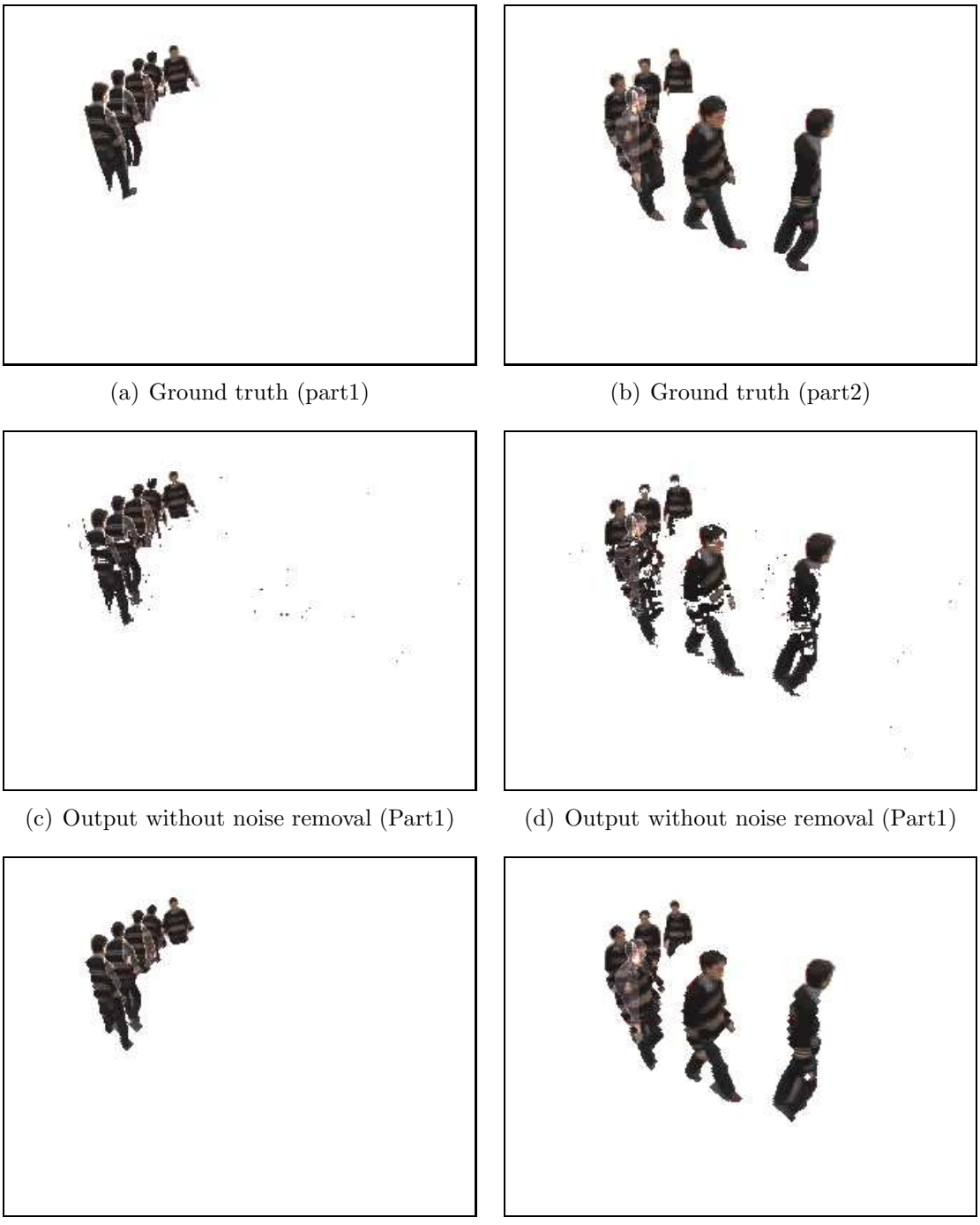
$$CR_{min} = min(CR_{min}, CR) \tag{5.30}$$

$$CR_{max} = max(CR_{max}, CR) \tag{5.31}$$

The value of the update rate used for the background model is $\alpha = 0.001$. Sample Results for the algorithm are shown in Figure 5.11. More discussion about the results as well as the comparison between this technique with one of the state-of-the-art techniques will be carried out as part of the evaluation chapter (Chapter 6).

Algorithm 2 Change detection algorithm based on surface spectral reflectance

Input: N-background frames
Output: Foreground mask
{Modelling Phase}
Step 1: Spectral Reflectance Conversion. Calculate the spectral reflectance weights using the transformation in 5.18.
Step 2: Background Model. Calculate the mean of the N background spectral reflectance.
Step 3: Correlation. Calculate the correlation between each background frame and the background model.
Step 4: Statistical. Calculate the maximum and the minimum of the correlation calculated in Step 3 (per pixel).
{Testing Phase}
Step 5: Spectral Reflectance Conversion. Calculate the spectral reflectance weights for the new video frame using the transformation in Equation 5.18.
Step 6: Correlation. Calculate the correlation between each new video frame and the background model.
Step 7: Segmentation. Compare the correlation calculated in Step 6 with the maximum and minimum calculated in Step 4 using Equation 5.29.
{Updating Phase}
Step 8: Updating the BGW_{mean}. Update the background SSR model BGW_{mean}.
Step 9: CCT Estimation. Calculate the CCT from each new frame.
Step 10: Updating Statistical Model. If illumination changes update the statistical model using Equation 5.30.



(a) Ground truth (part1)

(b) Ground truth (part2)

(c) Output without noise removal (Part1)

(d) Output without noise removal (Part1)

(e) Output with noise removal (Part2)

(f) Output with noise removal (Part2)

Figure 5.11: Samples of Algorithm 2 results for 'Intelligent Room' sequence

5.5 Summary

An investigation on the challenges of using physics-based models for image formation as a basis for change detection algorithms has been carried out in this chapter. Two novel algorithms have been developed which are based on a physical meaning of image formation and which use advances in colour constancy techniques.

The first algorithm models illumination variations using the ratio between foreground pixels and background pixels, based on shading model. The assumption which has been made is that only diffuse components are considered and narrow-band camera sensitivities are applicable. The Mahalanobis distance of these ratio triples are then used to segment foreground and shadows from static background.

The second algorithm is based on physics-based image representations which transforms the RGB colour space into weights which represent the surface spectral reflectance of the material. The assumptions are that only diffuse reflections are considered and the existence of a dominant illuminant is applicable.

The challenge of the change detection approach based on surface spectral reflectance arises from the new idea of processing the full-spectrum of the surface spectral reflectance instead of the three samples used by other colour spaces. The spectral representation uses a linear model, which consists of a number of basis functions (pre-trained from a set of materials) and weights (calculated for the object under investigation). A numerical estimation of the physics-based model for image formation and the real-time transformation from the camera output to the physical parameters is carried out. The segmentation between foreground and background objects is done through the correlation between the intrinsic spectral reflectance over the visible wavelengths of a reference background frame and each new frame.

"Segmentation results depend on the task at hand and on the visual content. No single segmentation technique is universally useful for all applications. Moreover different techniques are not equally suited for a particular application. An effective evaluation of segmentation is important in selecting the most appropriate technique for a specific application, and furthermore for optimally setting its parameters."

A. Cavallaro [158]

6

Performance Evaluation

6.1 Introduction

Change detection algorithms are a key component of a surveillance system. However, the required performance of such techniques differs depending on the target application [170], and on the particular characteristics of the monitored environment.

Due to the diversity of smart video surveillance applications and the increasing development of change detection algorithms, an automatic assessment procedure is desired to compare the results provided by different algorithms. This procedure is referred to as *objective evaluation*, which compares the output of each algorithm with the ground truth, obtained manually, and measures the differences according to objective metrics. The key challenge here is the presence of a decision process

parameter(s) (threshold(s)) that influence the outcome of the algorithms. So, such evaluation procedures should assist in both selecting optimal parameters for each algorithm and ranking different algorithms, depending on specific application requirements.

This chapter aims to evaluate the performance of the proposed change detection algorithms proposed in Chapter 5 and one of the state-of-the-art algorithms which derives the same functionality, targeting the requirements of the workplace surveillance application proposed in Chapter 2. Discrepancy empirical evaluation methods for change detection approaches, discussed in Chapter 3, are adopted in this chapter. An objective criteria which indicates the difference between the segmented frames and reference frames will be used. Real image data will be manually annotated, to specify an ideal frame segmentation (ground truth). In the case of physics-based algorithms, the synthetic data will not faithfully represent the physical parameters of the real monitored scene.

Three evaluation methods are adopted, illustrative visual comparison, objective, and computational complexity evaluation. Section 6.2 discusses the performance metrics which will be used in the threshold selection criterion and objective evaluation throughout this chapter. The experiments setup, and the data set used in this chapter are introduced in Section 6.3. Optimal algorithms decision rules (thresholds) that provide the best performance for each algorithm are chosen in Section 6.4. In section 6.7 the performance metrics are applied to objectively evaluate the proposed algorithms. The algorithms are evaluated in terms of their computational complexity in Section 6.8. Finally, a new performance evaluation metric is proposed and discussed.

6.2 Performance metrics

In this thesis, objective metrics are adopted to evaluate the performance of the proposed change detection algorithms by comparing the output of the algorithms with the ground truth obtained by manual edition. False alarms and detection failures are considered as types of errors. A false alarm occurs when a foreground

Table 6.1: Contingency table

		True Class	
		Foreground	Background
Output Class	Foreground	**TP**	**FP**
	Background	**FN**	**TN**

object is detected at a position where none is present. The detection failures are caused by missing regions which should have been detected. Ground truth based measures are used to evaluate the segmentation quality of different change detection algorithms, using pixel-based measures.

Given the contingency table where True Positives (TP), number of correctly detected foreground pixels, False Positives(FP), number of pixels falsely detected as foreground, and False Negatives (FN),, number of pixels falsely detected as background, as shown in Table 6.1. The Precision P and Recall R may be estimated as follows:

$$Recall = R(k) = \frac{TP}{TP + FP}$$

where k is the frame number. The Recall can be used to estimate the tendency of an algorithm to under-segment. The higher the recall the less likely is the under-segmentation.

$$Precision = P(k) = \frac{TP}{TP + FN}$$

The Precision can be used to analyse the tendency of an algorithm to over-segment. The higher the recall the less likely is over-segmentation.

$$F(k) = \frac{2}{\frac{1}{P(k)} + \frac{1}{R(k)}}$$

The Harmonic Mean (HM), which is called the F-measure, will be used as a single measure which combines Precision and Recall with equal weights, it assumes a high value only when both Precision and recall are high.

$$CDR = 1 - MDR = Precision$$

$$FAR = 1 - Recall$$

It is possible to estimate a pair ($CDR(k)$, $FAR(k)$) for each frame in a sequence under certain threshold value. Averaging of $CDR(k)$ and $FAR(k)$ over a whole sequence of K frames ($k = 1 \rightarrow K$), processed by the system, gives a more robust estimate of the Receiver operation characteristics (ROC) curve point($CDR(k)$,$FAR(k)$).

ROC curves

ROC curves were developed to characterise the ability of a radar operator to correctly distinguish between target echoes and false echoes.

ROC curves have been extensively used: in medical sciences, to evaluate and compare diagnostic systems; for the performance evaluation of image processing algorithms e.g. edge detection algorithms; and for the evaluation of machine learning algorithms. In this thesis ROC curves provide a well assessed tool that can be used for the purpose of evaluating the performance of change detection algorithms.

An ROC curve is generated by plotting the correct detection rate (CDR) on the vertical axis and the false alarm rate (FAR) on the horizontal axis. Each point on the ROC curve represents a specific CDR and FAR obtained at a specific value of the decision rule threshold.

In a complex algorithm, one main problem is that $CDR(k)$ and $FAR(k)$ depend on many thresholds. Varying these thresholds, different ROC curves describing the performance of the algorithm under different configurations can be obtained.

Consequently, the probability of correct detection and false alarms, used in the ROC curves can be defined for a quite large class of application requirements. On this basis, performances of different systems modules can be assessed in terms of their capability to produce intermediate results potentially capable of satisfying proposed workplace surveillance requirements.

The crucial point is the estimation of detection and false alarm rates for the considered task. For obtaining these quantities, the output of the algorithm under evaluation has to be examined. A ground-truth reference is required in order to compare the algorithm output with the segmentation produced by a surveillance operator monitoring the scene.

6.2.1 Parameter selection criterion

In change detection algorithms, the decision rule threshold determines if a pixel in the current frame corresponds to a moving object or not. Due to camera noise and illumination variations, a certain number of pixels are different from the background eventhough they do not belong to moving objects: a high value of the threshold allows maintaining a low number of false alarms. On the other hand, foreground objects presenting a low contrast with respect to the background risk to be eliminated if the threshold is excessively high.

The parameter selection criterion chooses thresholds which maximise the performance of the algorithm for a specific application. The parameters can be selected by two different methods, either by selecting the parameter which corresponds to a maximum F-measure or using ROC curves (minimax criterion).

Minimax criterion

ROC curves are used as a basic tool; a method is used that consists of tracing ROC curves for decision rule threshold(s). Such curves will be analysed in order to allow the maximisation of the global performance from the change detection algorithms under test according to a statistical criterion related to minimax one as the one used in [139]. The minimax criterion is a probabilistic criterion designed to minimise the maximum possible Bayesian risk when no knowledge about source probabilities is available [140]. Using this criterion gives:

$$CDR = 1 - (\frac{C_f}{C_m})FAR$$

where C_m and C_f are costs associated with mis-detection and false alarm errors, respectively and they can be fixed by using knowledge related to the specific video surveillance application [171].

For the application of the indoor workplace surveillance for example, C_m can be related to the cost paid when an event is not detected, while C_f can represent the detection of unimportant event, and the generation of a false alarm. A balance between the higher corrrect detection rate and lower false alarm rate is important. In the present evaluation C_f and C_m are considered equal without losing generality.

The minimax equation and the related line can be represented on the ROC curve (CDR,FAR) plane. In particular, it can be shown that the interaction between the minimax equation and an ROC curve gives the best operating point, i.e. the configuration of system parameters that minimise the maximum Bayesian risk under no knowledge of a-priori source probabilities.

6.3 Experiments setup

6.3.1 Data sets

The dataset used in this evaluation consists of three full-frame rate video sequences: two public domain video sequences ('Intelligent Room' [172] and 'Hall Monitor' [173]) and the video sequence 'Lab' captured by the author. Figure 6.1 shows samples of the used data set.

A video manipulation interface has been implemented on Matlab in order to allow a human operator to segment moving objects from the static background. A total number of 899 frames are segmented manually by marking-up the pixel boundary of foreground objects, for use as ground truth in experiments[1]. The ground truth data was validated for manual segmentation errors. The 'Intelligent Room' [172] test sequence consists of 300 frames and the frame rate is 25 fps. It is a (320×240) uncompressed video sequence. In the 'Intelligent Room' sequence the clothes of the person in the scene are textured which result in higher discrimination rates. The first 83 frames of the 'Intelligent Room' sequence contain no moving objects, and this

[1]Thanks to Mr Noaman and Mr Dawood who have performed the manual segmentation.

(a) 'Intelligent Room' (b) 'Hall Monitor' (c) 'Lab'

Figure 6.1: Samples of the used data set

means that these frames can be used to build the background model. The following frames of the sequence represent one participant who enters from a door on the left and moves toward the far end of the room away from the camera. The participant then start raising his hands up and lowering them down to the normal position, and then he proceeds back toward the camera, passing the door and walking toward the front end of the room. Appendix C show sample visual frames which represent this scenario (Figure C.1).

The famous 'Hall Monitor' [173] test sequence is a CIF (352×288) MPEG-4 which contains 299 frames and the frame rate is 25 fps. It is a challenging sequence for change detection, due to the colour content of objects and background, which are quite similar. The sequence also contains a cluttered background as well as a high level of noise. These features make the sequence a significant test case. The first 12 frames of the 'Hall Monitor' sequence contain no moving objects, and this means that these frames can be used to build the background model. The following frames of the sequence represent two participants, participant A and B. Participant A enters the scene from the left side door, while carrying a briefcase and walks through the corridor, away from the camera. After a few meters he places the briefcase on a bench and walks toward another door on the left. Before participant A exits, participant B enters the scene through a door on the right, and walks towards the camera. On his way he picks up a television set from another bench on the right. Appendix C shows sample frames which represent this scenario (Figure C.2).

The 'Lab' test sequence is a (320×240) uncompressed video sequence. It contains 300 frames and the frame rate is 25 fps. It is a challenging sequence for change detection due to several factors. Firstly, due to the camera field of view, the camera

is always situated near to the moving objects. Secondly, the participant is moving very near to the camera. Thirdly, the sequence contains a cluttered background and a high level of noise. All of these factors are known to impact upon the accuracy of change detection algorithms and make the sequence a significant test case. Appendix C shows selected frames which represent this scenerio (Figure C.3). They show a cluttered computer lab with tables, chairs and computer equipment seen through a camera situated on the ceiling at the near left of the room. The first 76 frames of the sequence contain no moving objects and this means that these frames can be used to build the background model. Following these static frames, the sequence shows one participant who enters the scene from the far right corner of the room. This participant then walks between the desks to the near left corner of the room and exits the scene.

6.3.2 Comparison of state-of-the-art algorithms in a previous independent study

Karaman et al. [174] carry out a comparison between nine of the state-of-the-art change detection methods each other in terms of the segmentation quality and computational complexity. Those methods are image differencing, Cavallaro et al. [48], Shen et al. [133], Haritaoglu et al. [11], Francois et. Al [58], McKenna et al. [71], Hong et al. [75], Jabri et al. [74] and Horprasert et al. [73]. Figure 6.2 shows the results of the objective evaluation as in [174]. These results are the average of the results obtained from five video sequences. As can be seen from the figure both Shen, Horprasert and Cavallaro outperforms other algorithms, see Figure 6.2 (a). However Horprasert and Cavallaro are three times as fast as Shen, see Figure 6.2 (b). Although both Horprasert and Cavallaro are very close in terms of their segmentation quality and computational complexity, the only difference lies in the balance in Horprasert method between Precision and Recall. Horprasert et al. utilize a pixel-based background model including luminance and chrominance information.

The approach of Horprasert et al. [73] was adopted, for the comparison, since it gives the best trade-off between segmentation quality and computational complexity,

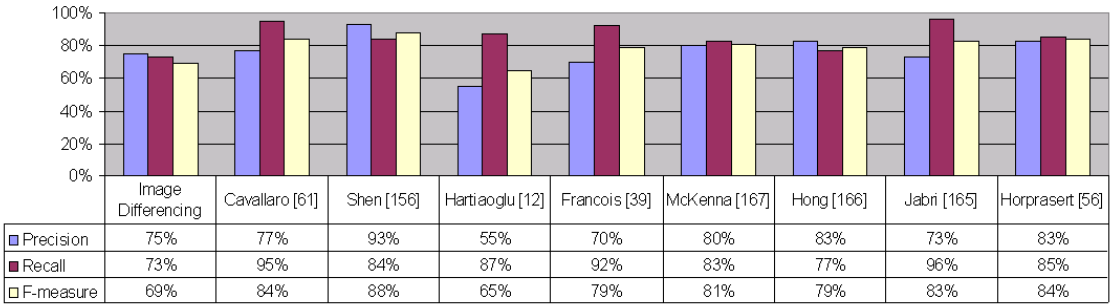


	Image Differencing	Cavallaro [61]	Shen [156]	Hartiaoglu [12]	Francois [39]	McKenna [167]	Hong [166]	Jabri [165]	Horprasert [56]
Precision	75%	77%	93%	55%	70%	80%	83%	73%	83%
Recall	73%	95%	84%	87%	92%	83%	77%	96%	85%
F-measure	69%	84%	88%	65%	79%	81%	79%	83%	84%

(a) Precision, Recall and F-measure

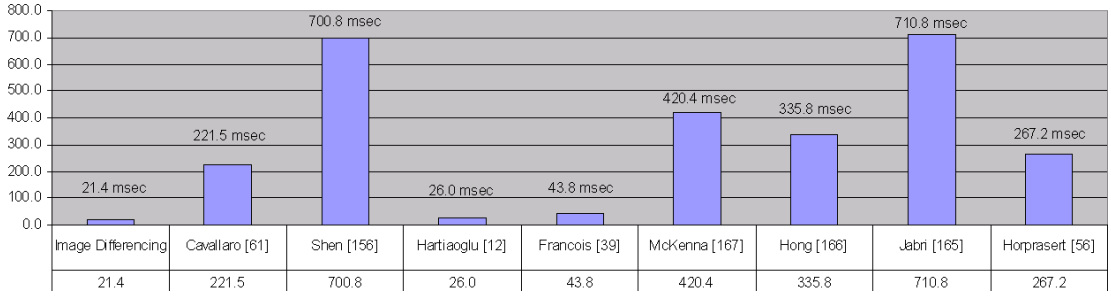


Image Differencing	Cavallaro [61]	Shen [156]	Hartiaoglu [12]	Francois [39]	McKenna [167]	Hong [166]	Jabri [165]	Horprasert [56]
21.4	221.5	700.8	26.0	43.8	420.4	335.8	710.8	267.2

(b) Average time to process one frame

Figure 6.2: Comparison of state-of-art algorithms [174]

allowing real-time performance. The method starts by building a background model from a number of static background frames. This method models the background using a model that separates luminance and chrominance information. Based on this model the luminance and chrominance distortion of an unknown pixel are calculated and used for classifying each pixel as foreground, background, highlight or shadow. This algorithm represents a multi-class approach that allows coping with several change detection problems.

6.4 Parameter selection

In this section, the parameter selection criteria defined in section 6.2.1 are used to select optimal parameters for the background subtraction and shadow detection algorithm developed by Horprasert [73], the change detection method based on the reflectance ratio (Algorithm 1), and the change detection algorithm based on the spectral reflectance (Algorithm 2). The F-measure as well as the ROC curves using the minimax criterion are applied for selecting the optimal threshold values for the

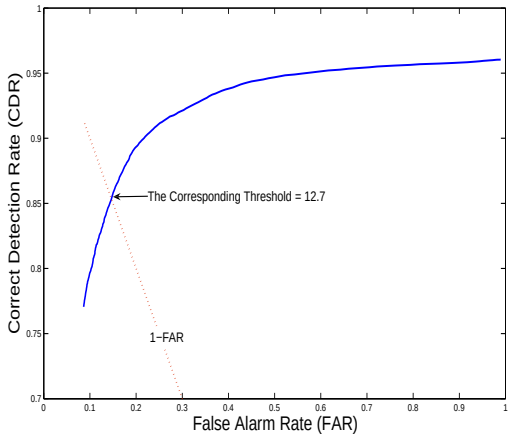
data sets used.

6.4.1 Horprasert algorithm

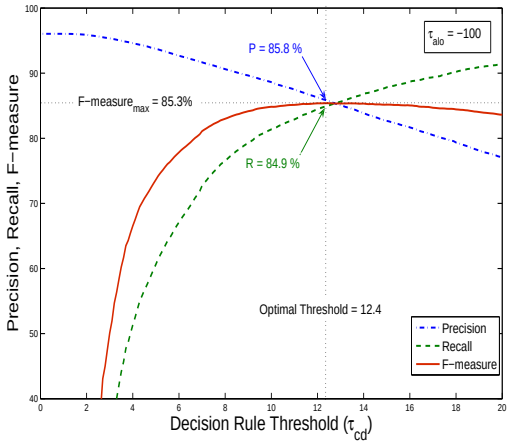
The performance of the Horprasert algorithm at different threshold values has been evaluated by computing the related ROC and PRF (Precision, Recall, F-measure) curves, according to the definitions introduced in Section 6.2 and traced for each threshold value. In particular, 200 different values of change detection threshold τ_{cd} have been selected which regulates the behaviour of the algorithm from 0.1 to 20 with step of 0.1. Figure 6.3 shows the ROC and PRF curves for the background subtraction and shadow detection method developed by Horprasert [73]. Taking $\tau_{alo} = -100$ and considering different video sequences in order to estimate the optimal τ_{cd}. For the 'Intelligent Room' sequence, the optimal value for τ_{cd} is 12.7 using ROC curves and 12.4 using the maximum F-measure. For the 'Hall Monitor' sequence, the optimal value for τ_{cd} is 8.7 using ROC curves and 8.8 using the maximum F-measure. For the 'Lab' sequence, the optimal value for τ_{cd} is 6.3 using ROC curves and 6.5 using the maximum F-measure.

6.4.2 Algorithm 1

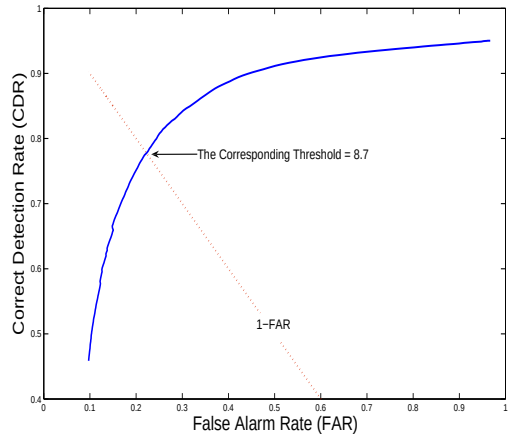
The performance of Algorithm 1 at different threshold values has been evaluated by computing the related ROC and PRF curves and traced for each threshold value. In particular, 276 different values of change detection threshold τ_m have been selected which regulates the behaviour of the algorithm from 25 to 300 with step of 1. Figure 6.4 (a) and (b) show the ROC and PRF curves for Algorithm 1 respectively. Considering different video sequences in order to estimate the optimal τ_m. For 'Intelligent Room' sequence, the optimal value for τ_m is 125 using ROC curves and 175 using the maximum F-measure. For the 'Hall Monitor' sequence, the optimal value for τ_m is 145 using ROC curves and 165 using the maximum F-measure. For the 'Lab' sequence, the optimal value for τ_m is 95 using ROC curves and 125 using the maximum F-measure. The other used thresholds τ_{sh}, τ_1 and τ_2 are fixed with $\tau_{sh} = 180$, $\tau_1 = -0.5$ and $\tau_2 = 0.3$ for all data sets.



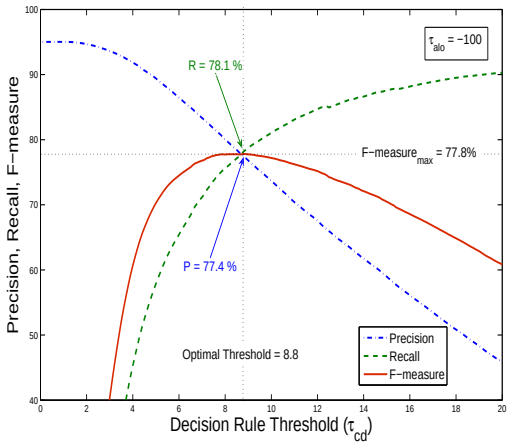
(a) ROC curves('Intelligent Room')



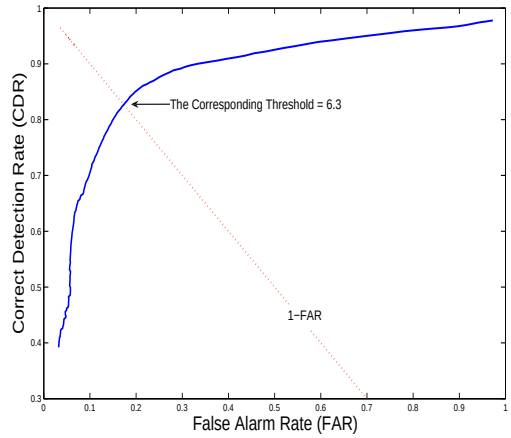
(b) PRF curves('Intelligent Room')



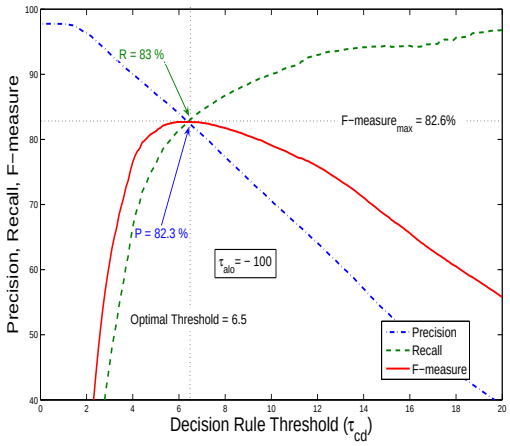
(c) ROC curves('Hall Monitor')



(d) PRF curves('Hall Monitor')

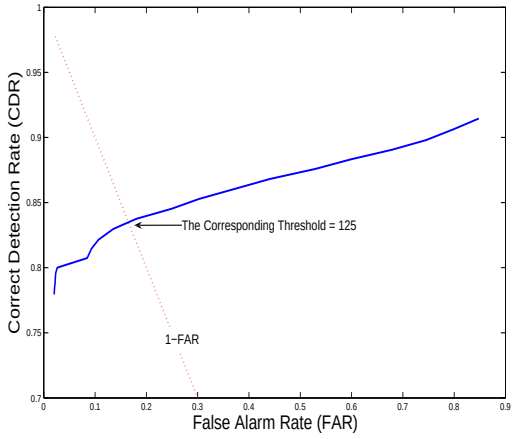


(e) ROC curves('Lab')

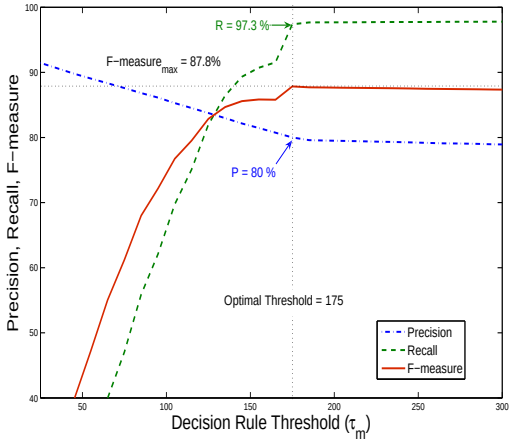


(f) PRF curves('Lab')

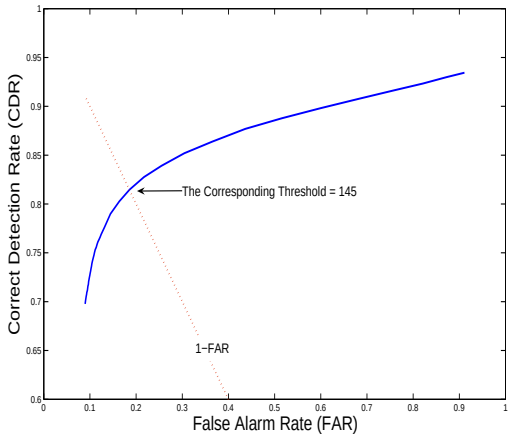
Figure 6.3: ROC curves and PRF curves for threshold selection of Horprasert method



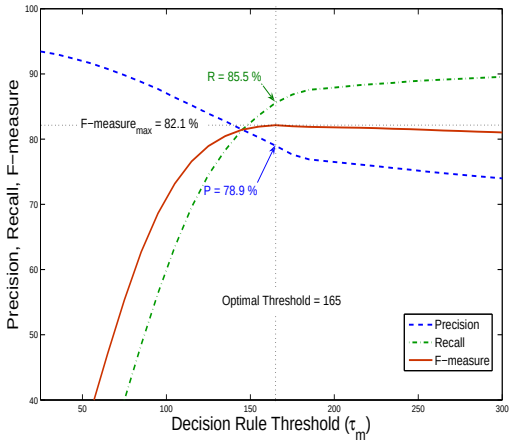
(a) ROC curves('Intelligent Room')



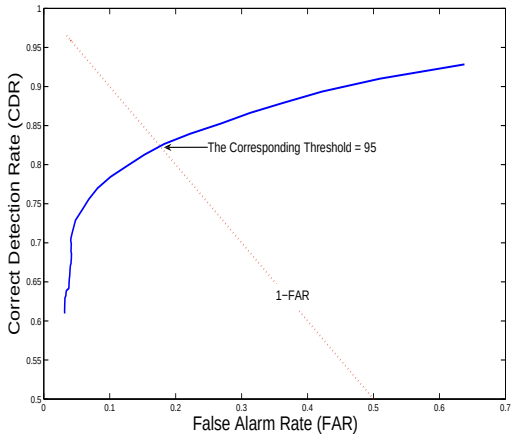
(b) PRF curves('Intelligent Room')



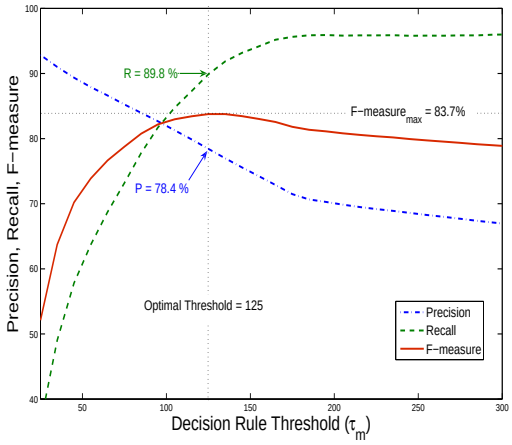
(c) ROC curves(Hall Mnitor)



(d) PRF curves(Hall Mnitor)

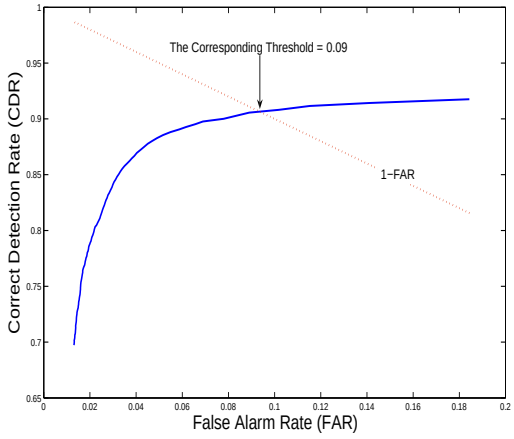


(e) ROC curves('Lab')

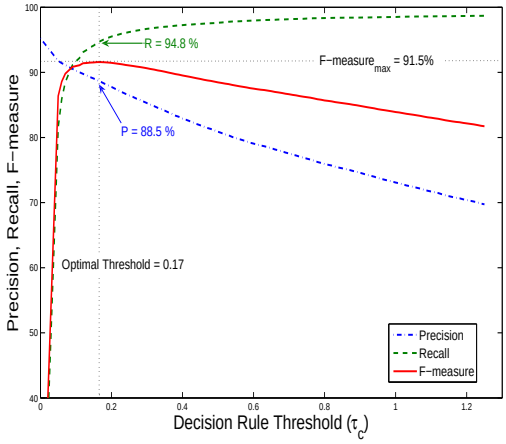


(f) ROC curves('Lab')

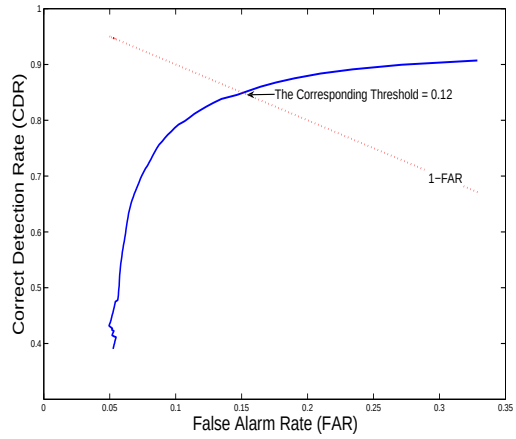
Figure 6.4: ROC curves and PRF curves for threshold selection of Algorithm 1



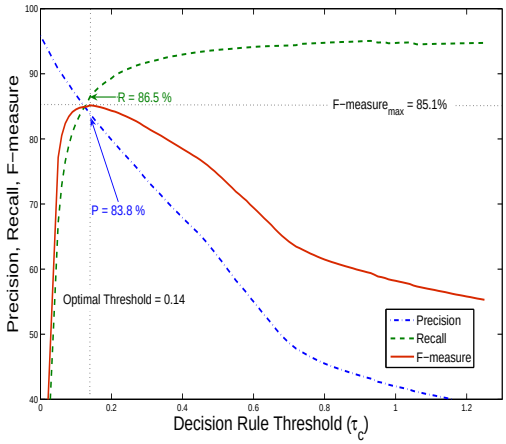
(a) ROC curves('Intelligent Room')



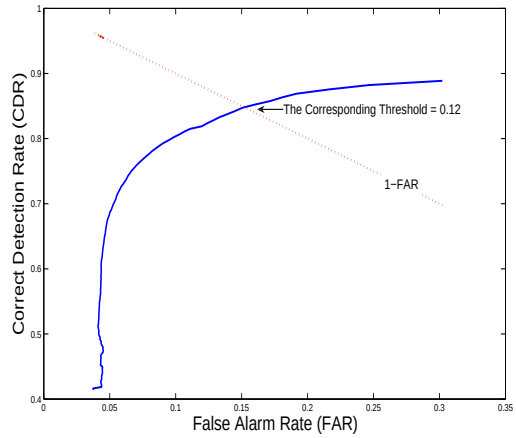
(b) PRF curves('Intelligent Room')



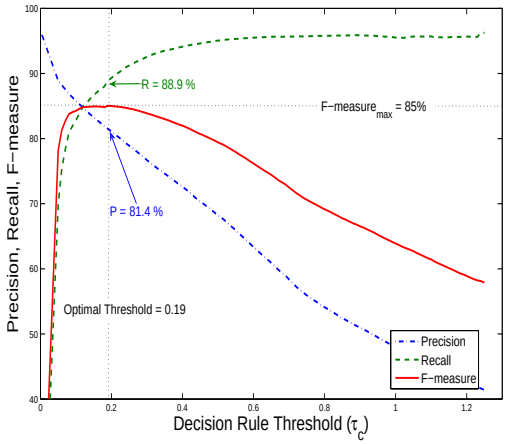
(c) ROC curves('Hall Monitor')



(d) RPRF curves('Hall Monitor')



(e) ROC curves('Lab')



(f) PRF curves('Lab')

Figure 6.5: ROC curves and PRF curves for threshold selection of Algorithm 2

6.4.3 Algorithm 2

The performance of Algorithm 2 at different threshold values has been evaluated by computing the related ROC and PRF curves. In particular, 126 different values of change detection threshold τ_c have been selected which regulates the behaviour of the algorithm from 0.05 to 1.25 with step of 0.01. Figures 6.5 (a) and (b) show the ROC and PRF curves for Algorithm 2, respectively. For the 'Intelligent Room' sequence, the optimal value for τ_c is 0.09 and 0.17 using ROC curves and the maximum F-measure respectively. For the 'Hall Monitor' sequence, the optimal value for τ_c is 0.12 using ROC curves and equals 0.14 using the maximum F-measure. For the 'Lab' sequence the optimal value for τ_c is 0.12 and 0.19 using ROC curves and the maximum F-measure respectively.

6.5 Tolerance to threshold variation

By plotting the F-measure of different data sets for the same algorithm, it is possible to evaluate the tolerance of the algorithm performance to variation of the threshold.

Figure 6.6 shows the overall performance metric (F-measure) of Horpraset algorithm for the 'Intelligent Room' video sequence plotted against the decision rule threshold. The figure includes three vertical markers which represent different optimal thresholds for 'Lab', 'Hall Monitor' and 'Intelligent Room' video sequences. Using the optimal threshold of the 'Intelligent Room' video sequence gives peak F-measure of 85.39%. However, using the optimal thresholds of either the 'Hall Monitor' or the 'Lab' sequences for the 'Intelligent Room' sequence results in the reduction of the F-measure. Using the optimal threshold of the 'Hall Monitor' sequence gives peak F-measure of 83.97%, while using the optimal threshold of the 'Lab' sequence gives peak F-measure of 79.38%. Figure 6.7(a) is similar to Figure 6.6 as it shows the same results of the Horprasert algorithm for the 'Intelligent Room' sequence, however this figure also includes the results of the 'Hall Monitor' and 'Lab' sequences. Figures 6.7(b) and (c) show similar plots for Algorithm 1 and Algorithm 2 respectively.

Table 6.2 summarises the results obtained from Figure 6.7. It is divided into

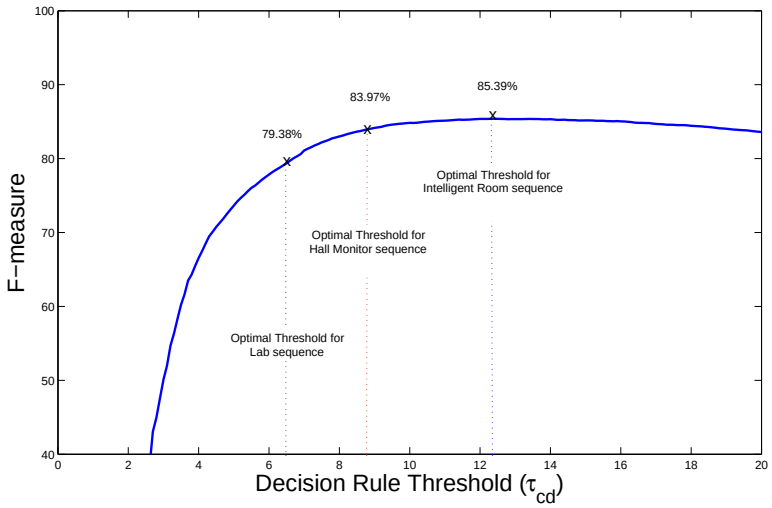
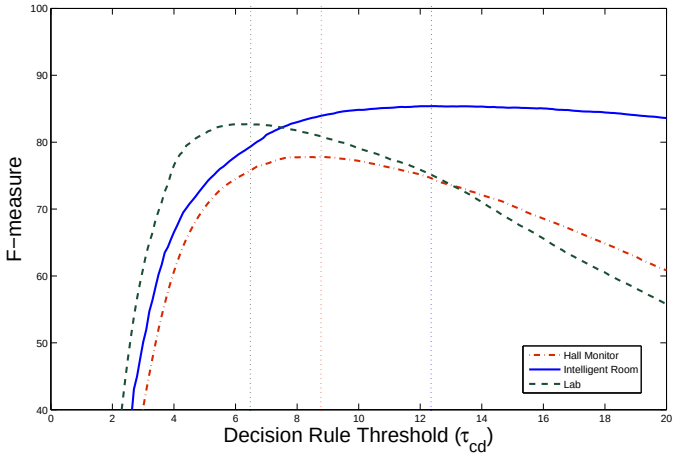


Figure 6.6: F-measure vs. Threshold example

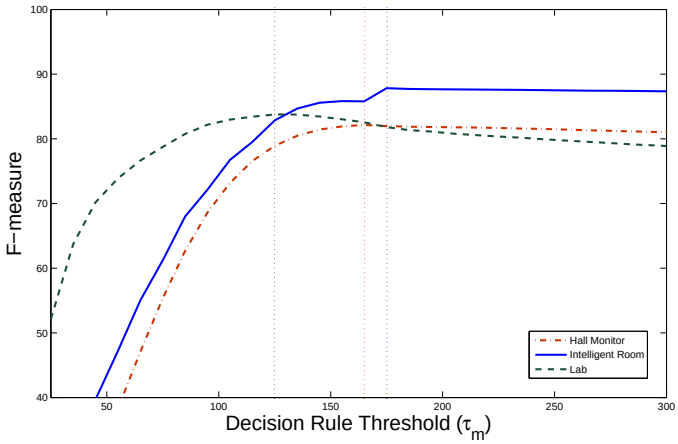
three parts; each part represents one of the three algorithms. The first row of table 6.2(a) summarises the process described above for the 'Intelligent Room' sequence using the Horprasert algorithm. The minimum and maximum deviations from the maximum F-measure are also shown. Subsequent rows show the results for the 'Hall Monitor' and 'Lab' sequences. The last row shows the mean of the maximum and minimum deviations. The same process is repeated in part (b) for Algorithm 1 and in part (c) for Algorithm 2.

From Table 6.2(a) it can be seen that the performance of Horprasert algorithm declines between -1.75% and -5.67% on average. The performance of Algorithm 1 declines between -1.16% and -3.51% on average, as can be seen from Table 6.2(b). However, table 6.2(c) shows that the performance of Algorithm 2 declines only between -0.11% and -0.24% on average.

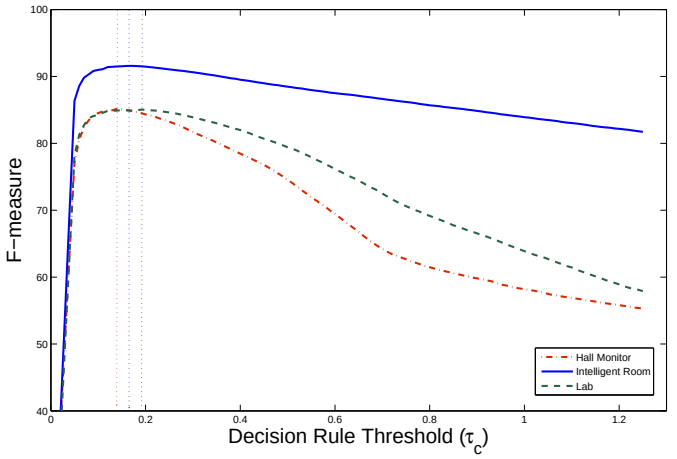
These results can be visualised from Figure 6.7, by comparing the peaks of different F-measure curves for the three sequences. For Algorithm 2 the peaks are clearly coincident (occur at similar range of threshold values), while the peaks of the Horprasert algorithm and Algorithm 1 are much more dispersed. This shows that Algorithm 2 is more flexible and, once optimised for one environment, it can maintain its performance when applied to different environments.



(a) Horprasert Algorithm



(b) Algorithm 1



(c) Algorithm 2

Figure 6.7: F-measure vs. Threshold for threshold sensitivity comparison

Table 6.2: Threshold Sensitivity comparison

(a)Horprasert

	'Intelligent Room' Threshold	'Hall Monitor' Threshold	'Lab' Threshold	Min	Max
'Intelligent Room'	85.39%	83.97%	79.38%	-1.42%	-6.01%
'Hall Monitor'	74.52%	77.80%	75.77%	-2.03%	-3.28%
'Lab'	75.03%	80.81%	82.60%	-1.79%	-7.57%
Mean				-1.75%	-5.62%

(b)Algorithm 1

	'Intelligent Room' Threshold	'Hall Monitor' Threshold	'Lab' Threshold	Min	Max
'Intelligent Room'	87.84%	85.80%	82.85%	-2.04%	-4.99%
'Hall Monitor'	81.98%	82.14%	78.53%	-0.16%	-3.61%
'Lab'	81.84%	82.50%	83.77%	-1.27%	-1.93%
Mean				-1.16%	-3.51%

(c)Algorithm 2

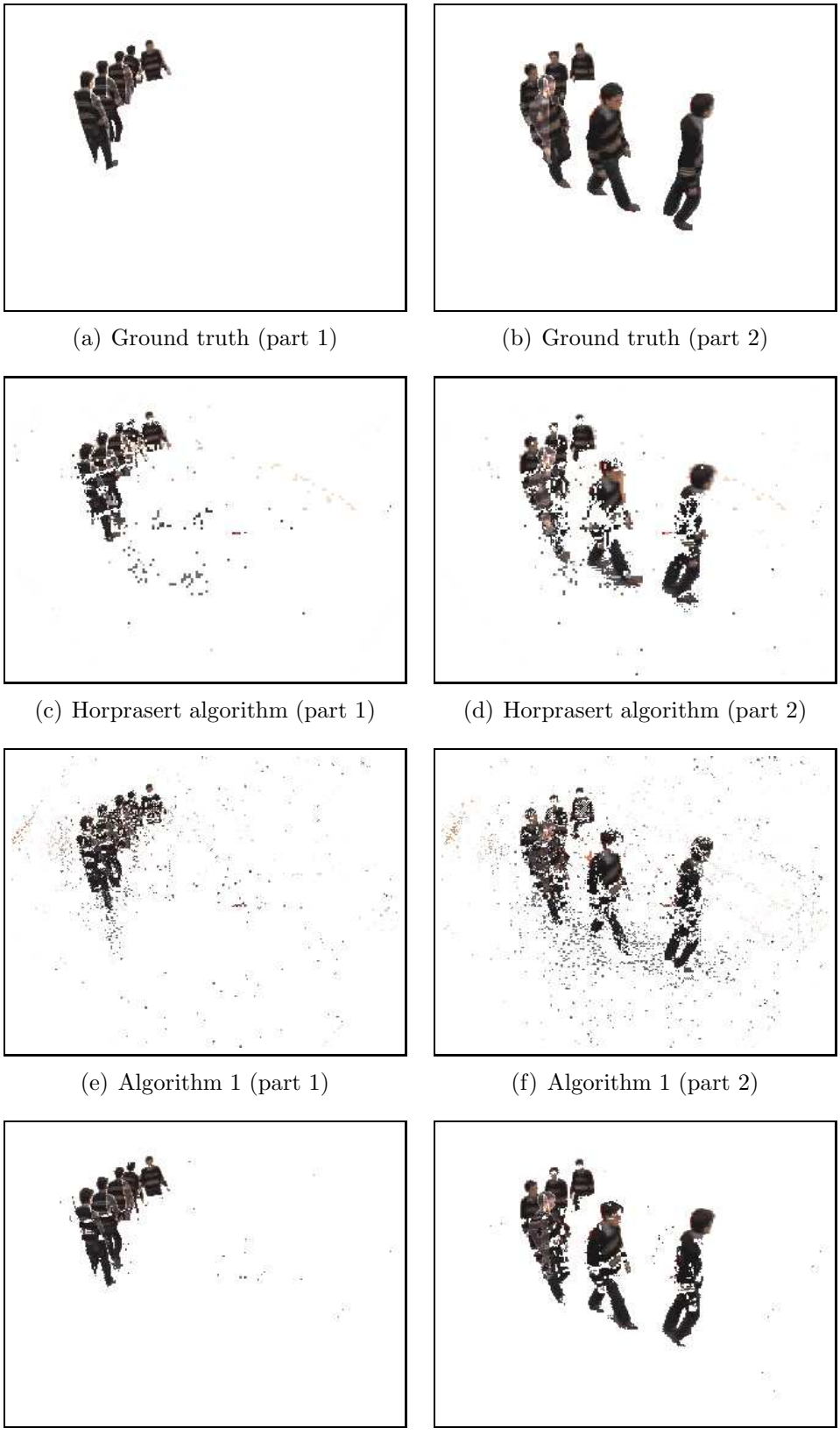
	'Intelligent Room' Threshold	'Hall Monitor' Threshold	'Lab' Threshold	Min	Max
'Intelligent Room'	91.50%	91.50%	91.50%	0%	0%
'Hall Monitor'	84.51%	85.10%	84.86%	-0.24%	-0.59%
'Lab'	84.86%	84.9%	85.00%	-0.1%	-0.14%
Mean				-0.11%	-0.24%

6.6 Illustrative visual comparison

The results are evaluated by visual comparison between one of the state-of-the-art techniques (e.g. Horprasert [73]) and the proposed algorithms. The segmentation result is compared visually to reference segmentation. This reference segmentation represents the ground truth which is generated manually. The optimal threshold values obtained from the PRF curves are used. The same noise removal procedure, discussed in 5.3.2 is applied for all outputs.

Due to space constraints, only certain frames from the processed sequences are presented.

Figures 6.8 and 6.9 divide the 'Intelligent Room' scenario into two parts. Part 1 shows a combination of five sample frames which represents movement of the participant from the door to the far end of the room. Part 2 is a combination of six



(a) Ground truth (part 1)

(b) Ground truth (part 2)

(c) Horprasert algorithm (part 1)

(d) Horprasert algorithm (part 2)

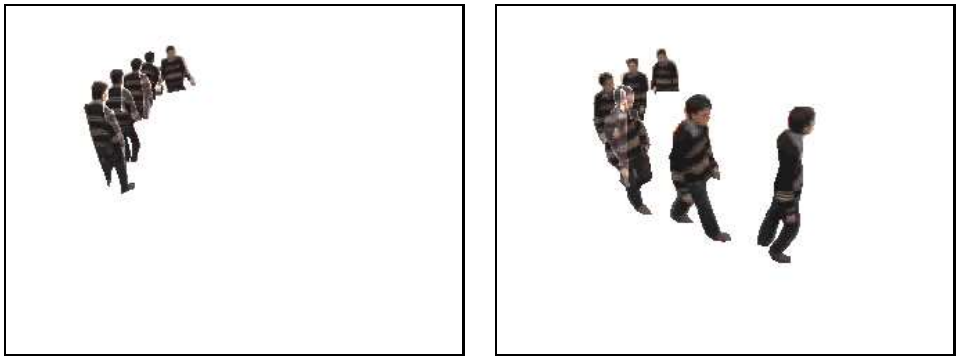
(e) Algorithm 1 (part 1)

(f) Algorithm 1 (part 2)

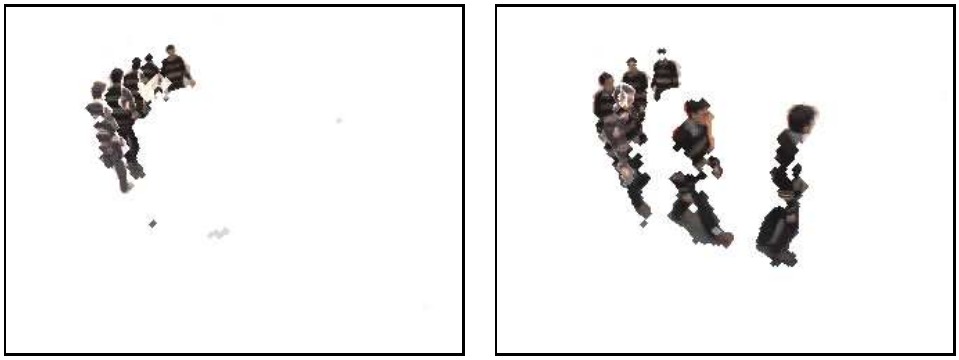
(g) Algorithm 2 (part 1)

(h) Algorithm 2 (part 2)

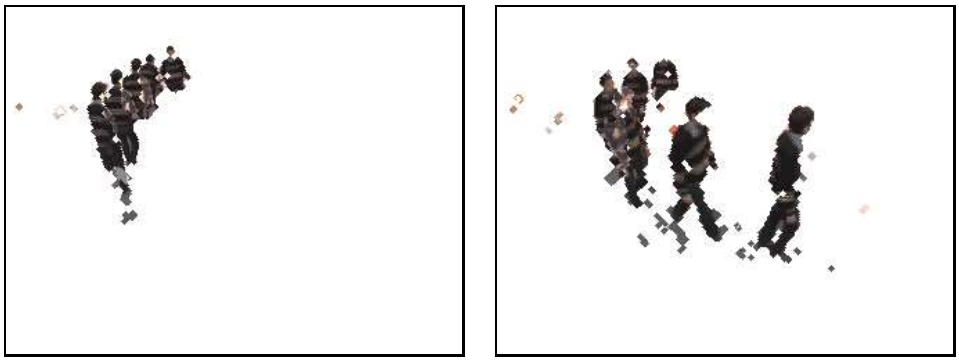
Figure 6.8: Algorithms comparison for 'Intelligent Room' sequence (without noise removal)



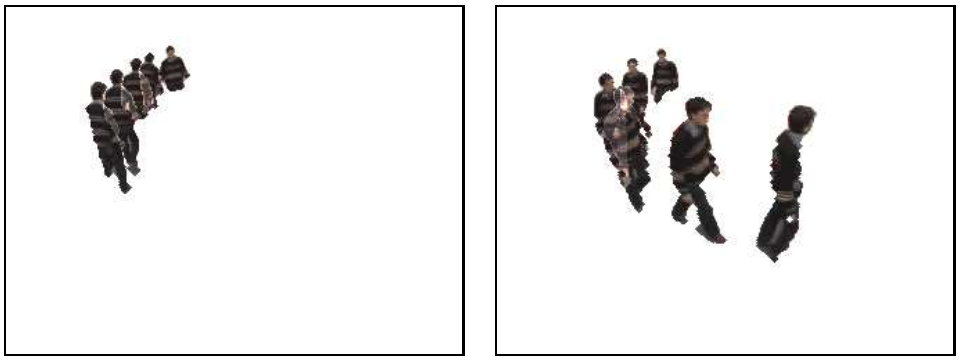
(a) Ground truth (part 1) (b) Ground truth (part 2)



(c) Horprasert algorithm (part 1) (d) Horprasert algorithm (part 2)



(e) Algorithm 1 (part 1) (f) Algorithm 1 (part 2)



(g) Algorithm 2 (part 1) (h) Algorithm 2 (part 2)

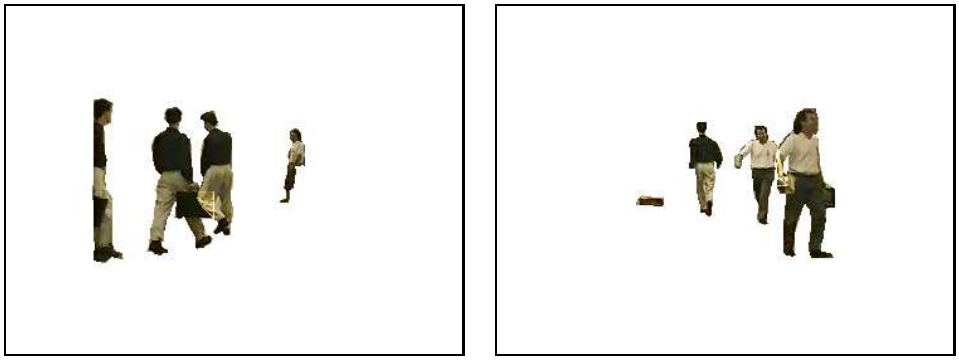
Figure 6.9: Algorithms comparison for 'Intelligent Room' sequence (with noise removal)

frames showing the movement of the participant from the far end to the front end of the room.

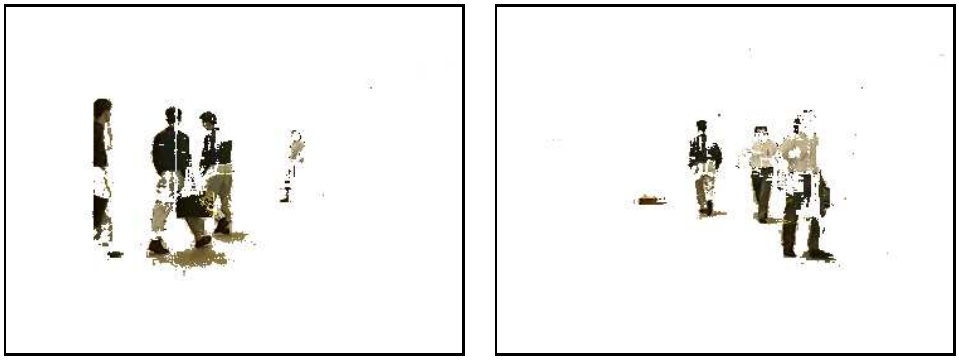
Figure 6.8 (a) and (b) show the hand-labelled ground truth of Part 1 and 2 of the Intelligent room equence respectively. Figure 6.8 (c) and (d) show the output of Horprasert algorithm without noise removal for Part 1 and 2 respectively. As can be seen in the figure, the segmented participant is fragmented and parts of the shadows are included, as can be seen below the participant. Figure 6.8 (e) and (f) show the output of Algorithm 1 without noise removal for Part 1 and 2 respectively. As expected, the shadows have been reduced significantly due to the inclusion of a shadow detection module which tries to eliminate shadows from the segmented objects. The output of Algorithm 2 for Part 1 and 2 without noise removal are shown in Figure 6.8 (g) and (h) respectively. The segmented objects in Figure 6.8 (g) are much closer to the ground truth shown in Figure 6.8 (a) than those of Horprasert algorithms and Algorithm 1. The same applies for Figure 6.8 (h), however the segmented object is more fragmented due to the fact that the participant is closer to the camera. A very small amount of isolated noise and shadows have been captured.

Figure 6.9 shows the output of the three algorithms after the noise removal stage. The noise disappears from Algorithm 2; however the output of the Horprasert algorithm and Algorithm 1 still suffers from some noise. The output of Horprasert algorithm includes shadows. The segmented participant is fragmented in both Algorithm 1 and Horprasert as Figures 6.9 (c) and (e) show. Algorithm 2 output shows much better segmentation in terms of less fragmentation and fewer shadows beeing captured, and is very similar to the ground truth output.

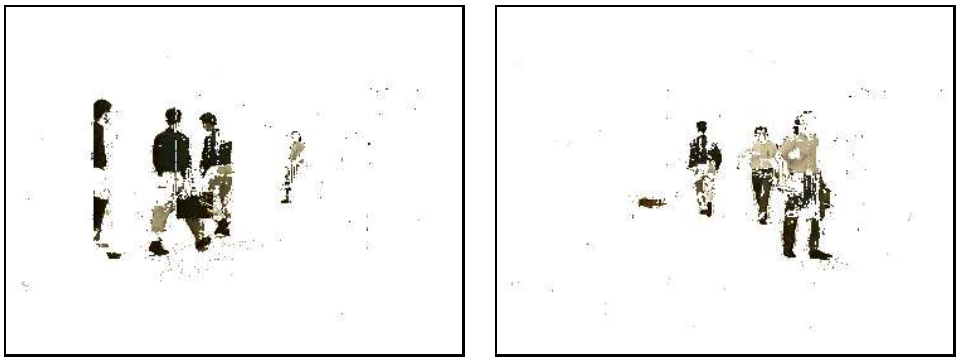
Figure 6.10 (a) and (b) show the hand-labelled ground truth of Part 1 and 2 of the Hall Monitor sequencerespectively. Figure 6.10 (c) and (d) show the output of Horprasert algorithm without noise removal for Part 1 and 2 respectively. As can be seen in the figure, the segmented participants are fragmented and parts of the shadows are included. Figure 6.10 (e) and (f) show the output of Algorithm 1 without noise removal for Part 1 and 2 respectively. As expected, the shadows have been reduced significantly, due to the inclusion of a shadow detection module which tries to eliminate shadows from the segmented objects. However, this has an



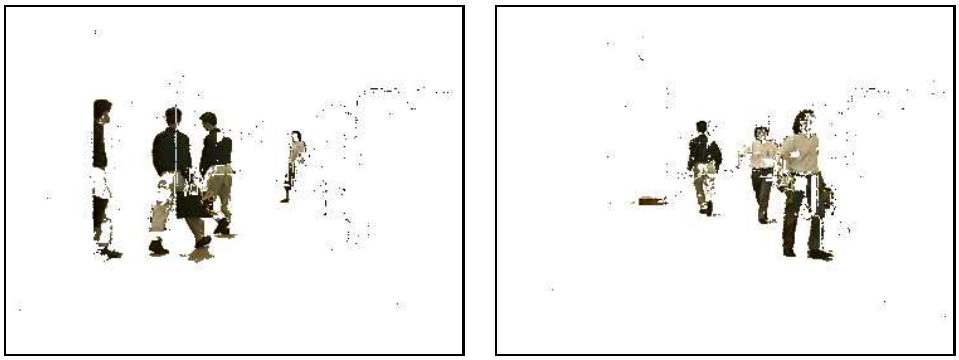
(a) Ground truth (part 1) (b) Ground truth (part 2)



(c) Horprasert algorithm (part 1) (d) Horprasert algorithm (part 2)

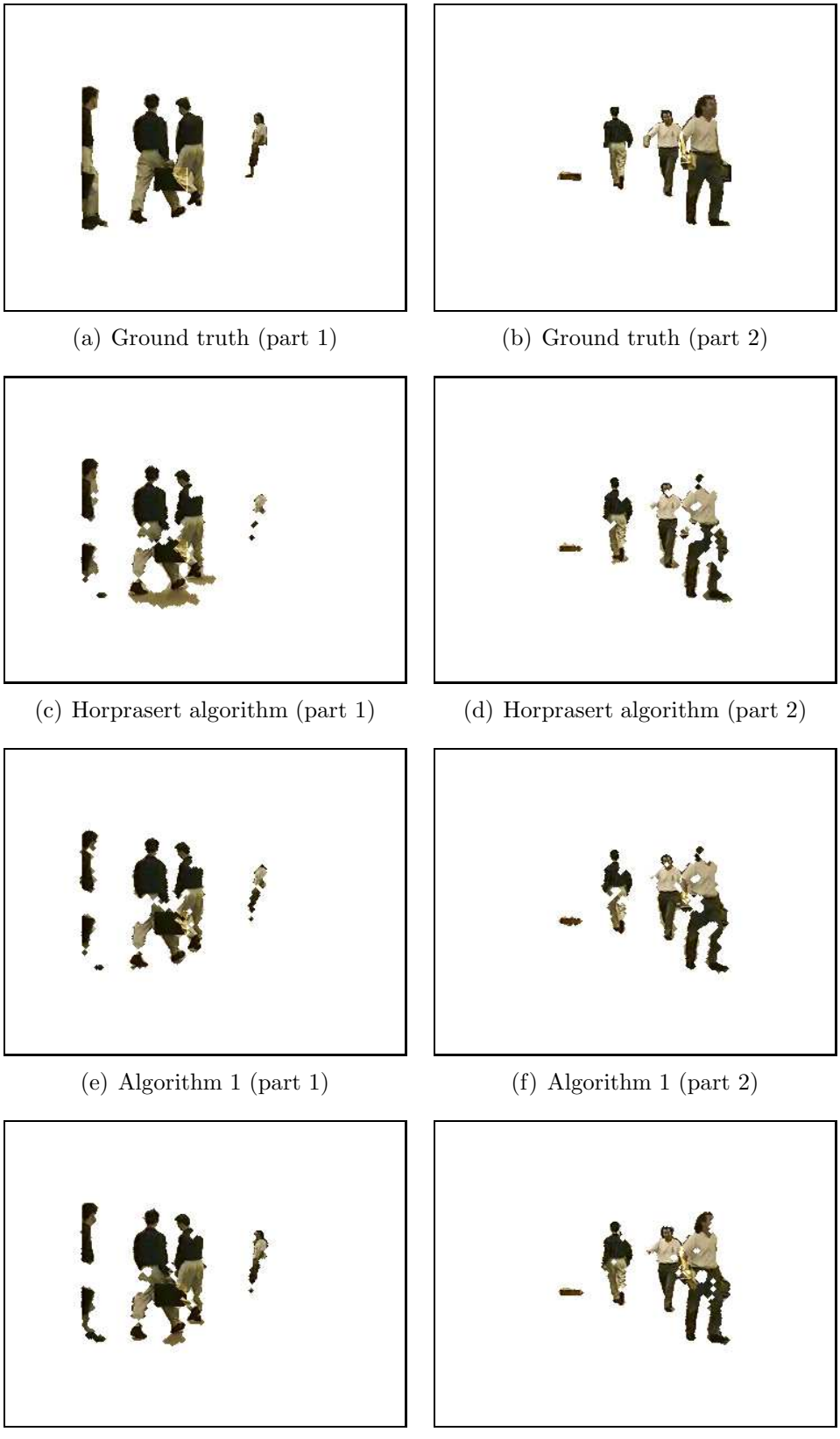


(e) Algorithm 1 (part 1) (f) Algorithm 1 (part 2)



(g) Algorithm 2 (part 1) (h) Algorithm 2 (part 2)

Figure 6.10: Algorithms comparison for 'Hall Monitor' sequence (without noise removal)



(a) Ground truth (part 1)

(b) Ground truth (part 2)

(c) Horprasert algorithm (part 1)

(d) Horprasert algorithm (part 2)

(e) Algorithm 1 (part 1)

(f) Algorithm 1 (part 2)

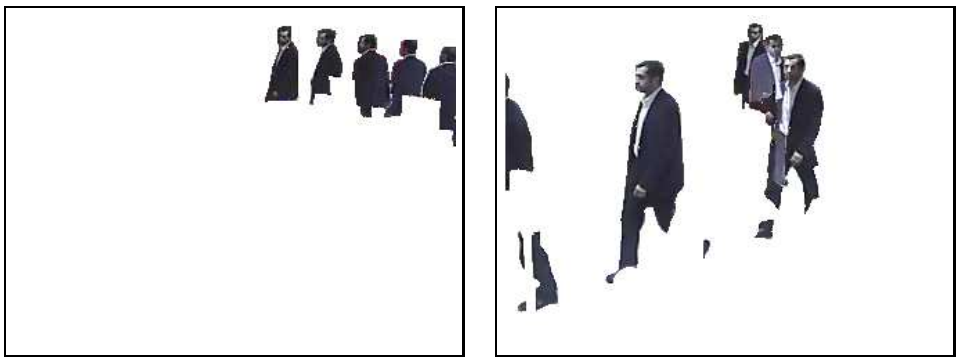
(g) Algorithm 2 (part 1)

(h) Algorithm 2 (part 2)

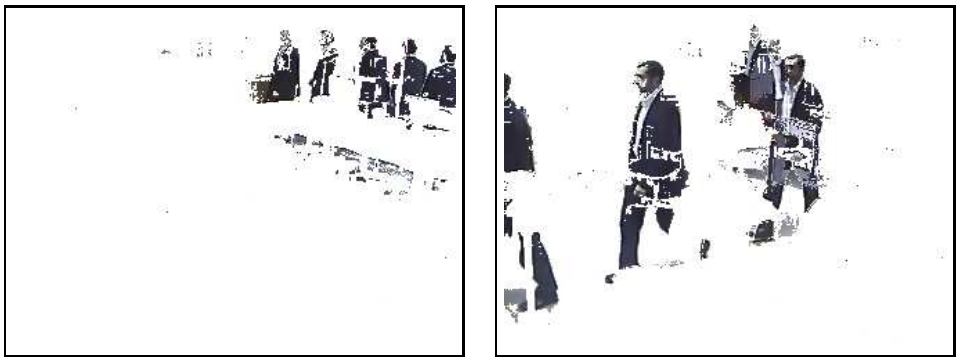
Figure 6.11: Algorithms comparison for 'Hall Monitor' sequence (with noise removal)

impact upon the segmentation as the removal of shadows from foreground objects means that the participants appear more fragmented. A future improvement to this algorithm is to exploit geometric shadow properties related to shadow boundaries and to the adjacency of each object and its cast shadow as in [18]. The output of Algorithm 2 for Part 1 and 2 without noise removal are shown in Figure 6.10 (g) and (h) respectively. Although the segmented objects are less fragmented, a few shadows are captured. However the shadows captured by Algorithm 2 are not as significant as in Horprasert algorithm. Algorithm 1 and 2 include a small amount of isolated noise.

Figure 6.11 shows the output of the three algorithms after the noise removal stage. The noise disappears from all the algorithms; however the output of the Horprasert algorithm still suffers from intensive shadow. The segmented participants are fragmented in both Algorithm 1 and Horprasert algorithm e.g. the segmented participant B is largely corrupted as Figures 6.11 (c) and (e) show. Algorithm 2 output shows better segmentation in terms of less fragmentation and fewer shadows being captured. The lab scenario is divided into two parts in Figure 6.12 which shows the sequence without noise removal and Figure 6.13 which shows the sequence with noise removal. Part 1 represents a combination of five sample frames which represents the movement of the participant as he enters the scene from the far right of the scene. Part 2 is a combination five frames showing the movement of the participant as he walks to the near left of the scene. Figure 6.12 (a) and (b) show the hand-labelled ground truth of Part 1 and 2 respectively. Figure 6.12 (c) and (d) show the output of Horprasert algorithm without noise removal for Part 1 and 2 respectively. As can be seen in the figure, the segmented participant is fragmented and parts of the shadows are included. Figure 6.12 (e) and (f) show the output of Algorithm 1 without noise removal for Part 1 and 2 respectively. As expected, the shadows have been reduced significantly. However, the participant appears more fragmented. The output of Algorithm 2 for Part 1 and 2 without noise removal are shown in Figure 6.12 (g) and (h) respectively. Although the segmented objects are less fragmented than for Algorithm 1, a small amount of shadow is captured. However the shadows captured by Algorithm 2 are not as significant as in Horprasert



(a) Ground truth (part 1) (b) Ground truth (part 2)

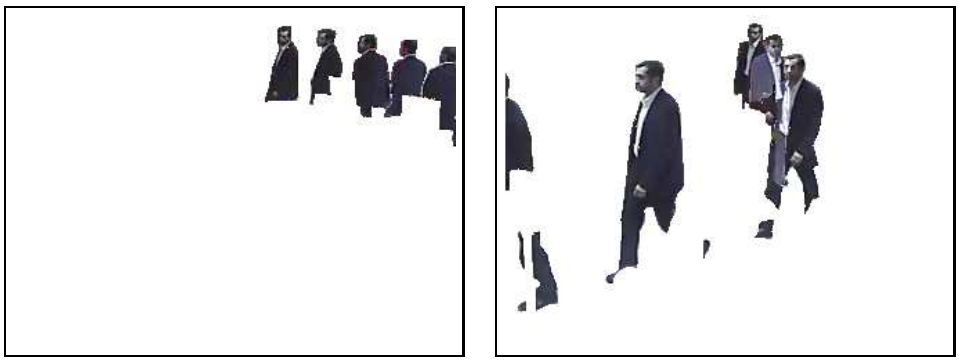


(c) Horprasert algorithm (part 1) (d) Horprasert algorithm (part 2)

(e) Algorithm 1 (part 1) (f) Algorithm 1 (part 2)

(g) Algorithm 2 (part 1) (h) Algorithm 2 (part 2)

Figure 6.12: Algorithms comparison for 'Lab' sequence (without noise removal)



(a) Ground truth (part 1) (b) Ground truth (part 2)

(c) Horprasert algorithm (part 1) (d) Horprasert algorithm (part 2)

(e) Algorithm 1 (part 1) (f) Algorithm 1 (part 2)

(g) Algorithm 2 (part 1) (h) Algorithm 2 (part 2)

Figure 6.13: Algorithms comparison for 'Lab' sequence (with noise removal)

algorithm. Both Algorithms 1 and 2 include a small amount of isolated noise more than Horprasert algorithm.

Figure 6.13 shows the output of the three algorithms after the noise removal stage. The noise disappears from all the algorithms; however the output of the Horprasert algorithm still suffers from intensive shadow. The segmented participant is fragmented in both Algorithm 1 and Horprasert algorithm. The three algorithm achieve comaratively similar performance, however, the object structures are more clear in the output of Algorithm 2.

6.7 Objective evaluation

In this section, the objective evaluation metric defined in Section 6.2 is used to compare and rank the proposed change detection algorithms with the Horprasert algorithm. These results complement the illustrative visual comparison results in Section 6.6 and those presented in Chapter 5. The evaluation of the accuracy of detection is carried out using $Precision$, $Recall$ and the overall metric $F-measure$, before and with noise removal. The same noise removal procedure, discussed in 5.3.2 is applied for all outputs. The optimal threshold values obtained from the PRF curves are used.

By analysing the F-measure without noise removal, change detection techniques could be ranked as shown in Figures 6.15(a), 6.14(a) and 6.16(a). The best performance on average is shown by Algorithm 2 with $74.98 - 88.94\%$, followed by Horprasert algorithm with $72.04 - 76.35\%$, then Algorithm 1 with $70.26 - 77.93\%$.

The analysis of the F-measure with noise removal, change detection techniques could be ranked as shown in Figures 6.15(b), 6.14(b) and 6.16(b). The best performance on average is shown by Algorithm 2 with $84.5 - 91.5\%$, followed by Algorithm 1 with $82.1 - 87.8\%$, then Horprasert algorithm with $77.8 - 85.3\%$.

The algorithm with the lowest oversegmentation (the highest Precision) is Algorithm 2 with $78.9 - 88.5\%$, it is followed by Horprasert with $77.4 - 84.9\%$, then Algorithm 1 with $77.8 - 80\%$. On the other hand, the algorithm with the lowest undersegmentation (the highest Recall) is Algorithm 2 with $84 - 97.3\%$, it

is followed by Algorithm 1 with $86.5 - 94.8\%$, then Horprasert with $77.4 - 84.9\%$.

The illustrative visual comparison performed in Section 6.6, supports the results of the objective evaluation. It can be seen that Algorithm 2 achieved the best segmentation quality with both low false positives and false negatives.

The results achieved show that Algorithm 1 enhances by an average increase of the recall by 8.85%, while the average decrease of the Precision declines by 1.45%, However the overall performance is improved by an average increase of 2.7%.

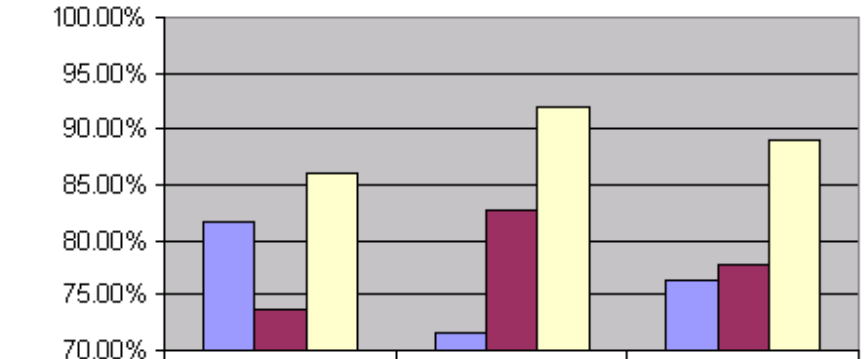
The results achieved by Algorithm 2 show that the proposed method enhances by an average increase of the recall by 7.45%, the Precision by 2.5% and the overall performance is improved by 7.85%, this can be considered a significant ratio.

It is possible to notice that the use of the reflectance ratio to represent the image as in Algorithm 1, and the decomposition of the pixel colour into chromaticity and luminance as in Horprasert algorithm leads to comparable results. the results of the two above mentioned techniques are however less accurate than those of the proposed Algorithm 2 that is based on surface spectral reflectance that better succeeds in detecting moving objects under varying light conditions.

6.8 Computational complexity evaluation

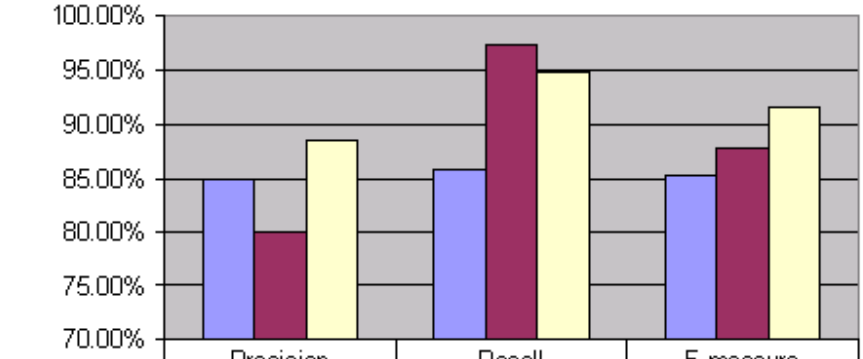
Computational complexity is important for change detection algorithms, especially for real-time applications where limited computational power is available and a given processing time must be attained. Computational complexity can be divided into two parts: time and memory consumption. A trade-off between both parts exists. Large data structures can be used to increase the speed at the cost of higher memory requirements and the reverse is also true. Time consumption is considered by measuring the processing time for the segmentation of one frame.

The computational complexity for the proposed algorithms is evaluated in terms of average processing time for the segmentation of one frame, time required to build the background model and memory consumption. Memory consumption is considered in terms of the amount of memory that is used to store one background model.



	Precision	Recall	F-measure
Horprasert	81.72%	71.64%	76.35%
Algorithm 1	73.71%	82.67%	77.93%
Algorithm 2	86.05%	92.03%	88.94%

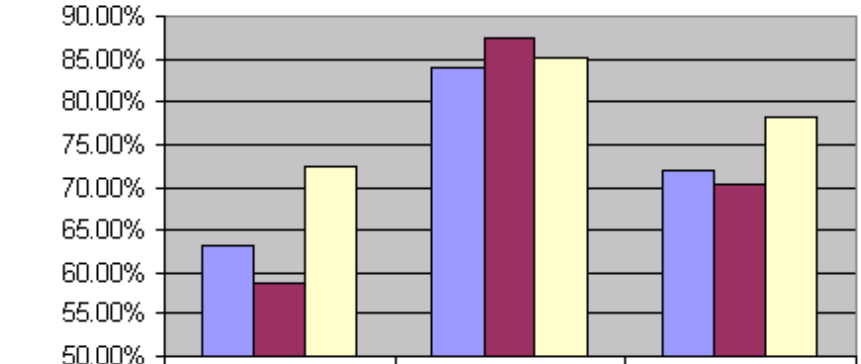
(a) without noise removal



	Precision	Recall	F-measure
Horprasert	84.90%	85.80%	85.30%
Algorithm 1	80.00%	97.30%	87.80%
Algorithm 2	88.50%	94.80%	91.50%

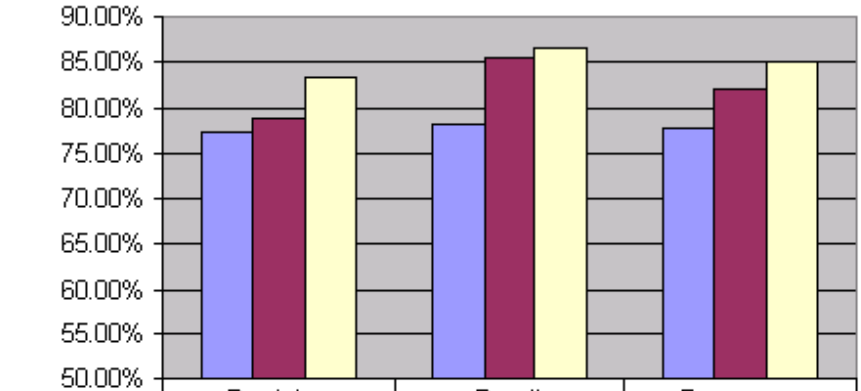
(b) with noise removal

Figure 6.14: PRF for the 'Intelligent Room' sequence



	Precision	Recall	F-measure
Horprasert	63.10%	83.94%	72.04%
Algorithm 1	58.73%	87.41%	70.26%
Algorithm 2	72.36%	85.15%	78.24%

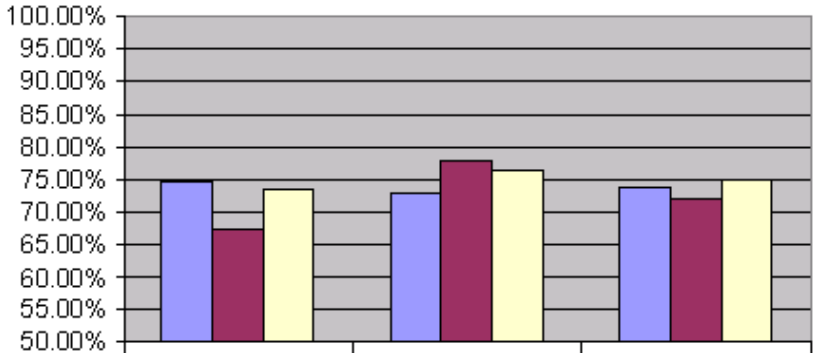
(a) without noise removal



	Precision	Recall	F-measure
Horprasert	77.40%	78.10%	77.80%
Algorithm 1	78.90%	85.50%	82.10%
Algorithm 2	83.30%	86.50%	85.10%

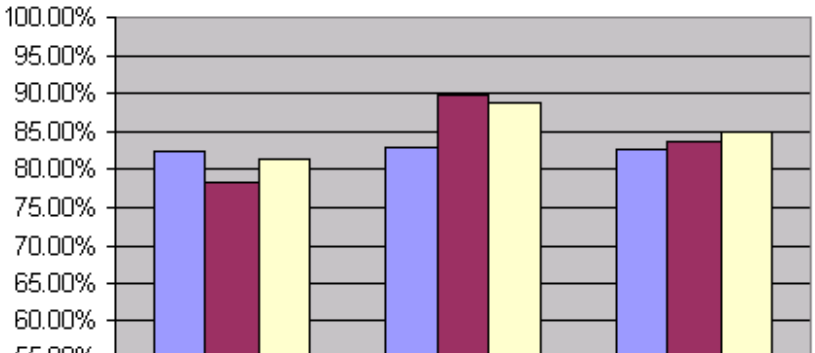
(b) with noise removal

Figure 6.15: PRF for the 'Hall Monitor' sequence



	Precision	Recall	F-measure
Horprasert	74.75%	72.97%	73.85%
Algorithm 1	67.32%	77.80%	72.18%
Algorithm 2	73.50%	76.51%	74.98%

(a) without noise removal



	Precision	Recall	F-measure
Horprasert	82.30%	83.00%	82.60%
Algorithm 1	78.40%	89.80%	83.70%
Algorithm 2	81.40%	88.90%	85.00%

(b) with noise removal

Figure 6.16: PRF for the 'Lab' sequence

The results presented in this thesis are obtained using a Xeon personal computer (processor:Intel Xeon 3GHz, memory: 2GB DDR RAM), an implementation in Matlab (version: 7.4, release: R2007a). All values are stored using double Precision (32 bit). Karaman et al. [174] carried out some implementations of change detection methods in C++ and Matlab, they concluded that the overall increase in speed is of factor 13 between Matlab (version: 7, release: 14) and C++(compiler: Visual C++ 7.1).

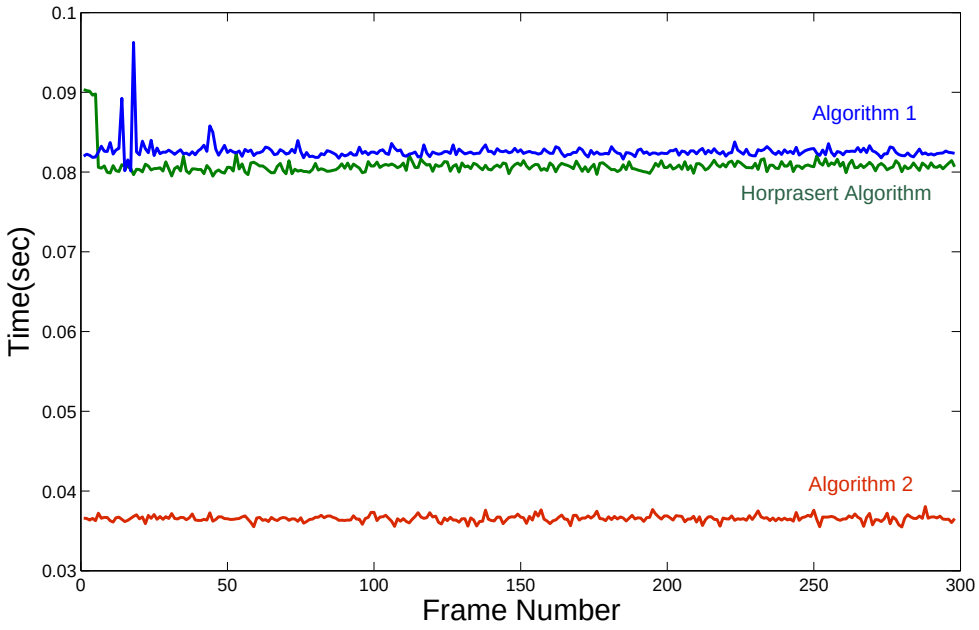
The memory consumption is proportional to the frame size of the video sequence. However the processing time is independent of the video sequence. The results obtained from the CIF 'Hall Monitor' video sequence are considered in this evaluation, each frame is 352×288 pixels.

Concerning the time consumption, Figures 6.17(a) and 6.18(a) show the processing time of the 300 frames of the 'Intelligent Room' sequence for all three algorithms, before and with noise removal respectively. Figures 6.17(b) and 6.18(b) show the average processing time over the whole video sequence, before and with noise removal respectively. The fastest algorithm is Algorithm 2 with 36.6 msec without noise removal and 40.8 msec with noise removal. The second is Horprasert algorithm with 81 msec without noise removal and 85.3 msec with noise removal. The third is Algorithm 1 with 82.7 msec without noise removal and 87.3 msec with noise removal.

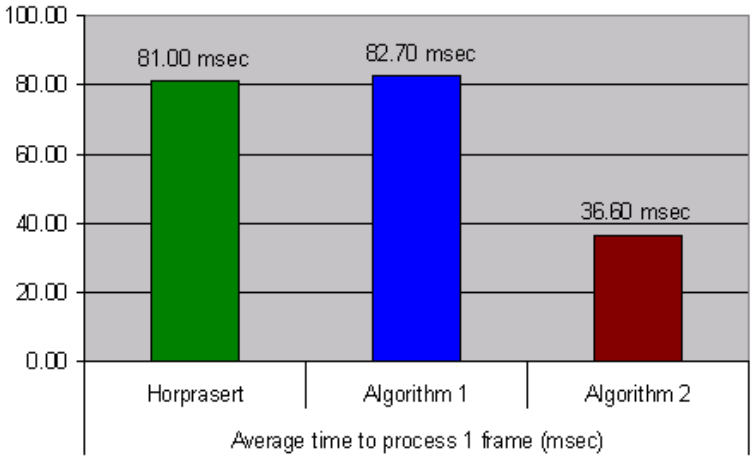
As it was expected, the speed of Algorithm 2 is approximately double the speed of the other algorithms. The reason is that both the other algorithms rely heavily on statistical models in the segmentation process, however Algorithm 2 is based on pre-trained models (the surface spectral reflectance models) and less sophisticated statistical manipulation.

Figure 6.19 shows the time required to build the background model using 50 background frames. Algorithm 2 requires 8.87 sec, Horprasert algorithm requires 9.98 sec and Algorithm 1 requires 17.5 sec. This is due to the computational complexity of the implementation of Mahalnobis distance in Algorithm 1.

Regarding the memory consumption, the amount of memory that is needed to store one background model for Algorithm 2 and for Horprasert algorithm is 3960

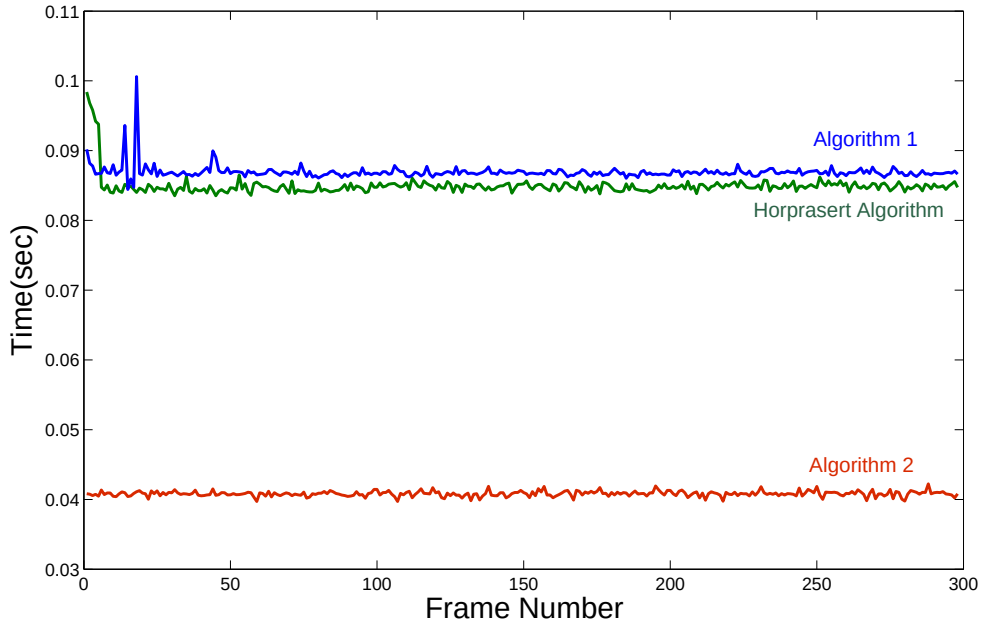


(a) 300 frames

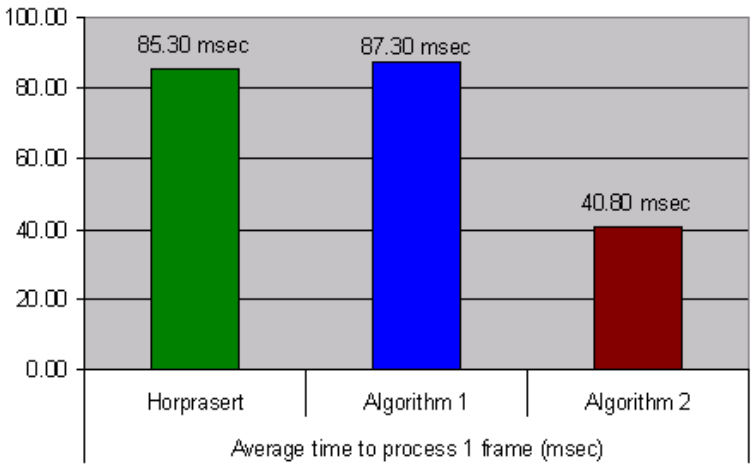


(b) Average

Figure 6.17: Processing time comparison (without noise removal)



(a) 300 frames



(b) Average

Figure 6.18: Processing time comparison (with noise removal)

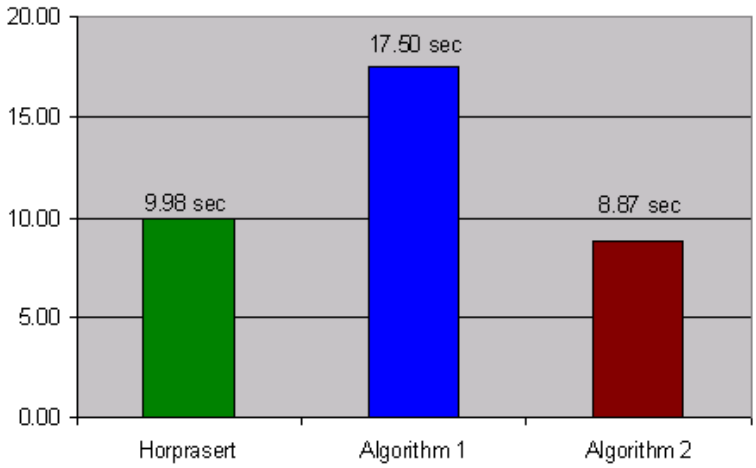


Figure 6.19: Time to build the background model (50 frames) comparison

Table 6.3: Computational complexity comparison

	Average Time (for one frame)	Time for Background Model (50 frames)	Memory for storing (the background model)
Horprasert	85.3 msec	9.98 sec	3960 Kbytes
Algorithm 1	87.3 msec	17.5 sec	9240 Kbytes
Algorithm 2	40.8 msec	8.87 sec	3960 Kbytes

Kbytes, however Algorithm 1 requires 9240 Kbytes. More memory is requires for Algorithm 1 due to the inclusion of the covariance matrix in the background model, which is part of the calculation of Mahalanobis distance.

Table 6.3 shows a comparison between all three algorithms in terms of the average processing time required for the segmentation of one frame(with noise removal), time to build the background model using 50 frames and the memory required for storing the background frame.

6.9 Proposed assessment criterion

Change detection algorithms based on background modelling require a number of static background frames to build their statistical background model. As discussed in Section 3.2, in real-world operational conditions sometimes it is difficult to capture a sufficient number of static background frames. In such situations, the algorithm

has to deal with the available number of background frames. If the system fails at any point, i.e due to changes in illumination, the algorithm has to rebuild its model. The system needs to wait for the required number of background frames to be captured and processed. During this period the system may miss some events. In normal operation, the background maintenance module updates the model with the scene variations, with an adaptation parameter α. This parameter is closely related to the sensitivity of the algorithm to the scene contents. If the system is able to capture those variations using fewer frames the adaptation process will be faster, which will affect the overall performance of the system.

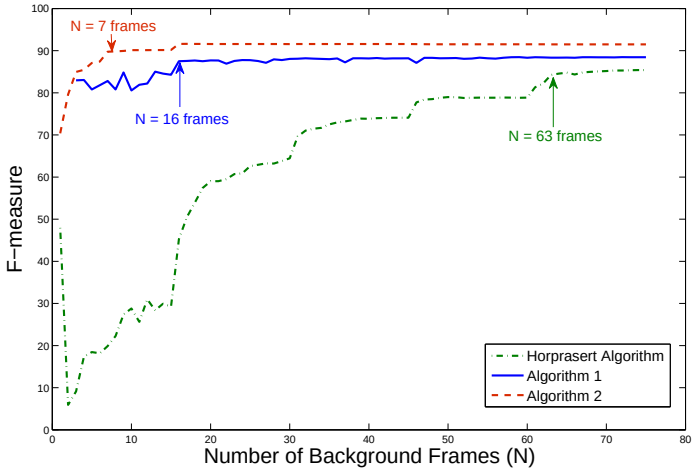
This thesis proposes the minimum number of background frames required by the algorithm to achieve its maximum performance to be used as a new evaluation metric. This new metric is important for the evaluation of change detection algorithms because it demonstrates the effectiveness of the algorithm, and underpins the robustness of the algorithm to scene variations.

The use of such metric can give a better indication of the time required by a specific algorithm to build its background model. When it comes to the comparison between different algorithms, instead of specifying a fixed number of background frames, each algorithm will be featured with its minimum number of background frames and this will be used to calculate the time required to build the model.

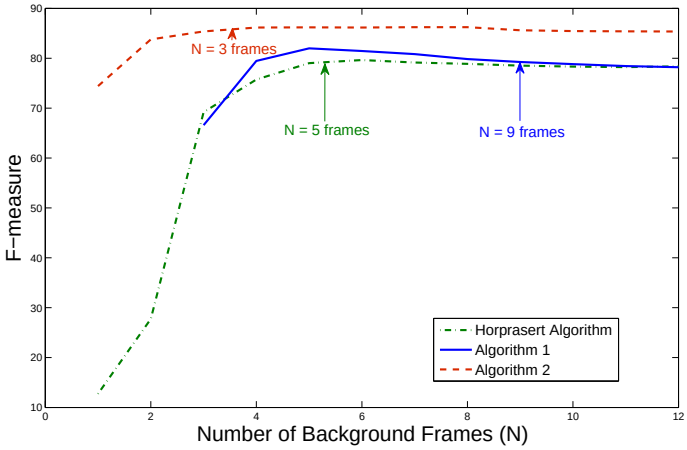
To the knowledge of the author, there is no evaluation criterion which identifies the minimum number of background frames required. In this criterion, it is proposed to plot the F-measure versus the number of background frames used to build the background model. The minimum number of background frames is identified from the curve as the minimum number of background frames that can achieve the maximum F-measure, as Figure 6.20 shows.

For the 'Intelligent Room' sequence [172], while the Horprasert algorithm requires a minimum of 63 frames to achieve 85.3% (its maximum F-measure), Algorithm 1 requires a minimum of 16 frames to achieve 87.8% (its maximum F-measure), where as the Algorithm 2 requires a minimum of 7 frames to achieve 91.5% (its maximum F-measure).

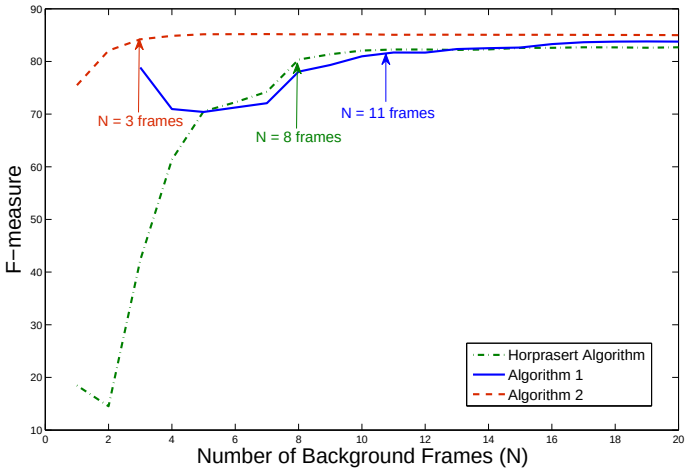
For the 'Hall Monitor' sequence, while Horprasert algorithm requires a minimum



(a) 'Intelligent Room' Sequence



(b) 'Hall Monitor' sequence



(c) 'Lab' sequence

Figure 6.20: F-measure vs. number of background frames

of 5 frames to achieve 81.45% (its maximum F-measure), the Algorithm 1 requires a minimum of 9 frames to achieve 84.4% (its maximum F-measure), where as the Algorithm 2 requires a minimum of 3 frames to achieve 86.53% (its maximum F-measure).

For the 'Lab' sequence, while Horprasert algorithm requires a minimum of 8 frames to achieve 82.6% (its maximum F-measure), the Algorithm 1 requires a minimum of 11 frames to achieve 83.7% (its maximum F-measure), where as the Algorithm 2 requires a minimum of 3 frames to achieve 85% (its maximum F-measure).

6.10 Summary

In this chapter the proposed physics-based segmentation algorithms have been quantitatively evaluated. Three evaluation methods are adopted: illustrative visual comparison, objective, and computational complexity evaluations.

The optimal decision rule thresholds have been obtained using tow different criteria: the minimax criterion using ROC curves and the maximum F-measure criteria. A number of illustrative visual examples have been discussed. Manually segmented ground truth data have been used and pixel-based error measures have been applied for objective evaluation.

The method proposed by Horprasert et al. [73] is chosen, based on the comparison of various state-of-the-art methods in [174]. This method is quite robust and its complexity is rather low in comparison to other approaches, allowing for real-time performance. A comparison between the proposed methods and the state-of-the-art change detection method, (Horprasert [73]), for several types of video sequences was conducted.

This chapter proposes a new assessment criterion for change detection algorithms. The new criterion includes the calculation of the minimum number of background frames required by an algorithm to achieve its maximum performance. A more comprehensive comparison between different algorithms is achieved by plotting the F-measure versus the number of background frames used to build the

background model.

The proposed algorithms have been evaluated by means of objective quality assessments on a variety of video data. The evaluation shows that those algorithms have achieved accurate segmentation results in different real world indoor environments. The results achieved show that the first proposed method (Algorithm 1) enhances the Recall between 6.2% and 11.5%, the Precision between -4.9% and 2% and the overall performance is improved between 1.1% and 4.3%. The Recall is enhanced significantly in Agorithm 1 due to the shadow detection module which eliminates shadows from foreground objects. This method requires approximately 87.3 msec to process one frame, which is approximately the same time required by the Horprasert method (85.3 msec). The results achieved by the second method (Algorithm 2) show that the proposed method enhances the Recall between 5.9% and 9%, the Precision between -0.9% and 5.9% and the overall performance is improved between 2.4% and 7.3%. Algorithm 2 outperforms the other two methods due to the use of the surface spectral reflectance as a cue to model the scene, which indicates its effectiveness in representing different environments. This method is twice as fast as the Horprasert method, it requires approximately 40.8 msec to process one frame. This was expected as Algorithm 2 adopts a simple statistical manipulation approach, relying on the physical parameters derived from the image. Where this algorithm shows that it needs from 3 to 8 background frames in order to achieve its maximum performance, the first algorithm requires from 9 to 16 where the Horprasert algorithm requires between 9 to 63 background frames. This means that the Horprasert algorithm requires from 1.79 to 12.57 seconds to build the background model. While algorithm 1 requires from 3.15 to 5.6 seconds, Algorithm 2 requires only from 0.17 to 0.47 seconds. Tolerance to threshold variation was evaluated by considering the performance of each algorithm using the optimal threshold derived for different sequences. This evaluation shows that the performance of Horprasert algorithm declines between -1.75% and -5.67% on average, and the performance of Algorithm 1 declines between -1.16% and -3.51% on average, however the performance of Algorithm 2 declines only between -0.11% and -0.24% on average. This shows that Algorithm

2 is more flexible and, once optimised for one environment, it can maintain its performance when applied to different environments.

The results of the comparison between the second algorithm and the other algorithms investigated in this thesis show that it matches better the operational requirements of a workplace surveillance system, in terms of real-time performance and robustness to cope with changes in illumination, scene structure and scene activity, with little or no human intervention. Moreover, this solution not only improves the robustness of the algorithm but also enhances its flexibility.

"Researchers have developed a number of sub-systems and partial solutions which go some way towards solving elements of the problem of surveillance. Much progress has been made, but in a piecemeal fashion, and often without reference to the situations in which such systems might actually be used."

M. Hannah Dee and A. Sergio Velastin [175]

7

Conclusions and Further Work

7.1 Conclusions

Video surveillance is a growing industry which is expected to be boosted by the technological advances of smart algorithms and systems, combined with further take up of digital CCTV installations and networked solutions.

Practical requirements of smart video surveillance applications are identified as segmentation, indexing, content analysis and video retrieval. These requirements are fulfilled by a number of different techniques, namely: change detection, object classification, object tracking, object classification, activity classification and scene visualisation.

The research community has addressed, with various degrees of success, the technical problems associated with these stages. However, these techniques have not reached the maturity required for unconstrained real-world applications; their value to the video surveillance industry in the near- and mid-term is minimal.

Very few of the proposed systems are up to the challenges of a real world operating environment, where robustness, adaptability, and flexibility are essential. A particular deficiency of these systems is the inadequate performance of the front-end processing sub-system. Hence, the literature is preoccupied with a lot more mundane problems at the moment. Chief among these problems is the front-end processing, which comprises elements such as change detection and object segmentation. The reason for the preoccupation with reliable front-end processing is that failures at the segmentation level can propagate downstream of the surveillance processing chain, and result in incorrect outputs which may, for example, lead to incorrect responses in critical situations.

The objective of most of video surveillance systems was originally security, however in practice they are being used for other purposes. The link between application requirements and current smart video surveillance techniques leads us to propose a new classification for practical applications of smart video surveillance. Applications are split into: security, safety, entertainment, and efficiency improvement. Efficiency improvement applications, and entertainment applications are expected to share the video surveillance infrastructure with security and safety applications.

Workplace surveillance presents a very promising application, which can make use of the existing techniques in smart video surveillance to improve the efficiency of workplace. There are practical needs for smart video surveillance systems to assist the decision makers in managing resources, in the workplace.

There is a conflict in needs, rights and understanding capabilities and limitations of the techniques between managers, people exposed to surveillance and developers of surveillance system. A 3D workplace surveillance model is proposed which links capabilities of smart video surveillance algorithms versus needs and implications of workplace surveillance applications.

The majority of workplace scenarios covers indoor environments, where the surveillance area is confined, video sensors are assumed to be near to the moving objects, and appearance or disappearance of objects can affect the scene illumination. A set of practical requirements are identified such as robustness, adaptability, flexibility as well as being a real-time system.

The challenge of using smart video surveillance techniques in order to improve the efficiency of workplace is the development of a robust segmentation (extraction) of moving or displaced objects. The change detection module is responsible for eliminating illumination variation and camera noise from changes caused by meaningful objects. Other video processing modules, in the surveillance system, rely on the effectiveness of this module.

The change detection problem has been extensively studied in the last decade, however there is no generally accepted method to detect moving objects in image sequences. Approaches based on background modelling are widely used for surveillance applications. Background modelling may be decomposed into two major steps, image representation and statistical manipulation. Change detection based solely on statistical methods fails to tackle the practical requirements of indoor applications. The literature suggests that the best combination between these two steps is a key success to develop a real-time, robust and adaptive change detection algorithm.

The review of the literature outlined the fact that the majority of the change detection approaches are non-physics-based and use well-known colour spaces to represent the image. The selection of an image representation is a crucial issue; the ability of each image representation to detect changes differs according to the camera used, the captured scene, and the illumination.

This thesis argues that understanding the underlying physics which govern the image formation process is crucial to deal with the variation in imaging parameters. Image formation models offer interesting alternative physics-based cues for foreground segmentation compared with other representations used in conventional methods.

Among the various reflection models, which describe the image formation

process, proposed in the literature; the dichromatic reflection model represents a practical choice for image processing and computer vision. The main limitation of this technique is that it can be applied only to inhomogeneous dielectrics. By applying the dichromatic reflection model for Lambertian surfaces, three main photometric invariants which are sensitive to changes in surface reflectance can be extracted: reflectance ratio, intrinsic image and surface spectral reflectance. The reflectance ratio and intrinsic image are derived from the shading model, which represents an approximated form of the dichromatic model.

Few solutions have been proposed in the literature with regard to the use of physical models of image formation to solve the change detection problem. Only approximated models, such as the shading model, have been used to derive physics-based photometric invariants, namely reflectance ratio and intrinsic images.

It can be concluded that features such as surface spectral reflectance have not been applied yet in the field of change detection, which represents a gap in the literature. The reason is the computational complexity of such models, and hence possible unfeasibility of real-time implementation.

The linear models developed in [105–107] to estimate the surface spectral reflectance give a good starting point towards the solution for such a problem. This thesis is the first attempt to investigate the use of such models in change detection.

The most significant cues for the purpose of designing a robust change detection approaches have been selected and investigated. In the context of this thesis, 'significant' means that they can be exploited based only on image-derived information and with a limited number of assumptions about the scene, and they are suitable for real-time processing. The validity of the approach has been demonstrated through two novel algorithms which are based on a physical meaning of image formation and which use advances in colour constancy techniques.

The first algorithm is a change detection algorithm based on reflectance ratio. This algorithm proposes a new moving object segmentation which models the variation in illumination using the reflectance ratio between foreground pixels and background pixels. The correlation between this ratio triples are then used to

segment foreground and shadows from a static background using Mahalanobis distance. Furthermore, shadows are detected and segmented using explicit criteria which have been developed to define a method for shadow segmentation from ratio triples components using the Mahalanobis distance. This combination between an approximated physics-based model (the shading model) for image formation and a statistical method proves its ability to give a robust real-time segmentation for both foreground and shadows.

The second algorithm presents a novel physics-based image representation where a real-time transformation from RGB space to weights, which represent the surface spectral reflectance of objects, was developed. This method starts by proposing a method for highlight detection, which is then used to estimate illumination characteristics of a scene, represented by a parameter called 'correlated colour temperature', using McCamy's method. Then, the surface spectral reflectance of the objects in the scene is calculated, and the correlation between the full spectrum of a background model and surface spectral reflectance of foreground objects is adopted as the classification strategy, to segment the moving or displaced foreground objects. This method is different from those typically used in literature and this represents an element of originality of the proposed approach.

The criteria used in the evaluation method for judging the performance of the segmentation represent an essential element and critical factor in change detection evaluation. The review of the principles and methods for evaluating and comparing the performance of change detection approaches shows that there is no widely accepted evaluation test framework. Moreover, most of the goodness metrics proposed give a measure of the segmentation quality but not the robustness of the algorithm.

A new assessment criterion for change detection algorithms is proposed. A more comprehensive comparison between different algorithms is achieved by plotting the over-all performance metric (F-measure) versus the number of background frames used to build the background model. The minimum number of background frames required by the algorithm to achieve its maximum performance is proposed to be

used as an evaluation metric. This metric is important for the evaluation of change detection algorithms because it demonstrates the effectiveness of the algorithm, and underpins the robustness of the algorithm to scene variations.

A comparison between one of the state-of-the-art change detection methods and the proposed methods for several types of video sequences was conducted. The technique of Horprasert et al. [73] was adopted, for the comparison, since it gives the best trade-off between segmentation quality and computational complexity, allowing real-time performance. The objective assessment results show that the physics-based method has a higher performance over this conventional method, are more tolerent to variation of their thresholds. Moreover, the minimum number of background frames required to train the background model is much less for the physics-based methods.

The results achieved show that the first proposed method enhances the recall between 6.2% and 11.5%, the precision between -4.9% and 2%, and the overall performance is improved between 1.1% and 4.3%. The results achieved by the second method show that the proposed method enhances the recall between 5.9% and 9%, the precision between -0.9% and 5.9% and the overall performance is improved between 2.4% and 7.3%. Whereas this algorithm needs from 3 to 8 background frames in order to achieve its maximum performance, the first algorithm requires from 9 to 16 background frames and the Horprasert algorithm requires between 9 to 63 background frames.

Tolerance to threshold variation was evaluated by considering the performance of each algorithm using the optimal threshold derived for different sequences. This evaluation shows that the performance of Horprasert algorithm declines between -1.75% and -5.67% on average, and the performance of Algorithm 1 declines between -1.16% and -3.51% on average, however the performance of Algorithm 2 declines only between -0.11% and -0.24% on average. This shows that Algorithm 2 is more flexible and, once optimised for one environment, it can maintain its performance when applied to different environments.

7.2 Contribution to knowledge

The challenges which face the use of image formation models as a basis for change detection algorithms in order to target indoor workplace surveillance applications are challenges imposed by the operational requirements of the workplace application and challenges related to the mathematical complexity of the image formation models.

The challenge to develop robust change detection techniques which can deal effectively with different indoor environments is tackled by taking into consideration the physics which govern the image formation. The image is represented in the proposed change detection algorithms using physics-based cues. Information about the illumination and intrinsic features about the material contained in the scene are extracted and used to solve the problem of robustness.

The challenge related to the mathematical complexity of the image formation models is tackled. Firstly, by choosing an appropriate reflection model, the dichromatic reflection model, which proves it effectiveness in computer vision applications.

Secondly, setting a feasible set of assumptions for such model which best match the reduction of model complexity and does not contradict with real-world operational conditions. Two assumptions have been considered, diffuse-only reflection and the existence of a dominant illuminant. No assumption is made about surface geometries, surface texture, or types and shapes of shadows, objects, and background, which shows the applicability of the model for practical operational requirements.

Thirdly, to tackle the challenge of the variability of surface spectral reflectance for different materials, the linear model was adopted which consists of a number of basis functions, pre-trained from a set of materials; and weights, calculated for the object under investigation.

Fourthly, the challenge of the choice of suitable colour constancy techniques which can ensure a rapid and convenient estimation of the illumination and extraction of the surface spectral reflectance is solved by choosing an illumination estimation method (McCamy's method), which has been proven successful and useful for real-time processing. The surface spectral reflectance has been extracted

and adopted as a reference of the scene under varying illumination conditions.

The main contributions presented in this work can be divided into primary and secondary contributions.

1. The primary contributions are as follows :

   (a) A novel change detection algorithm using colour constancy techniques is proposed. The algorithm computationally estimates a consistent physics-based colour descriptor model of surface spectral reflectance from the camera output and then correlates the full-spectrum reflectance of the background and foreground pixels to segment the foreground from a static background.

   (b) A new change detection algorithm, based on a shading model, is proposed. The algorithm uses the reflectance ratio to model the illumination variation and applies the Mahalanobis distance to segment foreground and shadows from a static background.

2. The secondary contribution is :

   (a) A new classification of practical applications for smart video surveillance is presented. A new workplace surveillance model is introduced which relates the requirements of workplace surveillance applications, and their implications in relation to the capabilities of smart video surveillance technologies. A set of requirements for a workplace surveillance system using smart video surveillance techniques is proposed.

Performance assessment results show that the novel algorithm (contribution 1.a) matches better the operational requirements of a workplace surveillance system, in terms of real-time performance, flexibility and robustness to cope with changes in illumination, scene structure and scene activity, with little or no human intervention.

These match better with the operational requirements which boosts the applicability of the algorithm in real-world scenarios, in line with the 3D workplace surveillance model introduced (contribution 2.a), where technical capability and applicability are important considerations.

7.3 Recommendations for further work

The modular structure of the proposed approach to change detection makes it particularly suitable for extending its capabilities and performances. Each level of analysis can be independently extended and improved. Some directions for further work are proposed below.

- A future improvement to shadow detection module included in Algorithm 1, is to exploit geometric shadow properties related to shadow boundaries and to the adjacency of each object and its cast shadow as in [18].

- In estimating the surface spectral reflectance, in Algorithm 2, the specular reflectance component is neglected assuming diffuse only reflectance. The estimation of the surface spectral reflectance can be extended by taking into consideration both diffuse and specular reflections which would enhance the segmentation.

- In order to fulfil the practical requirements of the proposed workplace surveillance application, a suitable selection of skin detection algorithm is important. The goal of the moving skin detection module is to classify skin from non-skin areas of a moving object. The output of this phase is a skin mask, a set of connected pixels, which correspond to human skin-colour [176, 177]. The estimation of surface spectral reflectance gives us the opportunity to segment objects with same type of materials. Pixels are grouped based on their surface spectral reflectance of the object in order to extract meaningful semantic objects, where segmentation involves the grouping of pixels into regions that should represent semantic units.

- Apart from the results presented in this thesis, the proposed approaches have been tested on a variety of video data. Different video feeds, including recorded videos, live streams as well as streams from IP cameras have been used. The algorithms have been applied to a large number of video sequences, outside the scope of this work for testing purposes, showing a range of different indoor and outdoor scenes with different levels of complexity. Tests were conducted

on image sequences containing targets of interest in a variety of environments, e.g., offices, public buildings, subway stations, campuses, parking lots, airports, and sidewalks. All of the obtained results are consistent with those presented in the evaluation chapter. A normal extension, is the application of such algorithms for other application areas.

A

Computational Spectral Reflectance

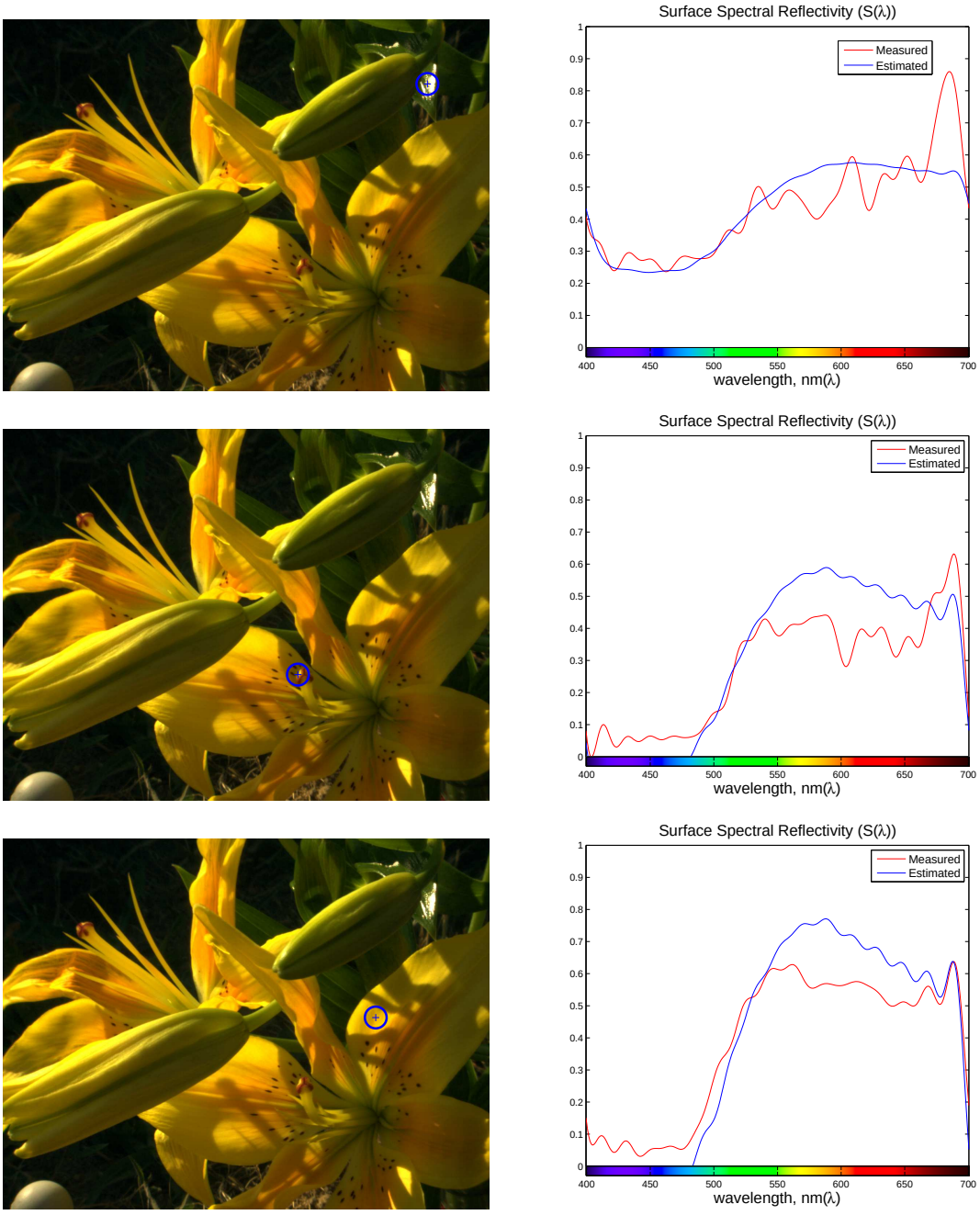
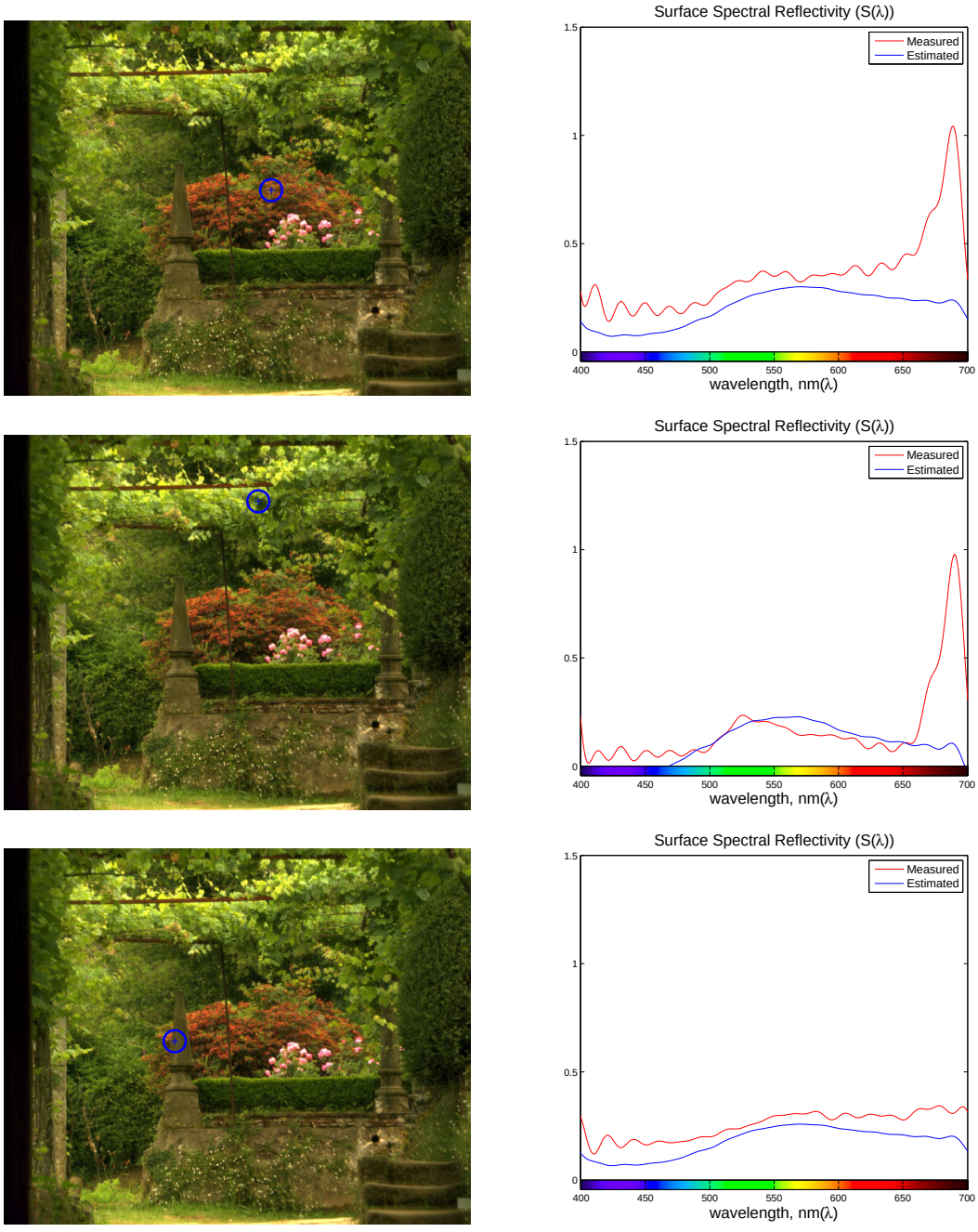


Figure A.1: Samples of measured and estimated surface spectral reflectance for Foster data set 1



(e) Scene no. 3 (f) Spectral Reflectance

Figure A.2: Samples of measured and estimated surface spectral reflectance for Foster data set 2

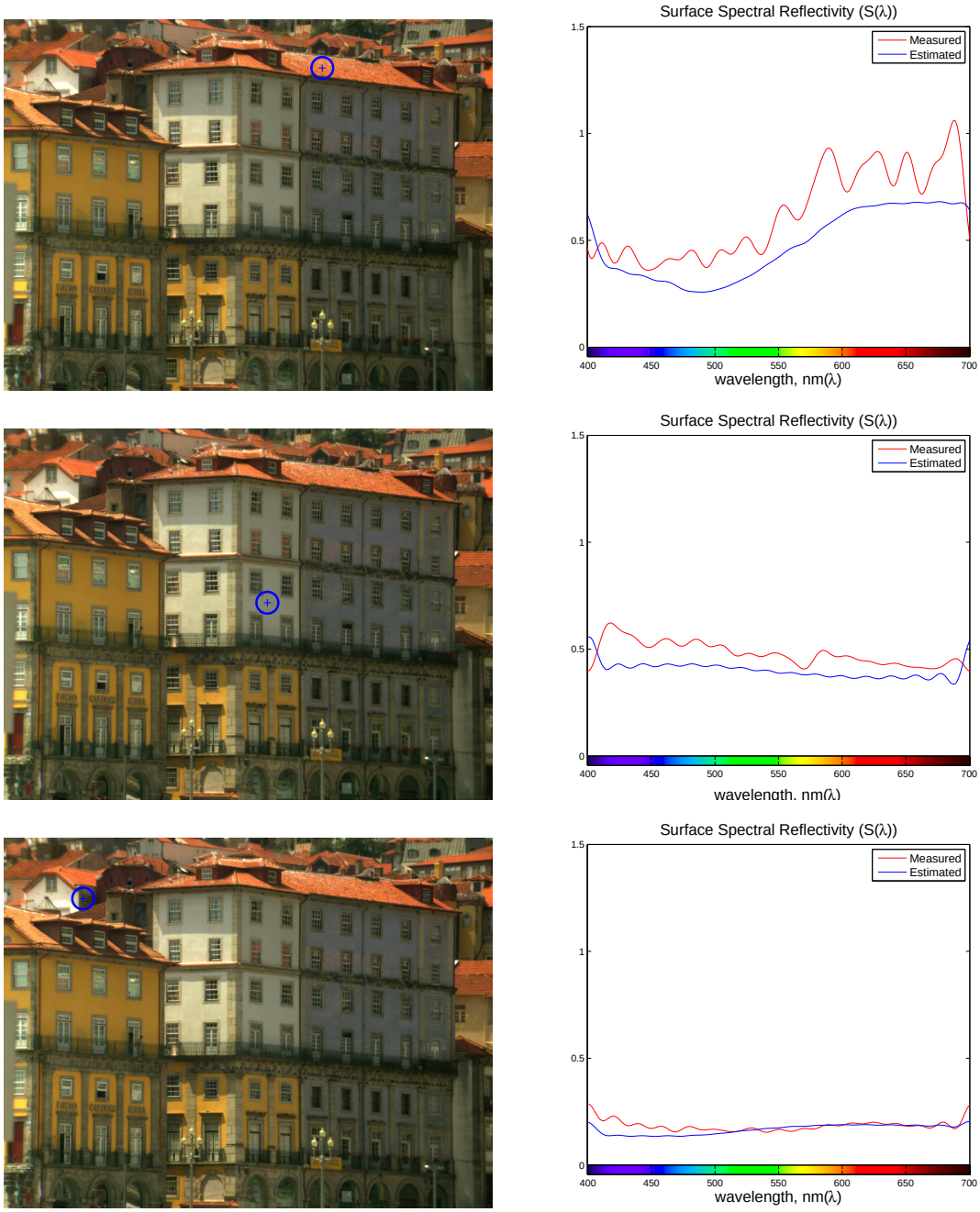


Figure A.3: Samples of measured and estimated surface spectral reflectance for Foster data set 3

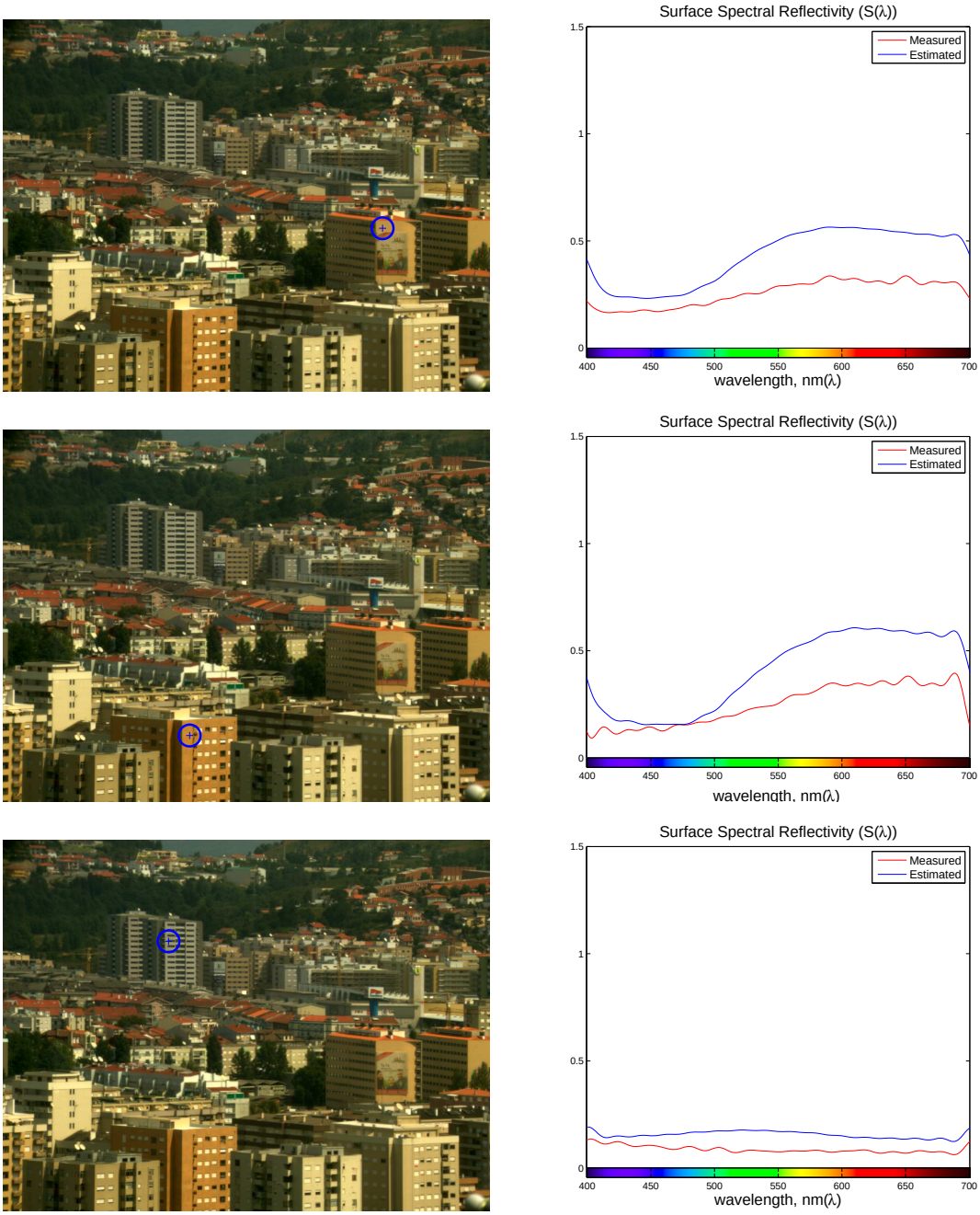


Figure A.4: Samples of measured and estimated surface spectral reflectance for Foster data set 4

B

Highlights Segmentation and CCT Estimation

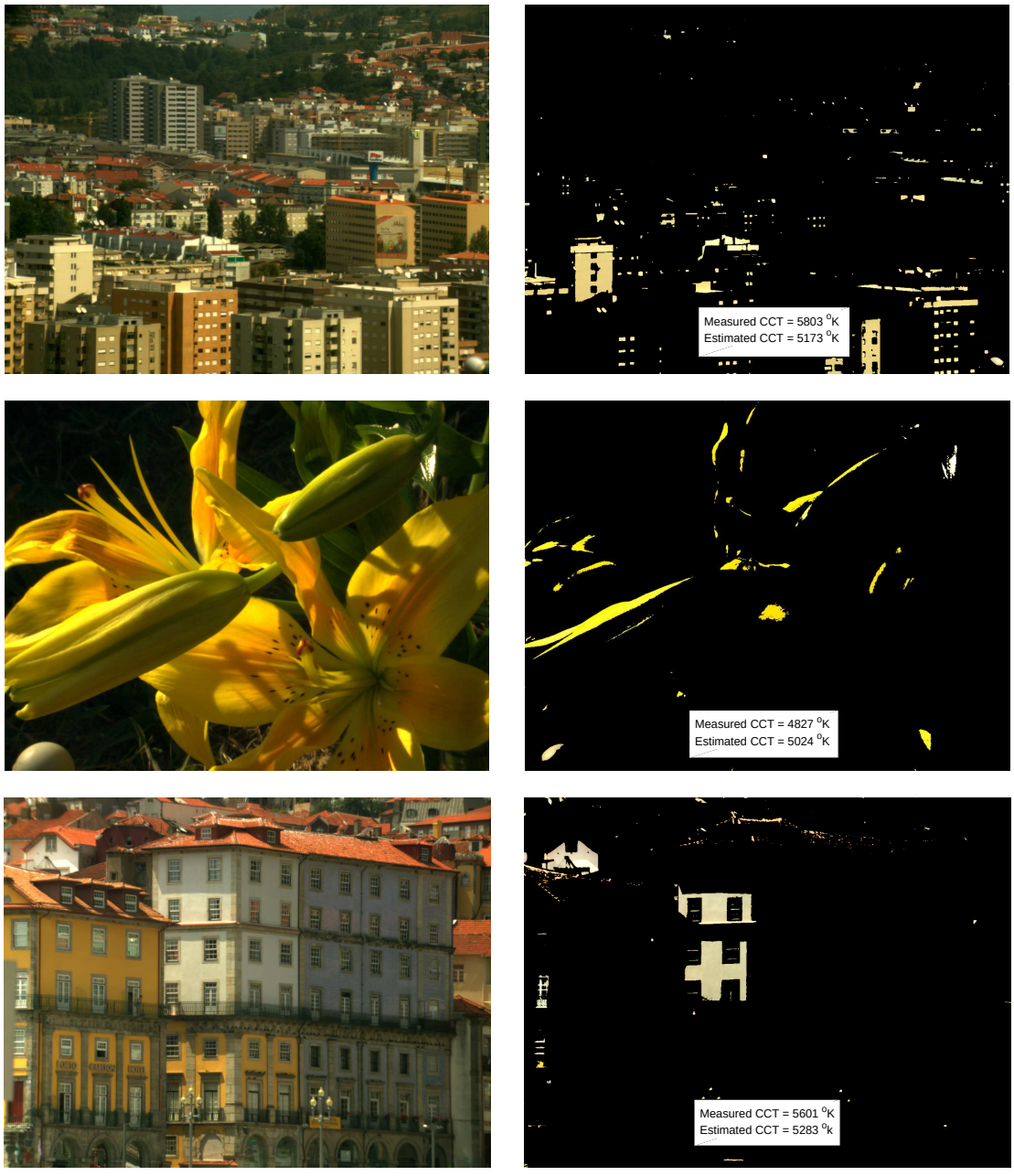


Figure B.1: Samples of segmented highlights from Foster data set

C

Data Sets

(a) frame # 1

(b) frame # 90

(c) frame # 150

(d) frame # 240

(e) frame # 300

Figure C.1: Samples of 'Intelligent Room' data set

(a) frame # 1)

(b) frame # 40

(c) frame # 100

(d) frame # 170

(e) frame # 240

Figure C.2: Samples of 'Hall Monitor' data set

(a) frame # 1)

(b) frame # 70

(c) frame # 140

(d) frame # 210

(e) frame # 260

Figure C.3: Samples of 'Lab' data set

D

List of Publications

1 M. Sedky, M. Moniri, and C. C. Chibelushi, "Classification of Smart Video Surveillance Systems for Commercial Applications," in IEEE International Conference on Advanced Video and Signal Based Surveillance, Como, Italy, pp. 638-643, 2005.

2 M. Sedky, M. Moniri, and C. Chibelushi, "Smart Video Surveillance for Workplace Applications: Implications, Technologies and Requirements," in The 5th IASTED International Conference on Visualisation, Imaging, and Image Processing, Benidorm, Spain, pp. 737-742, 2005.

Bibliography

[1] C. Regazzoni, V. Ramesh, and G. Foresti, "Scanning the issue/technology," In: Proceedings of IEEE Special Issue on Video Communications, Processing and Understanding For Third Generation Surveillance Systems, vol. 89, no. 7, pp. 1355-1366, 2001.

[2] "CCTV industrial report-UK- April 2008," Market & Business Development, Mintel, April, 2008.

[3] M. Sedky, M. Moniri, and C. C. Chibelushi, "Classification of smart video surveillance systems for commercial applications," In: IEEE International Conference on Advanced Video and Signal Based Surveillance, Como, Italy, pp. 638-643, 2005.

[4] `http://www.bsia.co.uk/LY8VMY74118`," (22^{nd} of April 2009, date last accessed)

[5] Hannah M. Dee, and Sergio A. Velastin, "How close are we to solving the problem of automated visual surveillance?," Journal of Machine Vision and Applications, Springer Berlin / Heidelberg, May, 2007.

[6] E. Wallace and C. Diffley, "CCTV control room ergonomics," Technical Report 14/98, Police Scientific Development Branch (PSDB), UK Home Office, 1998.

[7] A. Dick and M. Brooks, "Issues in automated visual surveillance," In: International Conference on Digital Image Computing Techniques and Applications, Como, Italy, pp. 195-204, 2003.

[8] R. T. Collins, A. J. Lipton, and T. Kanade, "A system for video surveillance and monitoring," In: Proceedings of American Nuclear Society (ANS), 8^th International Topical Meeting Robotic and Remote Systems, Technical Report, CMU-RI-TR-00-12, Carnegie Mellon University, 2000.

[9] A. J. Lipton, C. H. Heartwell, N. Haering, and D. Madden, "Automated video protection, monitoring and detection," In: IEEE Aerospace and Electronic Systems Magazine, vol. 18, no. 5, pp. 3-18, 2003.

[10] D. M. Gavrila, "The visual analysis of human movement: a survey," In: Proceedings of Computer Vision and Image Understanding, vol. 73, no. 1, pp. 82-98, 1999.

[11] I. Haritaoglu, D. Harwood, and L. Davis, "W4: real-time surveillance of people and their activities," In: IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 809-830, 2000.

[12] M. Sonka, R. Boyle, and V. Hlavac, Image Processing, Analysis and Machine Vision: PWS Publishing, NewYork, 1998.

[13] A. Hampapur, L. Brown, J. Connel, S. Pankanti, A. Senior, and Y. Tian, "Smart surveillance: applications, technologies and implications," In: IEEE Pacific-Rim Conference on Multimedia, 2003.

[14] A. J. Lipton, J. I. Clark, P. Brewe, P. L. Venetianer, and A. J. Chosak,"ObjectVideo Forensics: activity-based video indexing and retrieval for physical security applications," In: IEE Workshop on Intelligent Distributed Surveillance Systems (IDSS-04), pp. 56-60, 2004.

[15] C. A. Rahman, W. Badawy, and A. Radmanesh, "A real time vehicles license plate recognition system," In: IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 163-166, 2003.

[16] T. Alexandropoulos, S. Boutas, V. Loumos, and E. Kayafas, "Real-time change detection for surveillance in public transportation," In: IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 58-63, 2005.

[17] J. R. Radke, A. Srinivas, A. K. Omar, and R. Badrinath, "Image change detection algorithms: a systematic survey," In: IEEE Transactions on Image Processing, vol. 14, no. 3, pp. 294-307, 2005.

[18] E. Selvador, "Shadow segmentation and tracking in real-world conditions," PhD. Thesis, Trieste University, 2004.

[19] J. Ho, B. V. Funt, and M. S. Drew, "Separating a colour signal into illumination and surface reflectance components: Theory and applications," In: IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 12, pp. 966-977, 1990.

[20] G. J. Klinker, S. A. Shafer, and T. Kanade,"A physical approach to colour image understanding," In: International Journal on Computer Vision, vol. 4, no. 1, pp. 7-38, 1990.

[21] V. Cardei, B. Funt, and K. Barnard, "Estimating the scene illumination chromaticity using a neural network," In: Journal of the Optical Society of America A, vol. 19, no. 12, pp. 2363-2373, 2002.

[22] G. Sapiro,"Colour and illuminant voting," In: IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 21, no. 11, pp. 1210-1215, 1999.

[23] R. T. Collins, A. J. Lipton, and T. Kanade, "Introduction to the special section on video surveillance," In: IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 22, pp. 745-746, 2000.

[24] M. Saptharishi, C. S. Oliver, C. P. Diehl, K. S. Bhat, J. M. Dolan, A. Trebi-Ollennu, and P. K. Khosla, "Distributed surveillance and reconnaissance using multiple autonomous ATVs: CyberScout," In: IEEE Transactions on Robotics and Automation, vol. 18, no. 5, pp. 826-836, 2002.

[25] R. Cutler and L. S. Davis, "Robust real-time periodic motion detection, analysis, and applications," In: IEEE Transaction on Pattern Analysis Machine Intelligence, vol. 22, no. 8, pp. 781-796, 2000.

[26] G. L. Foresti, "Object recognition and tracking for remote video surveillance," In: IEEE Transactions on Circuits and Systems for Video Technology, vol. 9, no. 7, pp. 1045-1062, 1999.

[27] A. Birk and H. Kenn, "An industrial application of behaviour-oriented robotics," In: IEEE Conference on Robotics and Automation (ICRA), vol. 1, pp. 749-754, 2001.

[28] R. Cucchiara, C. Grana, A. Prati, G. Tardini, and R. Vezzan, "Using computer vision techniques for dangerous situation detection in domestic applications," In: IEE Workshop on Intelligent Distributed Surveillance Systems (IDSS-04), pp. 1-5, 2004.

[29] J. Kang, I. Cohen, and G. Medioni, "Tracking objects from multiple stationary and moving cameras," In: IEE Workshop on Intelligent Distributed Surveillance Systems (IDSS-04), pp. 31-35, 2004.

[30] F. Helten and B. Fisher, "Video surveillance on demand for various purposes?" In: the 5^{th} Framework Programme of the European Commission, In: Berlin Institute for Social Research: Berlin Institute for Social Research, Working paper no.11, 2003.

[31] B. Godfrey, "Electronic work monitoring: an ethical model," In: the $2^n d$ Australian Institute of Computer Ethics Conference. Canberra, Australia: Australian computer Society Inc., pp. 18-21, 2000.

[32] R. Tribbey, "Workplace privacy: audio and video surveillance," In: Public Administration, Master: University of Louisville, 1999.

[33] "A Guide to the Workplace Video Surveillance Act 1998 (NSW)," Privacy NSW, 2002.

[34] M. Sedky, M. Moniri, and C. C. Chibelushi, "Smart video surveillance for workplace applications: implications, technologies and requirements," In: The 5th IASTED International Conference on Visualisation, Imaging, and Image Processing, Benidorm, Spain, pp. 737-742, 2005.

[35] J. Rule and P. Brantly, "Surveillance in the workplace: a new meaning to 'personal' computing," In: International conference on Shaping Organisations and Shaping Technology, 1991.

[36] M. Vorvoreanu and A. H. Botan, "Examining electronic surveillance in the workplace: a review of theoretical perspectives and research findings," In: Conference of the International Communications Association, IN 47907, pp. 2000-14, 2000.

[37] C. Diehl, "Toward efficient collaborative classification for distributed video surveillance," PhD. Thesis, Carnegie Mellon University, 2000.

[38] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviours," In: IEEE Transactions on Systems MAN and Cybernetics, Part C: Applications and Reviews, vol. 34, no. 3, pp. 334-352, 2004.

[39] P. E. Walters, "CCTV systems thinking-systems practice," In: IEE European Convention on Security and Detection, pp. 64-69, 1995.

[40] M. Xu, L. Lowey, and J. Orwell, "Architecture and algorithms for tracking football players with multiple cameras," In: IEE Workshop on Intelligent Distributed Surveillance Systems (IDSS-04), pp. 51-55, 2004.

[41] A. Cripps, "Workplace surveillance," N. S. W. C. f. C. Liberties, Ed.: NSW Council for Civil Liberties, 2004.

[42] A. Crossman and L. Lee-Kelley, "High-tech surveillance in the workplace: the psychological contract revisited," In: the International HRM Conference, University of Ljubljana, 2004.

[43] J. Hogan, "Your every move will be analysed," New Scientist, vol. 179(2403), no. 4-5, 2003.

[44] S. Shan, Q. Liu, D. Tao, D. Xu, S. Yan, and X. Li, "Introduction to the special issue on video-based object and event analysis," In: Pattern Recognition Letters, vol. 30, no. 2, pp. 87, 2009.

[45] D. Rey, G. Subsol, H. Delingette, and N. Ayache, "Automatic detection and segmentation of evolving processes in 3D medical images: application to multiple sclerosis," In: Medical Image Analysis, vol. 6, no. 2, pp. 163-179, 2003.

[46] D. Corrall, "VIEWS: computer vision for surveillance applications," In: IEE Coll. on Active and Passive Techniques for 3D Vision, vol. 8, pp. 1-3, 1991.

[47] D. Farin, N. Mache, and P. H. N. With, "A software-based high-quality MPEG-2 encoder employing scene change detection and adaptive quantization," In: IEEE Transactions on Consumer Electronics, vol. 48, no. 1, pp. 887-897, 2002.

[48] A. Cavallaro and T. Ebrahimi, "Accurate video object segmentation through change detection," In: IEEE International Conference on Multimedia and Expo, ICME '02. Proceedings, vol. 1, pp. 445-448, 2002.

[49] A. Mittal, and N. Paragios, "Motion-based background subtraction using adaptive kernel density estimation," In: IEEE Computer Society Conference on Computer Vision and Pattern recognition, CPVR'04, vol. 2, no. 1, pp. 302-309, 2004.

[50] S. Elhabian, K. El-Sayed and S. Ahmed, "Moving object detection in spatial domain using background removal techniques - state-of-art," In: Recent Patents on Computer Science, vol. 1, no. 1, pp. 32-54, 2008.

[51] M. Piccardi, "Background subtraction techniques: a review," In: IEEE International Conference on Systems, Man and Cybernetics, vol. 28, no. 4, pp. 3099- 3104, 2004.

[52] P. J. Withagen, "Object detection and segmentation for visual surveillance," PhD thesis, University of Amsterdam, 2005.

[53] S. G. Narasimhan, V. Ramesh, and S. K. Nayar, "A class of photometric invariants: Separating material from shape and illumination," In: IEEE International Conference on Computer Vision, vol. 2, pp. 1387-1394, 2003.

[54] M. Mason and Z. Duric, "Using histograms to detect and track objects in colour video," In: Proceedings of IEEE Workshop on Applied Imagery Pattern Recognition, pp. 154-159, 2001.

[55] Y. Hwang and J-S Kim and I-S Kweon, "Change detection using a statistical model in an optimally selected colour space," In: Journal of Computer Vision and Image Understanding, vol. 112, no. 3, pp. 231-242, 2008.

[56] A. J. Lipton, H. Fujiyoshi, and R. S. Patil, "Moving target classification and tracking from real-time video," In: IEEE Workshop on Applications of Computer Vision, pp. 8-14, 1998.

[57] A. M. McIvor, "Background subtraction techniques," In: International Conference on Image and Vision Computing, pp. 3099-3104, 2000.

[58] A. R. Francois and G. C. Medioni, "Adaptive colour background modelling for real-time segmentation of video streams," In: Proceedings of the International Conference on Image Science, Systems, and Technology, pp. 227-232, 1999.

[59] C. OConaire, E. Cooke, N. O'Connor, N. Murphy, and A. Smeaton, "Background modelling in infrared and visible spectrum video for people tracking," In: Proceedings of IEEE International Workshop on Object Tracking and Classification in and beyond Visible Spectrum, vol. 1, no. 1, pp. 20-25, 2005.

[60] A. Colombari, A. Fusiello, and V. Murino, "Video Objects Segmentation by Robust Background Modelling," In: ICIAP 14^th International Conference on Image Analysis and Processing, pp. 155-164, 2007.

[61] A. Monnet, A. Mittal, N. Paragios, and V. Ramesh, "Background modelling and subtraction of dynamic scenes," In: IEEE International Conference on Computer Vision (ICCV), pp. 1305-1312, 2003.

[62] L. Li, W. Huang, I. Y. Gu and Q. Tian, "Statistical modelling of complex backgrounds for foreground object detection," In: IEEE Transactions on : Image Processing, vol. 13, no. 11, pp. 1459-1472, 2004.

[63] D-M Tsai and S-C Lai, "Independent component analysis-based background subtraction for indoor surveillance," In: IEEE Transactions on Image Processing, vol. 18, no. 1, pp. 158-167, 2009.

[64] T. Parag, A. Elgammal, and A. Mittal, "A framework for feature selection for background subtraction," In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 1916-923, 2006.

[65] M. Heikkila and M. Pietikinen, "A texture-based method for modelling the background and detecting moving objects," In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 4, pp. 657-662, 2006.

[66] O. Javed, K. Shafique, and M. Shah, "A hierarchical approach to robust background subtraction using colour and gradient information," In: Proceedings of IEEE Workshop on Motion Video Computing, pp. 22-27, 2002.

[67] T. E. Boult, R. J. Micheals, X. Gao, and M. Eckmann, "Into the woods: visual surveillance of non-cooperative and camouflaged targets In: complex outdoor settings," In: the IEEE Proceedings, vol. 89, no. 1, pp. 1382 -1402, 2001.

[68] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: principles and practice of background maintenance," In: the 7^{th} IEEE International Conference on Computer Vision, pp. 255-261, 1999.

[69] W. E. Grimson, C. Stauffer, R. Romano, and L. Lee, "Using adaptive tracking to classify and monitor activities in a site," In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR98, pp. 22-29, 1998.

[70] D. Wang, T. Feng, H. Shum, and S. Ma, "A novel probability model for background maintenance and subtraction," In: Proceedings of The 15^th International Conference on Vision Interface, vol. 1, no. 1, pp. 109-117, 2002.

[71] S. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking groups of people," In: IEEE Proceedings of Computer Vision and Image Understandings, vol. 80, no. 1, pp. 42-56, 2000.

[72] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modelling using nonparametric kernel density estimation for visual surveillance," Proc. IEEE, vol. 90, no. 7, pp. 1151-1163, 2002.

[73] T. Horprasert, D. Harwood, and L. S. Davis, "A statistical approach for real-time robust background subtraction and shadow detection," In: IEEE (ICCV-99) Frame-Rate Workshop, pp. 1-19, 1999.

[74] H. W. S. Jabri, Z. Duric and A. Rosenfeld., "Detection and location of people in video images using adaptive fusion of colour and edge information," In: Proceedings of the 15^th International Conference on Pattern Recognition, vol. 4, pp. 627-630, 2000.

[75] D. Hong and W. Woo, "A background subtraction for a vision-based user interface," In: IEEE Joint Conference Proceedings of the 4^th International Conference on Information, Communications and Signal, ICICS-PCM, vol. 1, pp. 263-267, 2003.

[76] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland,"Pfinder: Real-time tracking of the human body," In: IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 780-785, 1997.

[77] A.Cavallaro, E. Salvador and T. Ebrahimi, "Shadow-aware object-based video processing," In: Proceedings of IEE, Vision, Image and Signal Processing, vol. 152, no. 4, pp. 398-406, 2005.

[78] P. Blauensteiner, H. Wildenauer, A. Hanbury and M. Kampel, "Motion and shadow detection with an improved colour model," In: Proceedings of the IEEE International Conference on Signal and Image Processing, vol. 152, no. 4, pp. 627-632, 2006.

[79] Y. Weiss, "Deriving intrinsic images from image sequences," In: Proceedings of IEEE International Conference on Computer Vision, pp. 68-75, 2001.

[80] S.Nadimi and B. Bhani, "Physical models for moving shadow and object

detection in video," In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 8, pp. 1079-1087, 2004.

[81] M. Greiffenhagen, V. Ramesh, D. Comaniciu and H. Niemann, "Statistical modelling and performance characterization of a real-time dual camera surveillance system," In: Proceedings of International Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 335-342, 2000.

[82] C. Harris and M. Stephens, "A combined corner and edge detector," In: the 4^{th} ALVEY Vision Conference, pp. 147-151, 1998.

[83] D. Koller, J. Weber, T. Huang, J. Malik, G. Ogasawara, B. Rao and S. Russell, "Toward robust automatic traffic scene analysis in realtime," In: Proceedings of International Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 126-131, 1994.

[84] L. Li and M. Leung, "Integrating intensity and texture differences for robust change detection," In: IEEE Transactions on Image Processing, vol. 11, pp. 105-112, 2002.

[85] J. S. Zelek, "Towards Bayesian real-time optical flow," In: Image and Vision Computing, vol. 22, no. 1, pp. 1051-1069, 2004.

[86] L. Wixson, "Detecting salient motion by accumulating directionary-consistent flow," In: Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pp. 774-780, 2000.

[87] A. Senior, A. Hampapur, Y.-L. Tian, L. Brown, S. Pankanti, and R. Bolle, "Appearance models for occlusion handling," In: the $2^n d$ IEEE International Workshop on PETS, Kauai, Hawaii, USA, 2001.

[88] W. Zhiqiang, J. Xiaopeng, and W. Pengand, "Real-time moving object detection for video monitoring systems," In: Journal of Systems Engineering and Electronics, vol. 17, no. 4, pp. 731-736, 2006.

[89] T. Zickler, S. P. Mallick, D. J. Kriegman, and P. Belhumeur, "Colour subspaces as photometric invariants," In: Proceedings of the IEEE Computer Society

Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2000-2010, 2006.

[90] T. Zickler, S. P. Mallick, D. J. Kriegman, and P. Belhumeur, "Colour subspaces as photometric invariants," In: Proceedings of the International Journal of Computer Vision, vol. 79, no. 1, pp. 13-30, 2008.

[91] Y. Raja, S. J. MacKenn and S. Gong, "Segmentation and tracking using colour mixture models," In: Proceedings of Asian Conference on Computer Vision, 1998.

[92] Y. Hwang and J-S Kim and I-S Kweon, "Change detection using a statistical model of the noise in colour images," In: Proceeding of IROS'04, vol. 3, pp. 2713-2718, 2004.

[93] E. Durucan and T. EbrahimiKim, "Robust and illumination invariant change detection based on linear dependence for surveillance applications," In: Proceedings of X European Conference on Signal Processing, pp. 1041-1044,2000.

[94] E. Durucan, Y. Weilenmann and J. Snoeckx, "Illumination invariant background extraction," In: Proceedings of IEEE International Conference on Image Analysis and Processing, pp. 1136-1139, 1999.

[95] F. Porikli, "Multiplicative background-foreground estimation under uncontrolled illumination using intrinsic images," In: IEEE Workshop on Motion and Video Computing, vol. 2, pp. 20-27, 2005.

[96] Y. Matsushita, K. Nishino, K. Ikeuchi and S. Masao, "Illumination normalization with time-dependent intrinsic images for video surveillance," In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 68-75, 2004.

[97] L. M. Fuentes and S. A. Velastin, "People tracking in surveillance applications," In: Image and Vision Computing, pp. 1165-1171, 2006.

[98] S. Nayar and R. Bolle, "Computing reflectance ratios from an image," In: Pattern Recognition, vol. 26, no. 10, pp. 1529-1542, 1993.

[99] F. Bunyak, I. Ersoy and S. R. Subramanya, "Shadow detection by combined photometric invariants for improved foreground segmentation," In: IEEE Workshops on Application of Computer Vision, WACV/MOTIONS '05, vol. 1, pp. 510-515, 2005.

[100] K. Skifstad and R. Jain, "Illumination independent change detection from real world image sequences," In: Computer Vision, Graphics, Image Procecessing, vol. 46, pp. 387-399, 1989.

[101] G. D. Finlayson and S. D. Hordley, "Colour invariance at a pixel," In: Proceedings of Asian Conference on Computer Vision, pp. 13-22, 2000.

[102] S. A. Shafer, "Using colour to separate reflection components," In: Colour Research, pp. 210-218, 1985.

[103] T. Zickler, S. P. Mallick, D. J. Kriegman, and P. Belhumeur, "Determining shape and reflectance of hybrid surfaces by photometric sampling," In: Proceedings of the IEEE Journal of Robotics and Automation, vol. 6, no. 4, pp. 418-431, 1990.

[104] L. T. Maloney and B. A. Wandell, "colour constancy: a method for recovering surface spectral reflectance," In: Journal of the Optical Society of America A, vol. 3, no. 1, pp. 29-33, 1986.

[105] R. Bajcsy, S. Lee, and A. Leonardis, "Colour image segmentation with detection of highlights and local illumination induced by inter-reflections," In: Proceedings of ICPR, pp. 785-790, 1990.

[106] L. T. Maloney, "Evaluation of linear models of surface spectral reflectance with small numbers of parameters," In: Journal of the Optical Society of America A, vol. 3, pp. 1673-1683, 1986.

[107] D. H. Marimont and B. A. Wandell, "Linear models of surface and illuminant spectra," In: Journal of the Optical Society of America A, vol. 3, pp. 1673-1683, 1992.

[108] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," In: IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 25, no. 9, pp. 1063-1074, 2003.

[109] N. Tsumura, N. Ojima, K. Sato, M. Shiraishi, H. Shimizu, H. Nabeshima, S. Akazaki, K. Hori and Y. Miyake., "Imagebased skin colour and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin," In: ACM Transaction on Graphics, vol. 22, no. 3, pp. 770-779, 2003.

[110] X. Dai and S. Khorram, "The effects of image misregistration on the accuracy of remotely sensed change detection," In: IEEE Transactions on Geoscience Remote Sensing, vol. 36, no. 5, pp. 1566-1577, 1998.

[111] Y. Gong and M. Sakauchi, "Detection of regions matching specified chromatic features," In: Computer Vision and Image Understanding, vol. 61, no. 2, pp. 263-269, 1995.

[112] X. Lu and H. Zhang, "Colour classification using adaptive dichromatic model," In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '06), vol. 152, no. 4, pp. 3411-3416, 2006.

[113] P. Kumar, K. Sengupta, A. Lee, and S. Ranganath, "A comparative study of different colour spaces for foreground and shadow detection for traffic monitoring system," In: Proceeding of the 5^th IEEE International Conference on Intelligent Transportation Systems, pp. 100-105, 2002.

[114] Y. Hwang and J-S Kim and I-S Kweon, "Determination of colour space for accurate change detection," In: IEEE International Conference on Image Processing, pp. 3021-3024, 2006.

[115] P. Kumar, S. Ranganath, H. Weimin and K. Sengupta, "Framework for real-time behavior interpretation from traffic video," In: IEEE Transactions on Intelligent Transportation Systems, vol. 6, no. 1, pp. 43-53, 2005.

[116] S. Tominaga and B. A. Wandell, "Natural scene-illuminant estimation using the sensor correlation," In: Proceedings of the IEEE, vol. 90, no. 1, pp. 42-56, 2002.

[117] B. A. Maxwell and S. A. Shafer, "A framework for segmentation using physical models of image formation," In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 361-368, 1994.

[118] Z. Liu, K. Huang, T. Tan, and L. Wang, "Cast shadow removal with GMM for surface reflectance component," In: IEEE International Conference on Pattern Recognition, ICPR06, vol. 1, pp. 727-730, 2006.

[119] B. T. Phong, "Illumination for computer generated pictures," In: Communications, ACM 18, vol. 18, pp. 311-317, 1975.

[120] K. E. Torrance and E. Sparrow, "Theory for off-specular reflection from roughened surfaces," In: Journal of the Optical Society of America A, vol. 57, no. 9, pp. 1105-1114, 1967.

[121] S. Nayar and R. Bolle, "Reflectance based object recognition," In: International Journal of Computer Vision, vol. 17, no. 3, pp. 219-240, 1999.

[122] M. Tappen, W. Freeman and E. Adelson, "Recovering Shading and Reflectance from a single image," In: IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 27, no. 9, pp. 1459-1472, 2002.

[123] G.D. Finlayson and S.D. Hordley, "Colour constancy at a pixel," In: Journal of the Optical Society of America A, vol. 18, pp. 253-264, 2001.

[124] S. Tominaga and B. A. Wandell, "Standard surface-reflectance model physical models of image formation," In: Journal of the Optical Society of America, vol. 9, no. 4, pp. 576-584, 1989.

[125] J. Parkkinen, J. Hallikainen and T. Jaaskelainen, "Characteristic spectra of Munsell colours," In: Journal of the Optical Society of America A, vol. 6, no. 2, pp. 318-322, 1989.

[126] D. Judd, D. MacAdam, and G. Wyszecki, "Spectral distribution of typical daylight as a function of correlated colour temperature," In: Journal of the Optical Society of America A, vol. 54, no. 8, pp. 1031-1040, 1964.

[127] J. P. Renno, D. Markis, T. Ellis and G. A. Jones, "Application and evaluation of colour constancy in visual surveillance," In: International Workshop on Performance Evaluation of Tracking and Surveillance, pp. 301-308, 2005.

[128] S. Negahdaripour, "Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis," In: IEEE Transaction on Pattern Analysis and Machine Inteligence, vol. 20, no. 9, pp. 961-979, 1998.

[129] O. Pizarro and H. Singh, "Toward large-area mosaicing for underwater scientific applications," In: IEEE Journal of Ocean Engineering, vol. 28, no. 4, pp. 651-672, 2003.

[130] C. M.Onyango and J. A. Marchant, "Physics-based colour image segmentation for scenes containing vegetation and soil," In: Journal Image and Vision Computing, vol. 19, no. 8, pp. 523-538, 2001.

[131] Y-L Tian and M. Lu and A. Hampapur, "Robust and efficient foreground analysis for real-time video surveillance," In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, no. 1, pp. 1182- 1187, 2005.

[132] C. Stauffer and W. E. Grimson "Adaptive background mixture models for real-time tracking," In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR99, pp. 22-29, 1999.

[133] J. Shen, "Motion detection in colour image sequence and shadow elimination,"

In: Proceedings of SPIE on Visual Communications and Image Processing, pp. 731-740, 2004.

[134] D. Butler, S. Sridharan, V. Bove, and D. Harwood, "Real-time adaptive background segmentation," In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, ICASSP'03, vol. 1, no. 3, pp. 349-352, 2003.

[135] K. Kim, T. H. Chalidabhongse, L. S. Davis, and D. Harwood, "Real-time foreground-background segmentation using codebook model," In: Proceedings of The International Conference on Real-Time Imaging, vol. 11, no. 3, pp. 172-185, 2005.

[136] N. M. Oliver, B. Rosario, A. P. Pentland, "A bayesian computer vision system for modelling human interactions," In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pp. 831-843, 2000.

[137] P. Rosin, and E. Ioannidis, "Evaluation of global image thresholding for change detection," In: Pattern Recognition Letters, vol. 24, no. 14, pp. 2345-2356, 2003.

[138] P. Smits and A. Annomi, "Toward specification-driven change detection, " In: IEEE Transaction on Geosciences Remote Sensing, vol. 38, no. 3, pp. 1484-1488, 2000.

[139] F. Oberti, A. Teschioni and C.S. Regazzoni, "ROC curves for performance evaluation of video sequences processing systems for surveillance applications," In: Proceedings of International Conference on Image Processing, vol. 2, pp. 949-953, 1999.

[140] F. Oberto, F. Granelli and C.S. Regazzoni, "Minimax based regulation of change detection threshold in videosurveillance systems," In: Multimedia video-based surveillance systems: requirements, issues and solutions, G.L. Foresti, P. Mhnen, C.S. Regazzoni, pp 210-233, Kluwer Academic Publishers, 2000.

[141] A. Prati, I. Mikic, M.M. Trivedi, and R. Cucchiara, "Detecting moving

shadows: algorithms and evaluation," In: IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI, vol. 25, pp. 918-923, 2003.

[142] H. Junwei, H. Wenfong, and C. Chia-Jung, "Shadow elimination for effective moving object detection by Gaussian shadow modelling," Image and Vision Computing, vol. 2, no. 1, pp. 505-516, 2003.

[143] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Cast shadow segmentation using invariant colour features," In: Pattern Recognition Letters, vol. 26, no. 1, pp. 251-265, 2005.

[144] R. Cucchiara, C. Grana, M. Piccardi and A. Prati, "Detecting objects, shadows and ghosts in video streams by exploiting colour and motion information," In: Proceedings of IEEE International Conference on Image Analysis and Processing, pp. 360-365, 2001.

[145] J. Stauder, R. Mech and J. Ostermann, "Detection of moving cast shadows for object segmentation," In: IEEE Transaction on Multimedia, vol. 1, no. 1, pp. 65-76, 1999.

[146] J. A. Marchant and C. M. Onyango, "Shadow invariant classification for scenes illuminated by daylight," In: Journal of the Optical Society of America A, vol. 17, no. 11, pp. 1952-1961, 2000.

[147] C. Ridder, O. Munkelt and H. Kirchner, "Adaptive background estimation and foreground detection using Kalman-filtering, " In: International Conference on Recent Advances in Mechatronics, pp. 193-199, 1995.

[148] G. Welch and G. Bishop, "An introduction to the Kalman filter," Department of Computer Science, University of North Carolina at Chapel Hill, Technical Report TR 95-041, 1995.

[149] F. Khal, R. Hartley, V. Hilsenstein, "Novelty detection in image sequences with dynamic background," In: Proceedings of European Conference on Computer Vision, $2^n d$ Workshop on Statistical Methods in Video Processing (SMVP), 2005.

[150] P. KaewTraKulPong and R. Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detection," In: Proceedings European Workshop Advanced Video Based Surveillance Systems, $2^n d$ AVSS, vol. 1, pp. 149-158, 2001.

[151] J. Nascimento and J.Marques, "Performance evaluation of object detection algorithms for video surveillance," In: IEEE Transaction on Multimedia, vol. 8, no.4, pp. 761-774, 2006.

[152] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," In: Proceedings of International Conference on Pattern Recognition, pp. 28-31, 2004.

[153] Y. Zhang, Advances in Image and Video Segmentation, IRM Press, Hershey, PA. , 2006.

[154] C.E. Erdem, B. Sankur and A.M. Tekalp, "Performance measures for video object segmentation and tracking," In: IEEE Transactions on Image Processing, vol. 13, no. 7, pp. 937-951, 2004.

[155] X. Desurmont, R. Wijnhoven and E. Jaspers, "Performance evaluation of realtime video content analysis systems in the CANDELA project," In: Proceedings of SPIE, vol. 5671, pp. 200-211, 2005.

[156] P.L. Correia and F. Pereira, "Objective evaluation of video segmentation quality," In: IEEE Transactions on Image Processing, vol. 12, no. 2, pp. 186-200, 2003.

[157] T.H. Chalidabhongse, K. Kim, D. Harwood and L. Davis, "A perturbation method for evaluating background subtraction algorithms," In: Proceedings of Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, VS-PETS, 2003

[158] A. Cavallaro, "From visual information to knowldge: semantic video object segmentation, tracking and description," PhD. Thesis, Trieste University, 2002.

[159] McCamy and S. Calvin, "Correlated colour temperature as an explicit function of chromaticity coordinates," In: Journal of Colour Research & Application, vol. 17, no. 2, pp. 142-144, 1992.

[160] C. A. Poynton, "Frequently asked questions about gamma, `http://www.poynton.com/PDFs/GammaFAQ.pdf`," 1998, (8^{th} of June 2008, date last accessed)

[161] "`www.unibrain.com/download/pdfs/Fire-i_Board_Cams/ICX098BQ.pdf`" (8^{th} of June 2008, date last accessed)

[162] G. E. Healey and R. Kondepudy, "Radiometric CCD camera calibration and noise estimation," In: IEEE Transaction on Image Processing, vol. 16, no. 3, pp. 267-276, 1994.

[163] P. J. Withagen, F. C. A. Groen and K. Schutte, "Radiometric CCD camera calibration and noise estimation," In: IEEE Transaction on Instrumentation and Measurement, vol. 56, no. 1, pp. 199-203, 2007.

[164] S. Lin and S. Heung-Yeung, "Separation of diffuse and specular reflection in color images," In: The Proceedings of the IEEE Computer Society Conference on IComputer Vision and Pattern Recognition, vol. 1, pp. 341-346, 2001.

[165] R. T. Tan, "Illumination colour and intrinsic surface properties physics-based colour analysis from a single image," PhD. Thesis, University of Tokyo, 2002.

[166] S. Westland and C. Ripamonti, Computational colour science using Matlab: Wiley, 2004.

[167] D. H. Foster, "Does colour constancy exist?," Trends In: Cognitive Science, vol. 7, pp. 439-443, 2003.

[168] I. Ashdown, "Chromaticity and colour temperature for architectural lighting," In: in SPIE Proceding on Solid State Lighting II, vol. 4776,, pp. 51-60, 2002.

[169] C. A. Poynton, "Frequently asked questions about colour, `http://www.poynton.com/PDFs/ColourFAQ.pdf`," 1995. (8^{th} of June 2008, date last accessed)

[170] N. Lazarevic-McManus, J.R. Renno, and G.A. Jones, "Performance evaluation in visual surveillance using the F-measure", ACM Multimedia Workshop on Video Surveillance and Sensor Networks, October 27, Santa Barbara, CA, USA, pp. 41-55, 2006.

[171] G. L. Foresti, P. Mahonen, and C. Regazzoni, Multimedia video-Based surveillance systems: requirements, issues and solutions: Kluwer Academic Publishers, USA, 2000.

[172] "`http://its.ee.tsinghua.edu.cn/literatures/_Literature_incoming/`" (8^{th} of June 2008, date last accessed)

[173] "Europian project IST 10942 art.live.," `www.tele.ucl.ac.be/PROJECTS/art.live/`. (8^{th} of June 2008, date last accessed)

[174] M. Karaman, L. Goldman, and T. S. D. Yu, "Comparison of static background segmentation methods," In: Proceedings of SPIE, vol. 5960, pp. 2140-2151, 2005.

[175] S. A. Velastin, and P. Remagnino, Intelligent Distributed Video Surveillance Systems, IET Publishing, Institution of Electrical Engineers, ISBN 0863415040-9780863415043, 2006. W. Zhiqiang, J. Xiaopeng, and W. Pengand, "Intelligent distributed surveillance systems: a review," In: EE Proceedings: Vision, Image and Signal Processing, vol. 152, no. 2, pp. 192-204, 2005.

[176] Z. Xu and M. Zhu, "Colour-based skin detection survey and evaluation," In: the 12^{th} International Conference on Media Modelling, vol. 1, pp. 143-152, 2006.

[177] V. Vezhnevets, V. Sazonov, and A. Andreeva, "A Survey on pixel-based skin colour detection techniques," In: the 13^{th} International Conference on the Computer Graphics and Vision, Graphicon2003, Moscow, Russia, vol. 1, pp. 85-92, 2003.