



Environment-Independent Moving Cast Shadow Suppression in Video Surveillance

A dissertation submitted by **Ariel Amato** at Universitat Autònoma de Barcelona to fulfil the degree of **Doctor en Informàtica**.

Bellaterra, February 2012

Director	Dr. Mikhail G. Mozerov Centre de Visió per Computador Universitat Autònoma de Barcelona.
Co-director	Dr. Jordi González i Sabaté Centre de Visió per Computador & Dept. de Ciències de la Computació. Universitat Autònoma de Barcelona.
Thesis Committee	Dr. Rita Cucchiara Dipartimento di Ingegneria dell'Informazione. Università degli Studi di Modena. Dr. Angel D. Sappa Centre de Visió per Computador. Universitat Autònoma de Barcelona. Dr. Jose María Martínez Sánchez Escuela Politécnica Superior. Depto. de Tecno. Electrónica y Comunicaciones Universidad Autónoma de Madrid. Dr. Costantino Grana Dipartimento di Ingegneria dell'Informazione. Università degli Studi di Modena. Dr. Sergio Escalera Guerrero Dept. Matemàtica Aplicada i Anàlisi, Facultat de Matemàtiques. Universitat de Barcelona.
European Mention Evaluators	Dr. François Brémond INRIA Sophia Antipolis. Université de Nice. Dr. Michael Arens Fraunhofer IOSB.



This document was typeset by the author using L^AT_EX 2_ε.

The research described in this book was carried out at the Computer Vision Center, Universitat Autònoma de Barcelona.

Copyright © 2012 by Ariel Amato. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the author.

ISBN 978-84-938351-7-0

Printed by Ediciones Gráficas Rey, S.L.

Esta tesis está dedicada a la memoria de Milo, quien me acompañó y apoyó incondicionalmente en todos mis desafíos. De él aprendí que lo utópico no siempre es inalcanzable, y que el umbral se puede cruzar con vocación, dedicación y perseverancia. Has sido un hombre ejemplar, viviste llevando al límite tus convicciones e ideales. En estas escuetas y difusas líneas, quiero expresarte la gran admiración que he tenido y tendré hacia tu persona, y decirte que has sido un padre extraordinario.

*“Thoughts are the shadows of our feelings
- always darker, emptier and simpler.”*

Friedrich Nietzsche

Acknowledgements

Una sensación muy extraña y confusa aflora al saber que estas serán las últimas líneas que le entregue a esta tesis. Indudablemente, este trabajo pudo ser llevado a cabo gracias al esfuerzo, dedicación y determinación de muchas personas, y es por tal motivo que tengo el más sincero deseo de expresarles mi gratitud.

Mis primeros agradecimientos son para mis directores de tesis, Dr. Mozerov y Dr. González. I wish to thank my advisor, Mikhail. His advice and guidance during these years have made a substantial difference in my work. We have shared a lot of moments and discussions, starting by the 'glorious' three points up to the time travel dilemma, but in between we have worked hard. I think we did a good team. Agradecerle a Poal, por la confianza que siempre me ha brindado. El hecho de haber podido trabajar con absoluta libertad fue de un valor inmensurable para mí. A su vez, agradecerle por el enorme trabajo de gestión que realiza a diario, no solo para conmigo, sino también para con el grupo ISE.

El hecho de hoy poder estar aquí, en el CVC, escribiendo estos agradecimientos, es debido a la oportunidad que tanto el Dr. Villanueva como el Dr. Roca, en su momento me brindaron, les estoy sumamente agradecido y espero haber podido satisfacer sus expectativas.

A través de estos años, he tenido el placer de conocer gente extraordinaria, con la que pude compartir muchas vivencias. Agradecerles a Murad, Bhaskar, Marco, Ivan, Naveen y Piero, por aquellos valiosos momentos tanto dentro y fuera del marco profesional.

Agradecerles a los Doctores: Toledo, Parraga, Raducanu y Sappa por enriquecer mi mente durante tantas horas de discusiones. Agradecer también a los pasados y presentes miembros del grupo ISE: Dany, Ignasi, Pau, Carles, Andy, Miguel, Marc, Pep, Noha, Wenjuan, Zhanwu.

Como no agradecer a Montse, Mari y Gigi que constantemente están solucionando mis eternos problemas burocráticos. Agradecerles a las chicas de administración Anita, Raquel, Eva y Claire que siempre están dispuestas a ayudar. Agradecerle también a Claudia por las correcciones del inglés

Agradecerle por supuesto, al director de esta institución Dr. Lladós y hacer extensible dicho agradecimiento a todos los integrantes del CVC.

I also wish to express my gratitude to Dr. Cucchiara and her research group for welcoming me during my stay in Modena.

No olvidarme de amigos, que si bien no están aquí, igual seguimos debatiendo y compartiendo momentos a la distancia, Fabi, Ordoñez, y especialmente a Diego que

me ayudo con algunas de las ilustraciones de la tesis.

Un especial agradecimiento al Dr. Sappa, un referente ético y profesional, que en forma íntegramente desinteresada siempre puso a mi disposición su ayuda y experiencia.

Esta tesis esta escrita con dos tipos de páginas: las simples, las visibles, las que todos pueden leer, que son las que reflejan un contenido científico; y aquellas otras, las de atrás, las que sólo unos pocos pueden leer, son en definitiva las que esconden, lo que verdaderamente significó realizarla.

Sin lugar a duda, mis principales agradecimientos están dedicados, a aquellos pocos que realmente entendieron esas páginas de atrás. Para mis amigos de siempre: el Juampi, el Negro y el Fede, para un hermano Don-Angel y para mis amores: Milo, Clari, Moni y Silvi. . .

Abstract

This thesis is devoted to moving shadows detection and suppression. Shadows could be defined as the parts of the scene that are not directly illuminated by a light source due to obstructing object or objects. Often, moving shadows in images sequences are undesirable since they could cause degradation of the expected results during processing of images for object detection, segmentation, scene surveillance or similar purposes. In this thesis first moving shadow detection methods are exhaustively overviewed. Beside the mentioned methods from literature and to compensate their limitations a new moving shadow detection method is proposed. It requires no prior knowledge about the scene, nor is it restricted to assumptions about specific scene structures. Furthermore, the technique can detect both achromatic and chromatic shadows even in the presence of camouflage that occurs when foreground regions are very similar in color to shadowed regions. The method exploits local color constancy properties due to reflectance suppression over shadowed regions. To detect shadowed regions in a scene the values of the background image are divided by values of the current frame in the RGB color space. In the thesis how this luminance ratio can be used to identify segments with low gradient constancy is shown, which in turn distinguish shadows from foreground. Experimental results on a collection of publicly available datasets illustrate the superior performance of the proposed method compared with the most sophisticated state-of-the-art shadow detection algorithms. These results show that the proposed approach is robust and accurate over a broad range of shadow types and challenging video conditions.

Resum

Aquesta tesi està orientada a la detecció i l'eliminació d'ombres en moviment. Les ombres es poden definir com una part de l'escena que no està directament il·luminada, pel fet que la font d'il·luminació es troba obstruïda per un o diversos objectes. Sovint, les ombres en moviment que es troben en imatges o en seqüències de vídeo són causa d'errors en l'anàlisi del comportament humà. Això es deu a que les ombres poden causar una degradació dels resultats dels algorismes de processament d'imatges aplicats a: detecció d'objectes, segmentació, vídeo vigilància o en propòsits similars. En aquesta tesi primer s'analitzen exhaustivament els mètodes de detecció d'ombres en moviment, i després amb l'objectiu de compensar les seves limitacions es proposa un nou mètode de detecció i eliminació d'aquest tipus d'ombres. El mètode proposat no fa servir informació a priori de l'escena, ni tampoc es restringeix a un tipus d'escena en concret. A més, el mètode proposat pot detectar tant ombres acromàtiques com també les cromàtiques, fins i tot quan hi ha camuflatge (és a dir, quan hi ha una forta similitud de color entre el foreground i l'ombra). Aquest mètode explota una propietat de constància local de color aconseguida a causa de la supressió de la reflectància en les regions amb ombres. Per detectar les regions amb ombres en una escena, els valors de la imatge del background són dividits pels valors de la imatge actual, tots dos en l'espai de color RGB. Al llarg de la tesi es demostra com aquesta divisió serà utilitzada per detectar segments amb gradients baixos i constants, que al seu torn s'utilitzen per distingir entre ombres i foregrounds. Els resultats experimentals duts a terme sobre base de dades públiques mostren un rendiment superior dels mètodes proposats en aquesta Tesi, comparat amb els mètodes actuals més sofisticats de detecció i eliminació d'ombres. A més els resultats demostren que el mètode proposat és robust i precís a l'hora de detectar diferents tipus d'ombres en diferents tipus de vídeos.

Resumen

Esta tesis esta orientada a la detección y eliminación de sombras en movimiento. Las sombras pueden ser definidas como una parte de la escena que no esta directamente iluminada debido a que la fuente de iluminación se encuentra obstruida por uno o varios objetos. A menudo, las sombras en movimiento que se encuentran en imágenes o en secuencias de vídeo son indeseables. Esto se debe a que estas, pueden causar una degradación de los resultados esperados en algoritmos de procesamiento de imágenes aplicados a: detección de objetos, segmentación, video vigilancia o en similares propósitos. En esta tesis primero son exhaustivamente analizados los métodos de detección de sombras en movimiento, y luego con el objetivo de compensar sus limitaciones se propone un nuevo método de detección y eliminación. Dicho método, no usa información a priori de la escena, ni tampoco se restringe a algún tipo de escena. Además, el método propuesto puede detectar tanto sombras acromáticas como así también sombras cromáticas, incluso en presencia de camuflaje (es decir, cuando existe una fuerte similitud entre los colores de los pixeles del objeto y de los de la sombra). El método en cuestión explota una propiedad de constancia local de color, que se logra debido a la supresión de la reflectancia en las regiones con sombras. Para detectar las regiones con sombras en una escena, los valores de la imagen del fondo son divididos por los valores de la imagen actual, ambos en el espacio de color RGB. A lo largo de la tesis se demostrara como esta división será utilizada para detectar segmentos con bajos y constantes gradientes, que a su vez estos se usaran para distinguir entre sombras y objetos. Resultados experimentales llevados a cabo sobre base de datos públicas, ilustran una superior performance del método propuesto comparado con los más sofisticados métodos de detección y eliminación de sombras. Dichos resultados demuestran que el método que se propone en esta tesis, es robusto y preciso a la hora detectar diferentes tipos de sombras en videos con condiciones y características diversas.

Contents

1	Introduction	7
1.1	Motivation	9
1.2	Application Domains	10
1.2.1	Video Surveillance	10
1.2.2	Potential Applications	13
1.3	Context of this Research	15
1.4	Problem Statement and Objectives	17
1.4.1	Problem Statement	17
1.4.2	Objective	22
1.5	Method Outline	23
1.6	Contributions and Thesis outline	24
2	Related Work	25
2.1	Introduction	25
2.2	Moving Shadow Detection Methods	25
2.2.1	Taxonomy	27
2.2.2	Methods Review	33
2.3	Background Subtraction Methods	38
	Summary	43
3	Shadow Modeling and Analysis	45
3.1	Introduction	45
3.2	Shadow Modeling: The Reflection Model	46
3.3	Moving Shadow Analysis	50
3.3.1	Descriptions	51
3.3.2	Comparison	57
3.3.3	Limitations	60
3.4	Discussion	67
	Summary	69
4	Moving Cast Shadows Suppression	71
4.1	Framework Pipeline	72
4.2	Motion and Object Mask Formation	73
4.3	Reflectance Suppression	76
4.4	Shadow Region with Local Color Constancy	81

4.5	Gradient Space Connected Neighborhoods Segmentation	83
4.6	Classification Process	86
4.7	Experimental Results	91
4.8	Discussion	97
	Summary	100
5	Conclusion	101
5.1	Summary and Contributions	101
5.2	Future Lines of Research	103
A	Publications	105
	References	109

List of Tables

1.1	Thesis contributions	24
2.1	Comparison of different moving shadow detection algorithms.	39
3.1	Description of the sequences used in the comparative evaluation.	59
4.1	Description of the sequences Hallway, HW I and HW III.	92
4.2	Description of the sequences CVC Outdoor, Football Match and Pets 2009 V7.	93
4.3	Quantitative Results for different sequences.	94
4.4	Qualitative evaluation for different methods.	97
5.1	Thesis contributions	103

List of Figures

1.1	Shadows from static and moving objects.	8
1.2	Moving Cast Shadow: a challenging issue in scene understanding.	9
1.3	Images from two non-color calibrated cameras	12
1.4	Images from two cameras recording at the same time with two different viewpoints.	12
1.5	Sport Event Interpretation.	14
1.6	Human computer interaction: <i>Applications</i>	15
1.7	Background Subtraction representation.	17
1.8	Motion Segmentation.	18
1.9	Shadows Types: Self and Cast (Umbra and Penumbra).	19
1.10	Shadow-object Location.	20
1.11	Negative effect of shadow in surveillance scenarios.	21
1.12	Objective formulation.	22
1.13	Block diagram of the proposed cast shadows detection method.	23
2.1	Illustrative image of the Moving Cast Shadow Taxonomy.	27
2.2	Representation of level selection.	28
2.3	Taxonomy of different approaches.	30
3.1	The light distribution $L(\lambda)$	46
3.2	Visual illustration of the Phong model.	47
3.3	Umbra-Penumbra.	48
3.4	Umbra Penumbra spatial source variation.	49
3.5	Visual illustration of achromatic and chromatic shadow effect over a pixel information in the RGB color space.	50
3.6	Brightness distortion and color distortion in the RGB color space.	51
3.7	Thresholding zone.	53
3.8	Chromaticity space.	54
3.9	Hue Saturation Value color space.	55
3.10	Angular similarity measurement $\Delta\theta$ and Euclidean distance similarity measurement ΔI	58
3.11	Comparison of moving cast shadow detection methods in different sequences.	61
3.12	Comparison of moving cast shadow detection methods.	62

3.13	Frame with achromatic shadows	64
3.14	Frame with shadow camouflage.	65
3.15	Frame with chromatic shadows.	66
3.16	Foreground-shadow segmentation.	68
4.1	Framework Pipeline.	72
4.2	Difference in angle and magnitude in 2D "Polar RGB Color Space"	74
4.3	(a) Background image, (b) Current frame, (c) Binary moving pixel mask.	75
4.4	Thresholds evaluations.	77
4.5	Object Mask	78
4.6	Object Mask Examples	79
4.7	Umbral-penumbral transition in terms of intensity and spatial distance from the object. (a) test image; (b) illustration of the shadow parameter.	80
4.8	(a) Background image; (b) current image; (c) luminance ratio image in the RGB color space.	82
4.9	Illustration of the GSCN concept	84
4.10	Illustration of shadow splitting and over-segmentation effects.	85
4.11	SCN segmentation results in the luminance ratio space.	85
4.12	Point-wise border representation.	88
4.13	Results of different steps of the method.	89
4.13	Results of different steps of the method.	90
4.14	Results from the proposed method in different sequences: Hallway; HWI; HWIII.	95
4.15	Results from the proposed method in different sequences: CVC outdoor; Football Match; Pets 2009 V7.	96
4.16	Results from the proposed method in challenging scenarios.	98
4.17	Characteristic of the proposed approach.	99

Chapter 1

Introduction



*“The Thinker”
by Auguste Rodin.*

Motion perception is an amazing innate ability of the creatures on the planet. This adroitness entails a functional advantage that enables species to compete better in the wild. The motion perception ability is usually employed at different levels, allowing from the simplest interaction with the *'physis'* up to the most transcendental survival tasks. Among the five classical perception system ¹, vision is the most widely used in the motion perception field. Millions of years of evolution have led to a highly specialized visual system in humans, which is characterized by a tremendous accuracy as well as an extraordinary robustness. Although humans and an immense diversity of species can distinguish moving object with a seeming simplicity, it has proven to be a difficult and non trivial problem from a computational perspective.

In the field of Computer Vision, the detection of moving objects is a challenging and fundamental research area. This can be referred to as the 'origin' of vast and numerous vision-based research sub-areas. A particular domain-of-interest is represented by the semantic evaluation of human behavior in image sequences, in which different tasks are involved (e.g., detection, identification, tracking, action recognition and behavior understanding). Nevertheless, from the bottom to the top of this hierarchical analysis, the foundations still relies on *when* and *where* motion has occurred in an image.

¹Philosophy's conception was based on analyze the information obtained from the human perception. 'The Thinker' usually is used as a symbol to depict philosophy. Rodin based his theme on The Divine Comedy of Dante.



Figure 1.1: Shadows from static and moving objects. (a) Shadows casted by static objects. (b) Shadow casted by a moving object.

Pixels corresponding to moving objects in image sequences can be identified by measuring changes in their values. However, a pixel's value (representing a combination of color and brightness) could also vary due to other factors such as: variation in scene illumination, camera noise and nonlinear sensor responses among others. The challenge lies in detecting if the changes in pixels' value are caused by a genuine object movement or not. An additional challenging aspect in motion detection is represented by moving cast shadows. The paradox arises because a moving object and its cast shadow share similar motion patterns. However, a moving cast shadow is not a moving object. In fact, a shadow represents a photometric illumination effect caused by the relative position of the object with respect to the light sources.

Shadow detection methods are mainly divided in two domains depending on the application field. One normally consists of static images where shadows are casted by static objects, whereas the second one is referred to image sequences where shadows are casted by moving objects (see Fig. 1.1). For the first case, shadows can provide additional geometric and semantic cues about shape and position of its casting object as well as the localization of the light source. Although the previous information can be extracted from static images as well as video sequences, the main focus in the second area is usually change detection, scene matching or surveillance. In this context, a shadow can severely affect with the analysis and interpretation of the scene.

The work done in this thesis is focused on the second case, thus it addresses the problem of detection and removal of moving cast shadows in video sequences in order to enhance the detection of moving object.



Figure 1.2: Moving Cast Shadow: a challenging issue in scene understanding.

1.1 Motivation

Nowadays, we witness a significant increase in demand of visual information processing systems. The most desired areas are: visual surveillance, sport event interpretation, and human-computer interaction, among others. The huge amount of recorded data requires an automatic analysis and understanding of the scene, with a particular focus on moving objects.

In Computer Vision, motion analysis refers to detection, identification and tracking of objects in video sequences, where one of the most important goal is to understand and to predict the objects behavior. Even though there has been much progress in moving object detection [17, 57] during the past decades, robust and accurate moving object detection (or segmentation) still remains an open problem.

Moving Object Detection (MOD) process segments the scene into foreground (moving) and background regions. MOD is often one of the first tasks in vision-based applications, making it a critical part of the system. The success of many vision-based applications is highly related to how accurately the MOD algorithm performs the segmentation of the moving objects. However, in many circumstances the task of MOD is strongly hindered due to factors such as: global illumination changes, local illumination changes (moving cast shadows), camera noise etc. Moving Cast Shadows are one of the principal factors affecting vision-based system's performance [62, 80], since they can easily be misclassified as foreground (see Fig. 1.2).

This misclassification undoubtedly leads to a drastic and severe degradation in the moving object segmentation. For these reasons, an effective shadows detection algorithm is highly desirable for a wide range of real-world applications.

The work presented in this dissertation is motivated by the necessity to obtain a precise and accurate foreground segmentation by removing the negative effect caused by moving cast shadows. Therefore, in this thesis a moving cast shadow detection framework is presented; this can perform well over a broad range of shadow types and challenging real-video conditions.

1.2 Application Domains

Moving Cast Shadow detection is a key element in several applications. This section will briefly introduce some real-world applications where a robust motion segmentation approach (including an accurate moving cast shadow removal method) takes a fundamental role in the applications' performance.

1.2.1 Video Surveillance

We live in a Surveillance Society. Video surveillance is more prevalent in Europe than anywhere in the world. For instance, in the past decade, successive UK governments have installed over 2.4 million surveillance cameras (about one for every 14 people).² The average Londoners are estimated to have their picture recorded more than three hundred times a day³.

Video Surveillance has been in our society for a long time [14, 6]. It began in the twentieth century to assist prison officials in the discovery of escape methods. However, it was not until the late-twentieth century that surveillance expanded to include the security of property and people.

Traditionally it was used to display images on monitors inspected by guards or operators. This has allowed the observation of an increase number of places using less people and also to perform patrolling duties from the safety of a control room. However, a single operator can only monitor a limited amount of scenes simultaneously and for a limited amount of time, because the process of manual surveillance is very time-consuming and is a really tedious task.

The new breakthroughs in technology have led to a new generation of video surveillance. The current generation of video surveillance systems uses digital computing and communication technologies to improve the design of the original architecture, with the ultimate goal to create an automatic video surveillance system.

Recent trends in computer vision has delved into the study of cognitive vision systems, which uses visual information to facilitate a series of tasks on sensing, understanding, reaction and communication. In other words, video surveillance systems aim to automatically identify people, objects or events of interest in different kinds of environments. Although video surveillance is probably one of the most popular areas for research and much effort has been made to achieve an automatic system, this goal has yet to be reached.

Nowadays, the task of a video surveillance system aims to provide support to the human operator. The system warns an operator when an event, e.g., possible risks or potential dangerous situations, is detected. Despite the fact that the long-term goal is to build a completely automated systems, the short-term one is to increase the robustness of the current systems in order to reduce false alarms. This can only be achieved if the systems are able to interpret the interaction of events in the scene. This task includes detection, localization, tracking and high and low level event-reasoning.

A method that is able to perform the basic tasks, namely detection, localization and tracking with high accuracy, can highly benefit the process of scene understand-

²http://news.bbc.co.uk/2/hi/uk_news/6108496.stm

³<http://epic.org/privacy/surveillance/>

ing. Typically, a video surveillance system consists of multiple stationary cameras in offices, parking lots, banks, airports and other places of interest. In these different uncontrolled environments, many difficulties may occur. Some typical issues, which any kind of video surveillance system has to deal with, are:

- Acquisition of images: the majority of cameras belonging to a surveillance scenario provide low resolution information.
- Camera networks belonging to surveillance scenarios are not always calibrated in terms of colors or in terms of spatial geometry.
- Occlusion problems often arise because the topology of the scene, or the interaction between objects.
- There are cases where the number of objects to be detected is extremely large. Moreover, the spatial discontinuity between objects is imperceptible or it does not exist. In this overcrowded situation, the individual identification became a hard task.
- Surveillance systems, which are supposed to work 24/7, are obviously affected by different global illumination conditions (time of the day) as well as changing in the weather conditions (mostly for outdoor scenarios).
- The illumination changes are not always produced by external conditions, also local illumination modification occurs due to shadows casted by objects in the scene. Moving Cast Shadows are very frequent and extremely undesirable in video surveillance applications.

Several stages in video surveillance systems' architecture are highly affected when low resolution images are acquired. Firstly, the sensibility of movement detection method is affected due to the fact that the dynamic range of camera is diminished by noise. On the other hand, in the appearance-based recognition stage, the possibility to extract robust features or robust images descriptors is highly reduced when low resolution images are used [34].

Generally surveillance systems consist of several cameras, some of them sharing the field of view while others are pointed to different areas. Such systems should optimally relate the information acquired from different cameras. The camera calibration process' objective is to find the relationships between cameras. Cameras can be calibrated in terms of color [38, 31] or in terms of spatial localization [21, 53]. Fig. 1.3 shows images from a set of non-color calibrated cameras, the images look different despite the fact that they were all taken at the same time, under the same illumination conditions. The camera responses are different. This fact can seriously impair any kind of detection based on appearance color model. On the other hand, a spatial camera calibration can highly benefit a surveillance system. For instance, an obvious way to solve the occlusion problem is by using multiples views as it was described in the works of Amato et al. [54, 55, 3]. This fact is shown in Fig. 1.4. This figure shows two images simultaneously recorded from two different cameras located at different position. In Fig. 1.4(a) one agent is practically occluded by the other, while that in Fig. 1.4(b) the occlusion does not exist.



Frame 18 View 7



Frame 18 View 6

Figure 1.3: Images from two non-color calibrated cameras recording at the same time. *Source: Pets2010.*



(a) Frame 767 View 5



(b) Frame 767 View 3

Figure 1.4: Images from two cameras recording at the same time with two different viewpoints. (a) Pedestrian Occluded, (b) free-occlusion view. *Source: Hermes Outdoor*

Surveillance systems must be active during long periods of time, meaning to solve global illumination problems such as: natural illumination variations during the course of the day, appearance and disappearance of clouds which distort the primary source of light, etc. Moreover, outdoor systems must perform under changing weather conditions.

One big limitation in the performance of video surveillance systems arises when part of the moving object detected is shadow or when the moving shadows are detected as object. In these cases, tracking, classifying and analyzing of target object often fail. Later on, we will present common problems caused by moving cast shadows affecting surveillance applications. Summarizing, the capability to discriminate moving cast shadows and objects is crucial for the motion analysis in the field of video surveillance.

1.2.2 Potential Applications

There are another applications that can be benefited by an accurate moving object segmentation, for example:

1. Sport Event Interpretation

During decades, sport competitions have played an important role in our society. Nowadays, the money invested in sports reaches hundreds of millions if not more; part of it goes to improving the performance of elite players. Performance analysis is effectively integrated into both teams and individual athletes. This is achieved by working with an inter-disciplinary team, which includes coaches and athletes as well as other sports sciences and medicinal professionals.

Research has shown that providing athletes with accurate feedback based on systematic and objective analysis, is a key factor on improving sporting performance. This desired improvement is achieved by a combination of deep knowledge of performance's theory with a range of Vision-based technology. In this context video motion analysis is becoming a primary training tool for many athletes, coaches and trainers. Competition today is steep and teams and athletes need an extra edge to succeed. Sport Event Interpretation provides that edge with state-of-the-art in the motion analysis research area.

Basically, the computer vision techniques used in this area include: motion segmentation, tracking and human pose estimation. As we pointed out before, an accurate object segmentation is highly desirable and beneficial for the performance analysis.

Big efforts are done to improve the performance of athletes; however, vision-based motion analysis also contributes to improve the competition evaluation during the performance. The best known example is Hawk-Eye⁴, which is used as an additional artificial umpire. Hawk-Eye is mostly applied in tennis and cricket games (Fig. 1.5). Tracking and moving object segmentation are the core of this technology.

2. Human-Computer Interaction (HCI)

HCI is the study of how people interact with computers and to what extent computers are or are not developed for a successful interaction with human beings. Recently the field of human-computer interaction (see Fig. 1.6) has broadened and has payed more attention to the processes and context for the user interface.

This interesting application field explores the manner to integrate the whole human communication skills such as: speech, human motion (gestures, body poses and facial expressions) and human perception in order to interact with a computational system. The focus of research and development is now on understanding the relationships among users' goals and objectives, their personal capabilities, the social environment and the designed artifacts with which they interact.

⁴<http://www.hawkeyeinnovations.co.uk/>

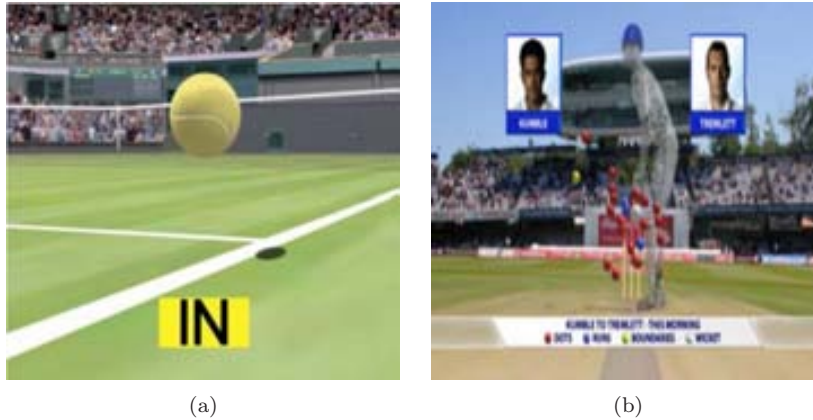


Figure 1.5: Sport Event Interpretation. Artificial umpire: (a) tennis application; (b) cricket application. *Source: <http://www.hawkeyeinnovations.co.uk>.*

As an applied field, see Fig. 1.6, is also concerned with the development process used to create an interactive system and its usefulness for the human user. The sub-areas involved in HCI are mainly Advanced Human Interface and Virtual Reality.

- **Advanced Human Interface**

The advanced human Interface can be described as the point of communication between the human user and the computer. Even though human communication is mainly based on speech, visual cues can noticeably improve the understanding of any communication process. In a human communication process, visual cues are obtained by watching the pose, gestures and facial expressions of the speakers.

In human-machine interaction, computer vision takes the role of the observer and interpreter during the dialog. Computer vision is a useful complement of speech recognition and natural language understanding, for a natural and intelligent dialog between human and machine. Additional visual information can create a more useful and friendly interface, which in turns allow for a more customized human-computer communication.

- **Virtual Reality**

Virtual reality is a form of human-computer interaction in which a real or imaginary environment is simulated and users interact with and manipulate that world. It can immerse people in an environment that would normally be unavailable due to cost, safety or perception restrictions. A successful virtual reality environment offers its users immersion, navigation and manipulation.

The realistic effects are achieved by using equipment with tiny computer screens, one in front of each eye and each giving a slightly different view so as to mimic stereoscopic vision. Sensors attached to the user, complemented with visual motion detection techniques, allow the creation of a



Figure 1.6: Human computer interaction: *Applications*.

realistic feeling of user been able to move in the scene. Virtual reality can be applied in a variety of ways. In scientific and engineering research, virtual environments are used to visually explore whatever physical world phenomenon is under study.

Training personnel for work in dangerous environments or with expensive equipment is best done through simulation. Airplane pilots, for example, train in flight simulators. Virtual reality can enable medical personnel to practice new surgical procedures on simulated individuals. As a form of entertainment, virtual reality is a highly engaging way to experience imaginary worlds and to play games.

Moving cast shadows are a major concern in today's performance from broad range of many vision-based applications because they made shape-based classification of objects very difficult. Furthermore, if objects are merged together due to their shadows then tracking cannot be accurately performed. Therefore, the success of applications that include motion analysis is highly related on the capacity to discriminate between moving shadows and moving objects.

1.3 Context of this Research

The work done in this thesis has been realized in an European project entitled HERMES IST-2006-027110 (*Human Expressive Representations of Motion and their Eval-*

uation in Sequences), under the umbrella of the Video Surveillance field.

HERMES was a consortium project that concentrated on how to extract descriptions of people behavior from videos, such as pedestrians crossing inner-city roads, approaching or waiting at stops of buses and even humans in indoor worlds like an airport hall, a train station, or a lobby.

These video recordings (from different parts of Europe in order to prevent over-adaptation to local habits) allowed exploring a coherent evaluation of human movements and facial expressions across a wide variation of scale.

The main objective of HERMES was to develop a cognitive artificial system based on a framework model which allows both recognition and description of a particular set of human behaviors arising from real-world events. Specifically, HERMES proposed to model the knowledge about the environment in order to make or suggest interpretations from motion events, and to communicate with people using natural language texts, audio or synthetic films. These events were detected in image data-streams obtained from arrays of multiple active cameras (including zoom, pan and tilt).

HERMES thus aimed to design a Cognitive Vision System for human motion and behavior understanding, followed by communication of the system results to end-users, based on two main goals. The first goal was to determine which interpretations are feasible to be derived in each category of human motion. Consequently, for each category, suitable human-expressive representations of motion have been developed and tested. In particular, HERMES interpreted and combined the knowledge inferred from three different categories of human motion, namely the motion of agent, body and face, in the same discourse domain.

The second objective of HERMES was set to establish how these three types of interpretations can be linked together in order to coherently evaluate the human motion as a whole in image sequences. Such evaluation has required to acquire human motion from video cameras, to represent the human motion using computational models, to understand the developments observed within a scene using high-level descriptions, and to communicate the inferred interpretations to a human operator by means of natural language texts or synthesized virtual agents as a visual language. Thus, the main procedure of HERMES was the combination of:

- Detection and tracking of agents while they are still at some distance away from a particular location (for example a bus station, a pedestrian crossing, or a passenger in an airport, or a guest in a lobby).
- When these agents come closer to the camera, when the active camera zooms in on these agents, their body posture will be evaluated to check for compatibility with behavior hypotheses generated so far.
- If they are even closer and their face can be modeled sufficiently well, facial emotions will be checked in order to see whether these again are compatible with what one expects from their movements and posture in the observational and locational context which has been accumulated so far by the system.

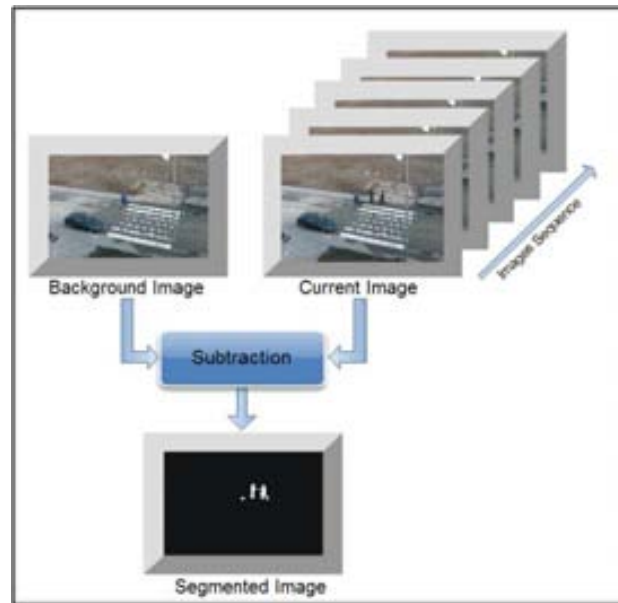


Figure 1.7: Background Subtraction representation.

1.4 Problem Statement and Objectives

1.4.1 Problem Statement

Moving object detection plays an important role in computer vision [17, 57]. It is a necessary pre-processing step for many real-world applications.

One of the most common, simple and effective approach to moving object segmentation is Background Subtraction, where a stationary camera is used to observe dynamic events in a scene. The methodology behind any background subtraction technique consists in subtracting a model of the static scene 'background' from each frame of a video sequence (see Fig. 1.7).

In general, a background subtraction technique can be divided into three phases: first, the generation of a suitable reference model, normally called background (training phase); second, the measurement procedure or classification (running phase) and finally; the model maintenance (updating phase).

For each of these phases, particular challenging exist. A deep case study together with solutions of some of these classical issues is reported in [30].

In moving object detection algorithms, moving cast shadows have a high probability to be misclassified as moving objects (foregrounds). Such error is due to the fact that a moving object and its moving shadow share similar motional characteristics. An example of motion segmentation image based on background subtraction process is shown in Fig. 1.8(c). The segmented image shows that the shadow was also segmented as a part of the object.

A shadow occurs when an object partially or totally blocks the direct light source.

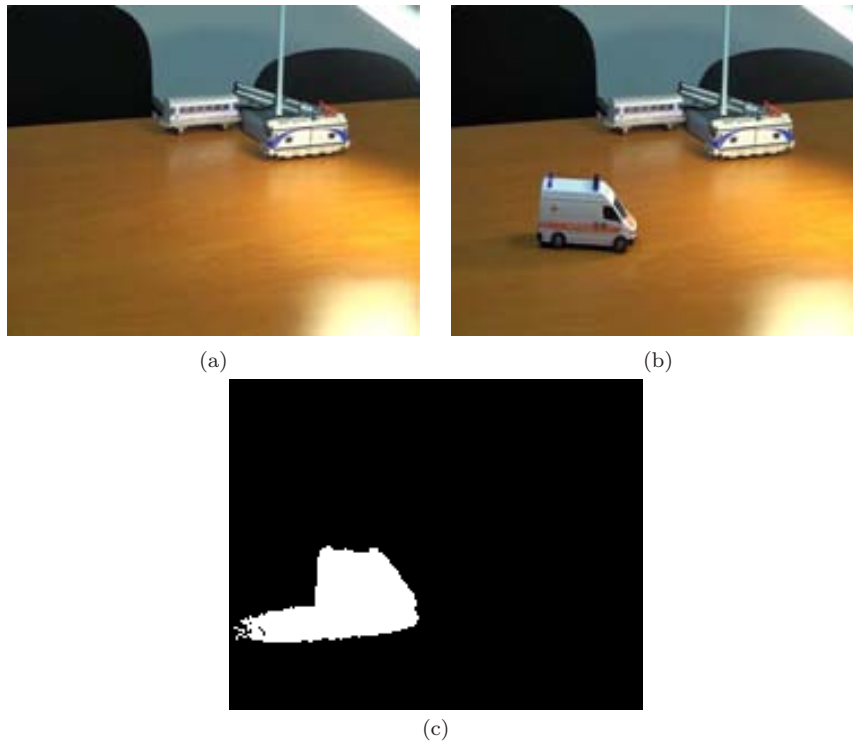


Figure 1.8: Motion Segmentation: (a) background Image; (b) current image; and (c) Segmented Image.

Shadows can take any size and shape. In general, shadows can be divided into two major classes: self and cast shadows. A self shadow occurs in the portion of an object that is not illuminated by direct light. Cast shadows are the areas projected on a surface in the direction of direct light. Cast shadows can be further classified into umbra and penumbra. The region where the direct light source is totally blocked is called the umbra, while the region where it is partially blocked is known as the penumbra. These definitions are visually represented in Fig. 1.9.

Shadows in images are generally divided into static and dynamic shadows. Static shadows are shadows due to statics objects such as building, parked cars, trees, etc. Moving object detection methods do not suffer from static shadows since static shadows are modeled as a part of background. In contrary, dynamic (moving) shadows, the subject of interest in this thesis, are harmful for moving object detection methods. These appear due to moving object such as vehicles, pedestrians, etc.

The shadows can be either in contact with the moving object, or disconnected from it (see Fig. 1.10). In the first case, shadows distort the object shape, making the use of subsequent shape recognition methods less reliable. In the second case, the shadows may be wrongly classified as an object in the scene.

For example, typical problems caused by moving shadows in surveillance scenarios are shown in Fig. 1.11. In Fig. 1.11-(I), a traffic surveillance scene, shadows cause

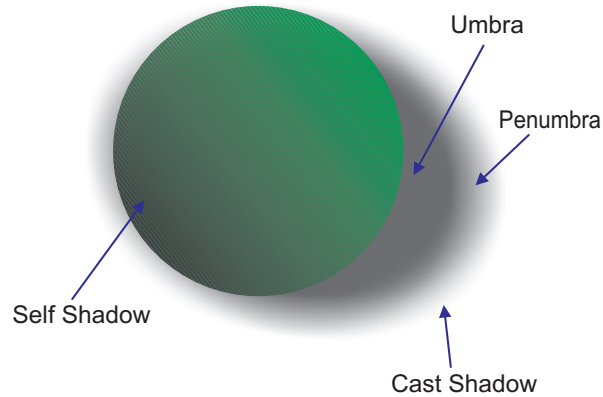


Figure 1.9: Shadows Types: Self and Cast (Umbra and Penumbra).

merging of multiple objects; in Fig. 1.11(II), an indoor scenario, shadows are projected on the floor and on the wall. In this case a false positive foreground (shadow casted on the wall) occurs; and in Fig. 1.11(III), a long shadow causes a severe object shape distortion in an outdoor scenario.

Clearly, in many image analysis applications, the existence of moving cast shadows may lead to an inaccurate object segmentation. Consequently, tasks such as object description and tracking are severely affected, thus inducing an erroneous scene analysis.

Shadows are normally considered as a local illumination problem. Obviously, areas affected by cast shadow experience a change of illumination. Often this illumination change is considered only as a decrease in brightness, without significant variation in chromaticity. However, the assumption that chromaticity is invariant to cast shadows is not always correct. It is correct, in fact, only when light sources are white and there is no color blending among objects.

This type of shadow is often called an achromatic shadow, while those that are not achromatic are referred to as chromatic shadows. Removing chromatic shadows is a particularly challenging task due to the fact that they are extremely difficult to distinguish from the foreground because they have not a clearly defined photometric pattern. The interplay between color and texture in the background and shadows is highly variable and difficult to characterize.

Another non trivial problem occurs when there is no difference in chromaticity between foreground object and background (e.g. black car is moving in highway), hence inducing a strong similarity between shadow-foreground regions. Such effect is called as shadow camouflage.

Despite of the fact that many articles of moving cast shadow detection have been published during the last years, only few works in the literature address these two major problems: **chromatic shadow identification**, and **shadow detection in camouflaged areas**.



(a)



(b)

Figure 1.10: Shadow-object Location.

Shadows location: (a) shadow is spatially connected to the object; (b) shadow is spatially unconnected to the object.

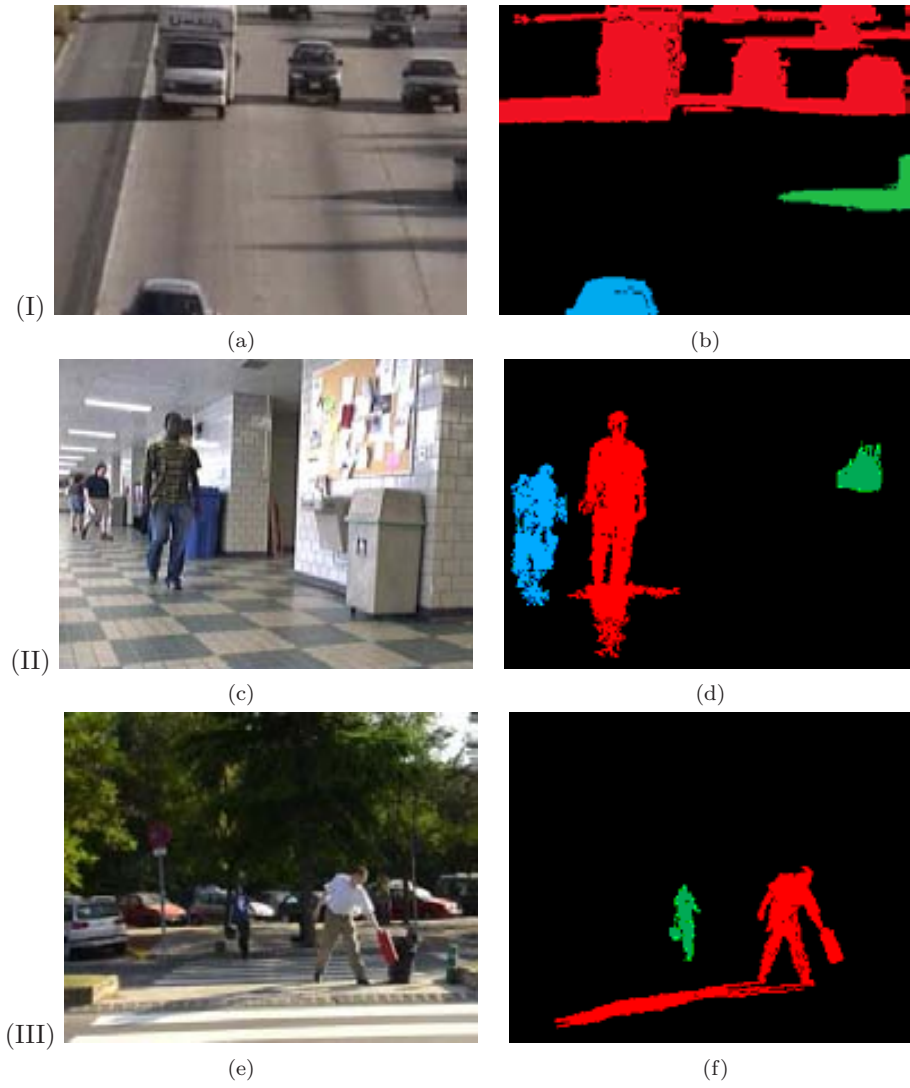


Figure 1.11: Negative effect of shadow in surveillance scenarios.

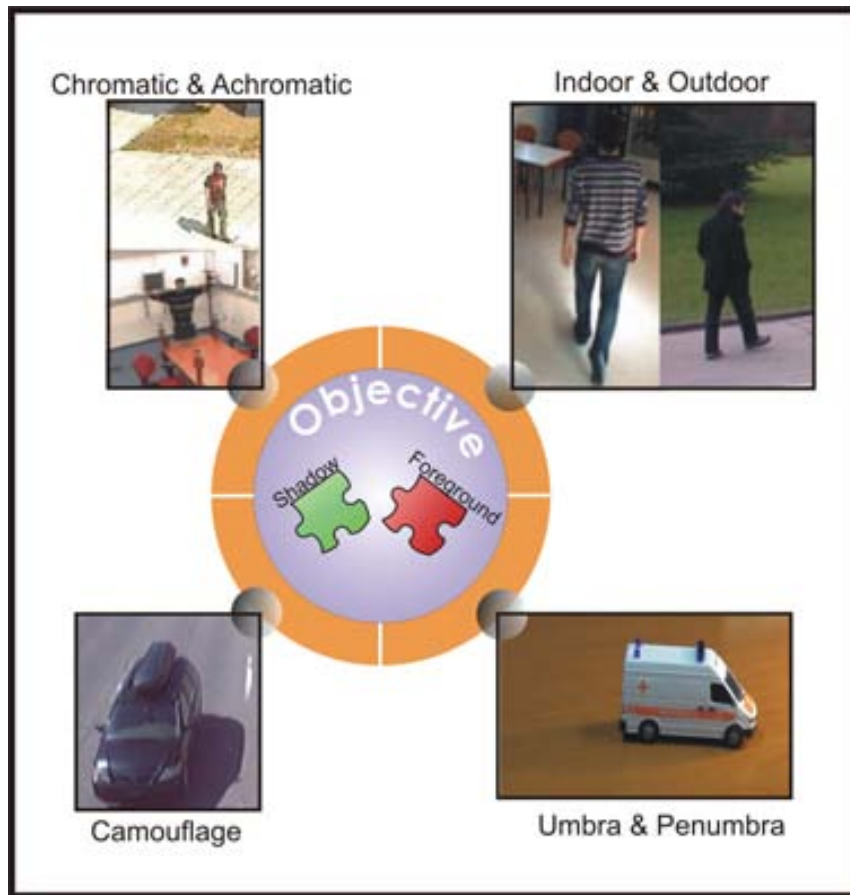


Figure 1.12: Objective formulation.

1.4.2 Objective

Previously, different moving cast shadows characteristics have been described together with the most common problems that normally moving shadows might cause for moving object segmentation algorithms. The main objective of this dissertation is to develop a moving shadow detector capable of detecting moving cast shadows for most possible scenarios occurring in real video sequences. Thus, the proposed method has to operate under the following conditions: *(i)* for indoor as well as outdoor scenarios; *(ii)* detect both umbra and penumbra; *(iii)* detect chromatic and achromatic shadows; and *(iv)* recognize shadows even in the presence of camouflage. Figure 1.12 illustrates this objective.

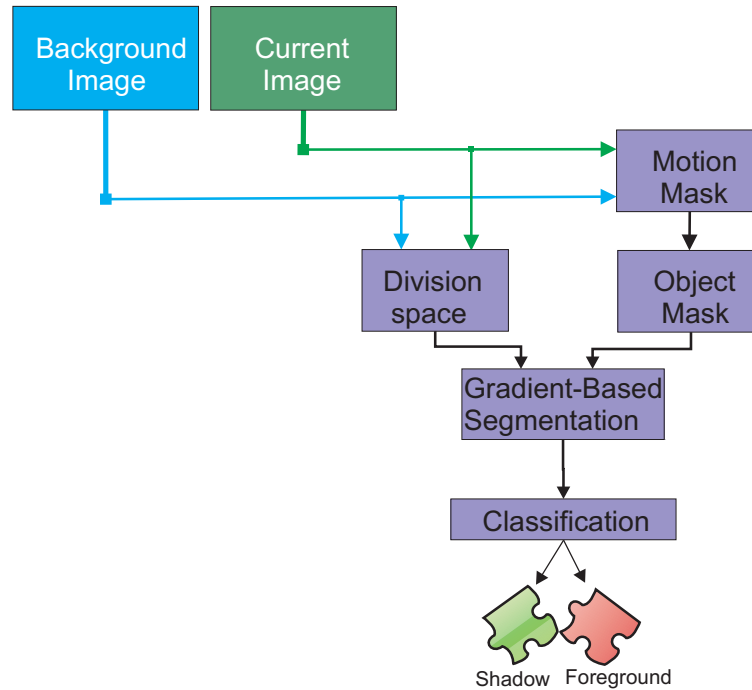


Figure 1.13: Block diagram of the proposed cast shadows detection method.

1.5 Method Outline

A novel approach to detect and remove moving cast shadows is presented in this thesis. This section depicts the outline of the proposed shadow detector. The method exploits local color constancy due to reflectance suppression over shadowed regions. Such a color constancy effect, over a shadow region, is achieved through dividing the values of the background image by the values of the current image to form a new image in the division space. In this space, segments with low gradients correspond to all shadow regions, as opposed to foreground regions which, in most cases, exhibit higher gradients.

A brief description of the main steps of the method are given next. First, an initial change detection mask containing moving objects and cast shadows is obtained using a background subtraction technique. Then, the objects' mask is computed by using connected component analysis. Next, the proposed method is used to detect cast shadow areas inside each detected moving object segment. To do this, luminance values of the background image are divided by the corresponding luminance values of the current frame, thus suppressing the reflectance component in shadow areas. Later, each object area is partitioned into a set of segments using a novel gradient-based segmentation algorithm.

Finally, these sub-segments are classified as foreground or shadow by analyzing the intrinsic parameters of sub-segments. Fig. 1.13 shows the block diagram.

LOCATION	CONTRIBUTIONS
Chapter 2	A comprehensive review of the literature. A new taxonomy of moving cast shadow methods.
Chapter 3	A deep analysis of shadow modeling. An exhaustive examination of relevant methods, which includes description, comparison and limitations.
Chapter 4	A motion detection algorithm based on angular and euclidean distance similarity measurement. A shadow feature based on low gradient region in the luminance ratio space. A gradient-based segmentation approach. A geometrical shadow supporter based on external terminal points.

Table 1.1
THESIS CONTRIBUTIONS

1.6 Contributions and Thesis outline

The outcome of this work ends with a new moving cast shadow detector, which is able to properly performs in most possible scenarios occurring in real video sequences. Furthermore, the proposed detector outperforms the state-of-art [2].

Table 1.1 summarizes an account of the specific contributions achieved throughout the development of this thesis and the chapters where such contributions appear.

This thesis is organized in the following chapters. **Chapter 2** gives a literature review on Moving Cast Shadow methods, describing different algorithms developed during the last years. In turn, some relevant Background Subtraction techniques are also revised. **Chapter 3** is divided into two parts. The first one, is focused on the shadow modeling from a photometric viewpoint. While, in the second, a deep analysis of common shadow detection methods is presented. In **Chapter 4** the description of the proposed moving cast shadow detection framework is given. Finally, **Chapter 5** briefly reviews the topics discussed in the different sections of this thesis, establishes future lines of research, and, as a conclusion, summarizes the contributions of the thesis.

Chapter 2

Related Work

*“Que otros se jacten de las páginas que han escrito;
a mí me enorgullecen las que he leído.”*

Jorge Luis Borges

2.1 Introduction

Several shadow detection methods have been reported in the literature during the last years. They are mainly divided into two domains. One usually work with static images [16, 63, 78], whereas the second one uses image sequences, namely video content.

In spite of the fact that both cases can be analogously analyzed, there is a difference in the application field. The first case, shadow detection methods can be exploited in order to obtain additional geometric and semantic cues about shape and position of its casting object (‘shape from shadows’) as well as the localization of the light source [40, 60]. While in the second one, the main purpose is usually change detection, scene matching or surveillance (usually in a background subtraction context). Shadows can in fact modify in a negative way the shape and color of the target object and therefore affect the performance of scene analysis and interpretation in many applications, such as video retrieval, as well as video analysis.

This chapter will mainly reviews shadow detection methods related with the second case, thus aiming at those shadows which are associated with moving objects (*moving shadows*). As it has been mentioned, the moving shadows detection problem is highly related to background subtraction techniques, therefore in this chapter classical and well-know background subtraction algorithms will also be studied.

2.2 Moving Shadow Detection Methods

Moving cast shadow detection algorithms are mainly based on the use of shadow descriptors. They basically model shadows by using properties such as: chromaticity

invariant, textural patterns, photometric physical models, or even by analyzing the projected areas in term of size, shape and direction.

The methodology of moving cast shadow detection can further includes geometrical-shadow-information or spatial-shadow-cues as well as a training-shadow-stage, or a sort of combination of them. In turn, the methods can perform at different levels, considering only the information of a single pixel, using a set of pixels, or even performing with the information of the whole frame.

Diverse information that characterizes moving shadows is exploited and in many cases such information is combined or used in a different way. This makes very difficult to classify in a unique manner the moving cast shadow methods.

Actually, in the literature we only found three taxonomies used to classify moving shadow detection methods: Salvador et al.[67] propose to divide shadow detection methods in *(i)* model-based and *(ii)* property-based. Model-based approaches work with models that represent a priori knowledge of the geometry of the scene, the object and the illumination. While property-based methods identify shadows by using properties such as the geometry, brightness and/or color of shadows. This is not a very useful classification. First of all this taxonomy is quite general. Furthermore, the majority of the state-of-art methods use some property/ies to detect moving shadows, even the model-based approaches.

Ullab et al.[80] state that moving shadow removal method can be partitioned into three categories: *(i)* intensity information, *(ii)* photometric invariant information and *(iii)* color and statistical information. The first classification concentrates in the brightness of the shadowed pixels. Typically a shadowed pixel decreases its brightness compared to the same pixel without shadow. The second classification includes those algorithms that exploit photometric-invariant-shadow property. Normally such photometric invariability can be obtained in normalized color spaces that can separately operate with the brightness and the chroma of the pixels. The last classification stands for methods which usually classify shadow by using statistical model of the pixel's information. In this classification two main drawback can be pointed out.

The first one, an ambiguity can occurs when the methods are classified as category *(i)*, since the majority of the methods exploit this property of shadows. The second one is that relevant information used to detect moving shadows are not included in the classification (e.g., spatial, temporal, prior-knowledge of scene, etc.). These are basically the reasons of why we have found that this taxonomy can not classifies known methods properly.

Undoubtedly, the most complete taxonomy available in the literature was proposed by Prati et al. [62], they presented two layers taxonomy (*algorithm-based taxonomy*). The first layer classification considers whether the decision process introduces and exploits uncertainty. Deterministic approaches use an on/off decision process, whereas statistical approaches use probabilistic functions to describe the class membership. In turn, both layers are further divided. For statistical approaches the authors include parametric and non-parametric separation. In the case of deterministic methods, algorithms are classified by whether the decision is supported by model-based knowledge or not. Additionally, spectral, spatial and temporal information are also considered.

Summarizing our review of the known taxonomies we can state that the greatest distinction behind moving cast shadow methodologies does not reside in the decision

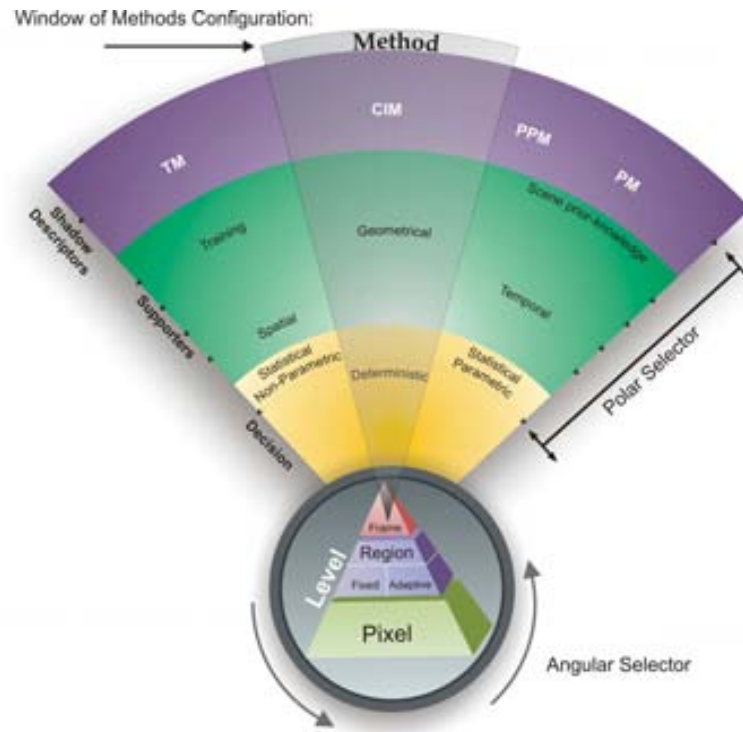


Figure 2.1: Illustrative image of the Moving Cast Shadow Taxonomy. TM, CIM, PPM and PM stand for Texture Model, Color and Intensity Model, Photometric Physical Model and Projection Model respectively.

process. The highest diversity reflected in the results as well as the capability of methods to perform under different scene's conditions are in the choice of this feature/s used, how the feature/s is used and combined, and essentially in the *manner* or *level* that this feature/s is implemented. This realization was the motivation to propose a new moving cast shadow taxonomy.

2.2.1 Taxonomy

The proposed taxonomy starts with three operative levels. In turn, the methods are classified based on the descriptors or properties used and how such descriptors are implemented. In Fig. 2.1 an illustrative graphic of the proposed taxonomy is shown. The illustration consist of three main components:

1. **Angular Selector:**

It is represented by the circle that surround the pyramid level. A rotation of 0° is used for *frame level*, 90° for *adaptive-region level*, 180° for *pixel level* and 270° for *fixed-region level*. (see Fig. 2.2)

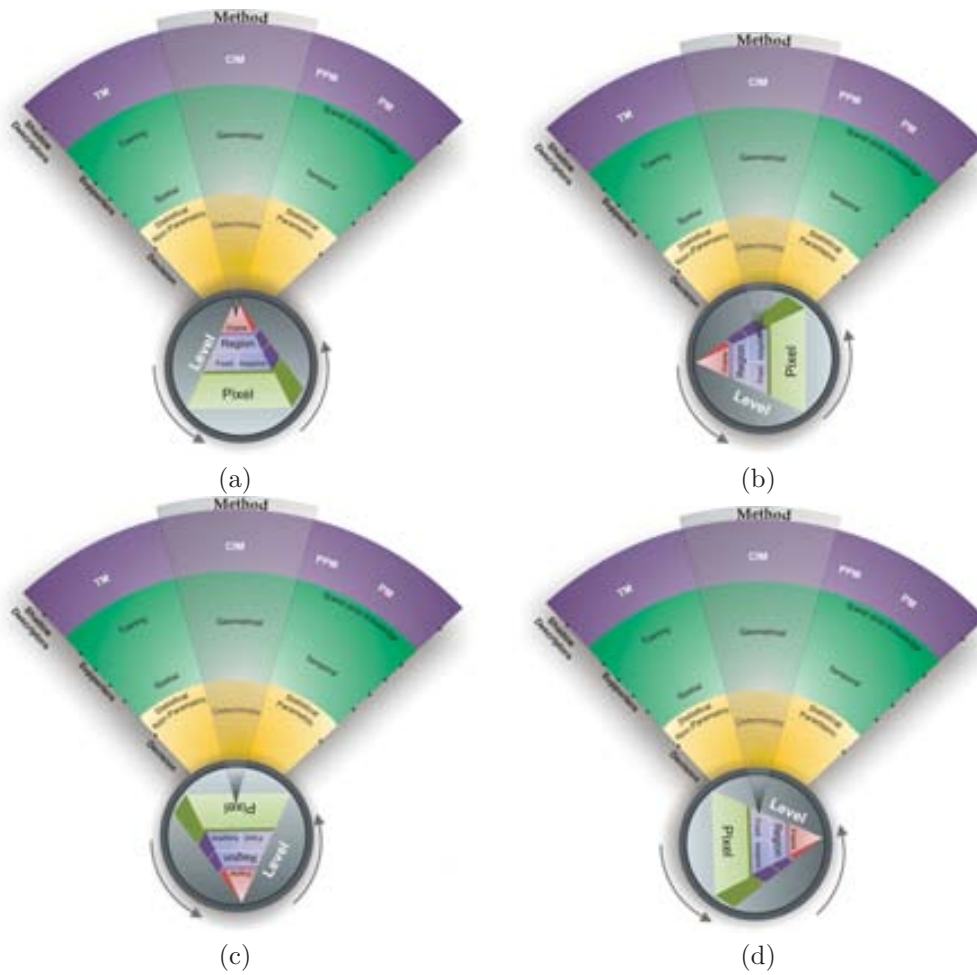


Figure 2.2: Representation of level selection. (a) Frame Level, (b) Adaptive-Region Level, (c) Pixel Level and (d) Fixed-Region Level.

2. Polar Selector:

It is used to set up the configuration of the method in term of the shadow descriptor used as well as the supporters and the decision process.

3. Window of Method Configuration:

It is represented by a fixed semi-transparent cone, that determines the configuration of the method. In Fig. 2.3 we show the taxonomies of four different methods of the literature.

The main layer of the taxonomy is related with the operative level of the method. In other words, it describes whereas the methods operate with a single pixel or with a group of pixels or using the whole frame information. A distinction must be done with the region level, meaning that the region level can performs within a fixed area (usually with a window of pixels) or within a local adaptive region meaning a particular region of pixels.

On the other hand, the methods can be classified based on shadow descriptor used. The principal shadow descriptors used in the literature are: chromaticity and intensity models (CIM), texture models (TM), photometric physical models (PPM) and projection models (PM).

In turn, some methods need to be supported by an extra information. Such information can be obtained by a training phase or by using some prior-knowledge of the scene or by exploiting geometrical, temporal, or spatial cues.

Finally the last category is based on the classification's decision of the methods. It was inspired by the work of Prati et. al. [62] where the classification considers whether the decision process introduces and exploits uncertainty or not. This category discriminates between deterministic and statistic algorithms.

A detailed components description of the proposed taxonomy is given below.

• Shadow Descriptors

– Chromaticity and Intensity Models (CIM)

Many shadow detection method assume that a region under shadow becomes darker but with a similar chroma that the same region without shadow. Chromaticity is a measurement of color that is independent of intensity component.

The invariability in chroma, between a (non-shadowed) pixel belonging to the background and the same (shadowed) pixel belonging to the current image, together with a brightness decrement, represent a distinctive shadow feature¹.

Often methods that are using this shadow descriptor perform in color spaces where the distinction between brightness and chroma is supported. These common spaces are: HSV, HSI, YUV, C1C2C3, normalized RGB, etc.[25, 13, 50, 67, 66].

¹Earlier approaches perform with gray scale images; they exploit the same intensity's decrement assumption. Therefore, they could be consider to be inside this category.

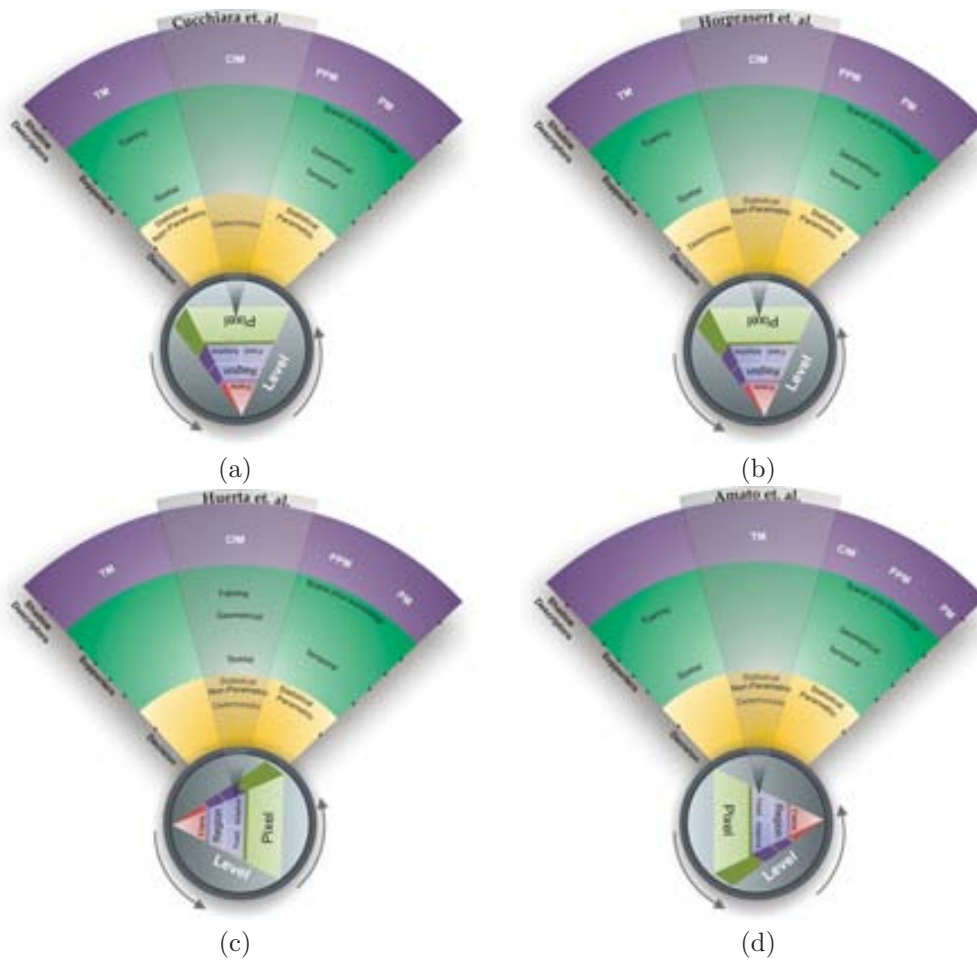


Figure 2.3: Taxonomy of different approaches. (a) Cucchiara et al. [13], (b) Horprasert et al. [25], (c) Huerta et al. [29] and (d) Amato et al. [1]

- *Texture Models (TM)*

Methods that use texture as shadow descriptor basically are based on the idea that a shadow is a semi-transparent region in the image. Thus, they assume that a strong correlation between two regions, one affected by shadow and the same region without the shadow's effect, must exist.

These methods try to obtain such a correlation using for example: local binary patterns (LBP), normalized cross-correlation (NCC), color cross covariant (CCC), Markov random field, etc. (e.g. [19, 91, 23, 89, 1]).

- *Photometric Physical Models (PPM)*

Methods in this category aim to obtain an inference of the pixel values in the shadowed areas. For such a purpose, a formulation can be achieved by using: a reflectance model, an illumination model or an adaptation of classical color models.

In order to obtain the appearance of the shadowed pixels some methods may need a training phase (it could be supervised or unsupervised), or/and some prior knowledge of the scene of interest. (e.g. [56, 46, 47, 28, 88]).

- *Projection Models (PM)*

Methods that use this shadow descriptor are focused on the characteristic of the casted shadow area. Usually the characteristics to be analyzed are direction, size and shape of the shadow. Often these methods operate in two steps. In the first one, the border points between the object and its shadow are computed. The second step aim to obtain a refined segmentation by estimating the line through the border points between objects and its shadow.

These methods can be hardly able to avoid the use of some prior knowledge of the scene (e.g. [8, 26, 27, 90]).

- **Supporters**

This category attempts to describe whereas methods combine some additional information or not, and which kind of information is used. The difference between the *supporters* and the *shadows descriptors* is highlighted next.

The *supporters* are not used as the core of moving cast shadow methods, they are only implemented as an additional information used to improve detection.

- *Scene prior-knowledge*

There are methods that make use of the information such as: camera location, ground surface, geometry of the objects and classes, illumination condition, etc. Obviously, the use of scene's prior-knowledge can significantly improve the detection. Nevertheless, the way in which the information of classes of objects is used in term of their photometric structure as well as their geometry is not a trivial problem for complex environments, where many objects interact in the scene.

- *Training*

Several methods make use of this supporter to obtain a better inference in the shadow-foreground classification. Basically, the idea behind this technique is to implement a training stage, which can be supervised or unsupervised, in order to obtain a more accurate approximation of the shadowed pixel appearance. However, a strict relationship exists between the accuracy in the achieved results and the training phase.

- *Geometrical*

A significant improvement can be achieved when the detection is enriched by using geometrical and topological clues that characterize a shadow. An example can be found in the consideration that a casted shadow pixel or region cannot be inside the foreground²; a shadowed region should surround the object.

Another assumption frequently used considers a fix light source; therefore, the shadow is casted in a certain direction. Often, this is used in outdoor scenarios to detect shadow casted by rigid objects (mainly in traffic surveillance) because the direction as well as the relative size of the shadow can be pre-computed.

- *Temporal*

Temporal information is also exploited by some moving cast shadow removal methods. These methods take advantage of certain temporal constancy of the shadowed pixels to infer the shadow-foreground classification of the consecutive frames (multi-frame integration).

Broadly speaking, a multi-frame integration methodology evaluates a confidence detection map to better infer the detection of consecutive frames. More sophisticated approaches use trackers of shadows.

- *Spatial*

Following the principle of the shadow descriptor based on texture, where a shadow can be considered as a semi-transparent area, methods that use the spatial supporter try to find an extra shadow clue by analyzing edges, gradient density or similar spatial-features under the shadowed area.

Undoubtedly, in many circumstances the detection of moving cast shadow can be highly benefited by the use of supporters. However, not all real video sequences can hold the constrains as well as the assumption needed for some supporters. Furthermore, in the majority of the cases, some of the aforementioned supporters attempt to loss generality, making the detection useful only for some specific and controlled sequences.

- **Decision**

This category discriminates between deterministic and statistic algorithms. Unlike Prati et al. [62], it should be observed that there are methods that combine the uses of statistical inferences with some deterministic decision. Therefore, we propose a not mutually exclusive separation.

²Note that the topic of interest in this research is not focused in the self-shadow, this work addresses moving cast shadow.

The decision category could be deterministic and/or statistic and this one can be parametric or non-parametric.

2.2.2 Methods Review

In this subsection some of the most classical moving cast shadow detection approaches are presented. They are organized in pixel-level, region-level and frame-level³.

Pixel-Level

In general, the main shadow descriptors used by a method that performs at the pixel-level, are CIM and PPM. Then, the supporter and the decision used by them can vary. Some methods can be chosen for their simplicity, for their speed, or for their accuracy as well as for a trade-off between these advantages. However, the pixel-level methodology in moving cast shadow detection methods intrinsically imply an inability to solve shadow camouflage.

Methods that exploit CIM as shadow descriptors were proposed by the following authors:

Horprasert et al. [25] propose a color model that compares intensity to the chromaticity component at each pixel. Each pixel is classified as background, shaded, highlighted or moving foreground through a combination of three threshold values, which are defined over a single Gaussian distribution. An extension of this work based on multiple background pixels organized in a codebook is done by Kim et al. [36].

Cucchiara et al. [13] use shadow properties in the HSV color space to distinguish shadows from moving objects. These properties show that cast shadows darken the background in the luminance component, while the hue and saturation components change within certain limits.

McKenna et al. [50] assume that cast shadows result in significant change in intensity without much change in chromaticity. Pixel's chromaticity is modeled using its mean and variance. In turn, the first-order gradient of each background pixel is also exploited. Moving shadows are then classified as background if the chromaticity or gradient information supports their classification as such.

The advantage of all the mentioned above methods reside in that they are fast (suitable for real-time applications), and easy to implement. However, they are specially restricted to achromatic shadows. Moreover, some of them often require explicit tuning of parameters for each scene.

PPM that implement statistical learning-based methodology have been developed to learn and remove cast shadows [46], [61], [47], [28]. For example, in the work of [47] a nonparametric framework to model surface behavior when shadows are cast on them is introduced. Physical properties of light sources and surfaces are employed in order to identify a direction in RGB space at which background surface values under cast shadows are found. However, these approaches are particularly affected by the training phase. These methods require a long training period.

In the work of Siala et al. [71] a statistical non-parametric shadow detection

³Frame-level methods are included, despite the fact that they are not widely used, in order to obtain a thorough review of the methods.

method is presented. First, in the learning phase, an image containing foreground, background and moving shadow is selected. The moving shadow regions are manually annotated. The information obtained from this annotation is then used to create a diagonal model that describes the shadow appearance in the RGB ratio color space. The shadow detection is obtained by performing a one class classification based on a support vector domain description (SVDD).

Region-Level

As it has been mentioned before, region-level methods are separated into fixed region and adaptive region. A fixed region, normally is predefined as a square or rectangular window with a fixed size, while the adaptive region has not predefined neither a shape nor the size of the area.

Fixed:

Although most of the methods that perform in a fixed region typically make use of texture information (TM), there are few methods that exploit other shadow descriptor such as: CIM or PPM. Next, a brief description of moving cast shadow detection methods that perform at fixed region level is presented.

Grest et al.[19] propose to tackle the moving cast shadows detection using two similarity measurements, one is based on the normalized cross correlation (NCC) and the other is the color cross covariant (CCC). Basically the authors are interested in comparing pixel values at the same position in two images, (the current image and a reference image) and then inferring if there is a correlation between the information of these pixels. The computation of these measurements are done over a given window size. The NCC is calculated using the brightness of the pixel, while the CCC is obtained in the biconic HSL (Hue, Saturation and Lightness) color space. The authors assume that: *(i)* a shadowed pixel is darker than the corresponding pixel in the background image; *(ii)* the texture of the shadowed region is correlated with the corresponding texture of the background image. Despite the fact that CCC is used to solve the limitation of the method to distinguish shadow from object over homogeneous areas, still the success of the approach under shadows-camouflage areas is far to be achieved.

Other approach based on NCC is proposed by Yuan et al. [91]. The authors proposed to include a multi-frame differencing strategy to improve the segmentation in those cases where the shadows cannot successfully be removed. This strategy is based on that shadowed regions differ a little in two consecutive frames. Therefore the biggest part of the shadows can be eliminated by frame difference, but only remain some shadow edges. These shadow edges are removed by using a new frame differencing step.

Jacques et al.[7] propose to detect shadows regions by using intensity measurement of a set of pixels. This measurement are computed by ratio-pixels (image/background) in a fixed 3×3 windows and the decision is based on a statistical non-parametric inference.

In the work of Yao et al. [89], textures are computed using the LBP combined with a RGB color model. The authors state that LBP can work robustly to detect moving shadows on rich texture regions. However, it fails when both the background image and the foreground objects share the same texture information. Therefore, to handle these situations, in this work the authors make use of a shadow invariant color distance in the RGB color space. They claim that pixel values changed due to shadows are mostly distributed along in the axis going toward the RGB origin point. Thus, they propose to compare the color difference between an observed color pixel and a background color pixel using their relative angle in RGB color space with respect to the origin and the changing range of the background color pixel up to last time instant.

Leone et al. [37] use a textural shadow descriptor by projecting the neighborhood of pixels onto a set of Gabor functions, extracted by applying a generalized scheme of the Matching Pursuit strategy. The methodology for shadow detection is based on the observation that shadows are half-transparent regions which retain the representation of the underlying background surface pattern. This approach assumes that shadow-regions contain same textural information, both in the current and in the background images.

In the work of Amato et al. [1] a method that introduces two discriminative features to detect moving cast shadow is presented. These features are computed based on angular and modular patterns, which are formed by similarity measurement between two sets of RGB color vectors. Unlike the most texture-based methods that often exploit spatial information, the patterns used in this approach are only photometric. This method could also be categorized as CIM since it makes use of chroma and intensity information of a set of pixels to form a textural pattern.

Salvador et al. [67, 66] introduce a two stage method for segmenting moving shadows. The first stage segments the moving shadows in each frame of the sequence. In this stage the property that shadows casted on a surface reduce the surface intensities is exploited by using the photometric invariant C1C2C3 color space. In addition, to obtain a more robust result, the authors propose two schemes: (i) analyze a set of pixels (neighborhood) instead of a single pixel, and (ii) include geometrical verification based on boundary analysis of the shadow-candidate regions and testing the position of shadows with respect to objects. The second stage is used to obtain a coherent description of the segmented shadows over time. Therefore, the authors introduce a tracking shadow algorithm. An extension of this work was presented in [68] where the algorithm can segment cast shadow for both still and moving images.

Yang et al. [88] propose a PPM moving cast shadow detection algorithm that combines shading, color, texture, neighborhoods and temporal consistency in the scene.

In comparison with methods that perform at the pixel-level, the aforementioned methods normally exploit texture information or use information from a set of pixels, making the detection more robust against noise and more efficient in those cases where ambiguity in the pixel's information occurs. However, the main drawback of these methods reside in the choice of the region's size that will be used. In other words, a strong dependency between the size of the region and the success of the method exists.

Many factors are involved in the choice of the region's size, for example: size of the object, textural composition of the background as well as of the object, etc. Consequently, an optimal region's size highly depends on the scene; moreover, the optimal size can change for different frames, even the optimal region size can change within the frame. Furthermore, in many cases the computational time will vary with the size of such a region.

Adaptive:

These methods perform with local adaptive regions. Basically they attempt to segment the moving area and then analyze and classify each segment based on shadow properties. These methods take advantage from pixel-level methods since they can make use of the information of a set of pixels. Additionally, they have also an advantage with respect to the fixed region-level methods since they can automatically adapt the area of analysis. A summary of some common adaptive methods is given below.

Toth et al. [76] propose a shadow detection algorithm based on color and shading information. They segment an image into several regions based on color information and the mean shift algorithm. They consider that the intensity values of a shadow pixel divided by the same pixel in the background image should be constant over a small segment.

In [18] an algorithm for outdoor scenarios is presented. Luminance, chrominance and gradient density information are exploited to create a shadow confidence score. Such a shadow score is based on three rules. The first rule claims that the luminance of the cast shadow is lower than the background. The second rule claims that the chrominance of the cast shadow is identical or slightly shifted when compared with background. And the last rule claims that the difference in gradient density between the cast shadow and background is lower than the difference in the distance of gradient between the object and background.

The final classification combines the shadow score with a geometrical supporter. The geometrical cue used is based on the fact that the cast shadow is at the boundary region of moving foreground mask. That is, the cast shadow can be formed in any direction of the object, but not inside the object. However, the method is restricted to: (i) the areas where the shadows are casted on are not textured and (ii) the object shape is a convex hull which makes inappropriate to detect non-rigid object.

Rosin et al. [65] present a method based on the notion of a shadow as a semi-transparent region in the image which retains a (reduced contrast) representation of the underlying surface pattern, texture or gray value. The method uses a region growing algorithm which apply a growing criterion based on a fixed attenuation of the photometric gain over the shadow region, in comparison to the reference image. The problem with this approach is that region growing algorithm cannot perform accurately in the penumbra part of the shadow due to the intensity's variations inside of the shadow region.

Xu et al. [87] detect shadow region in indoor environment. The proposed method assumes that the shadow often appears around the foreground object. A number

of techniques are used including initial change detection masks, Canny edge maps, multi-frame integration, edge matching, and conditional dilation. The method tries to detect shadow regions by extracting moving edges.

Chang et al. [8] propose a parametric Gaussian shadow model to detect and suppress pedestrian's shadow. The model makes use of several features including the orientation, mean intensity, and center position being estimated from the properties of object movements.

In the work of Hsieh et al. [26] a line-based shadow modeling process is proposed to detect moving shadows in traffic surveillance. When a vehicle moves along a lane, it will have several boundary lines parallel or vertical to this lane. Then, the lane can provide useful information for shadow elimination and do not destroy vehicle shapes. In the method first all lanes dividing lines are detected. These lanes dividing lines from video sequences are detected by vehicle's histogram. This histogram is obtained by accumulating different vehicles' positions in a training period. According to these lines and their directions, two kinds of lines are used to eliminate shadows. The first one is the lines that are parallel to the dividing lines and the second one is the lines vertical to the dividing lines.

In [56] the authors propose an outdoor shadows removal method. It is based on a spatio-temporal-reflection test and a dichromatic reflection model. The approach is divided in several sequential steps. The step one starts with the motion mask, which is computed based on mixture of Gaussians. The intensity test takes the second step. This is in charge to discard all the foreground pixels that are more brightness than their corresponding background pixels. The third step so-called blue ratio test exploits the observation that shadows pixels falling on neutral surfaces, tend to be more blueish (this step can only be performed in neutral surfaces the authors propose to define a neutral surface based on the saturation level). The fourth step so-called albedo ratio segmentation performs a segmentation based on a spatio-temporal albedo ratio. Basically, this step attempts to obtain segmented regions with uniform reflectance. Step five removes the effect of the sky illumination. The authors claim that the reflection due to sky illumination (ambient reflection) is considered as an additive component; therefore they subtract the foreground pixels from the background. The regions that belong to the foreground will result with a very different color vectors that it is the contrary of the pixels belonging to the shadow regions. The last step aim to classify those regions that could not be labeled in previous stages. This stage computes the dominant color of the unclassified-regions (body color estimation) and compare with the body colors of material surfaces pre-stored as a background model (using a supervised-learning phase).

Similar to [56] Huerta et al. [29] approach uses a multi-stage approach, however they use multiple cues: color and gradient information, together with known shadow properties. In this way, regions corresponding to potential shadows are grouped by considering the "bluish effect" and an edge partitioning. Additionally, temporal similarities between textures and spatial similarities between chrominance angle and brightness distortions are analysed for all potential shadows regions. Furthermore, geometrical shadow position is used to avoid a misclassification of moving shadows.

The challenge in these methods is not only in being able to properly analyze the segments, but also in the segmentation process. However, this is a promising

methodology since the analysis is done with the context of the shadowed area having all the shadow information.

Frame-Level

There are a very few moving cast shadow detection methods that perform at the frame-level. Normally, these few methods are not used in a background subtraction context. Hence, some of the closest methods related to the research line proposed in this thesis are briefly describe.

Liu et al. [42] detect shadows using pixel-level information, region-level information, and global-level information. Pixel-level information is extracted using GMM in the HSV color space. Local-level information is used in two ways. First, if a pixel gets a sample that is likely to be a shadow, then not only the GMM of that pixel is updated but the GMM of neighbor pixels is also updated. Second, Markov random fields are employed to represent the dependencies among neighboring pixels. For global-level information, statistical feature is exploited for whole scene over several consecutive frames.

Stauder et al. [73] use a physics-based luminance model to describe illumination changes. They assume a plain textured background and a cast shadow is determined by combining the results of change detection, static edge detection, shading change detection and penumbra detection.

A summary of the methods previously described is presented in Table 2.1. Notation in the table is as follows: Shadows Descriptors: TM, CIM, PPM and PM stand for texture models, chromaticity and intensity models, photometric physical models and projection models respectively. Supporters: Sp, Te, Ge, Tr, and SPK stand for spatial, temporal, geometrical, training and scene prior knowledge respectively. Decision: D, SP and SNP stand for deterministic, statistic parametric and statistic non-parametric respectively.

Next section reviews some of the most classical and well-know background subtraction algorithms.

2.3 Background Subtraction Methods

Background subtraction is the most commonly used technique for motion segmentation in static scenes [48, 59, 35, 5, 69]. It attempts to detect moving regions in an image by subtracting the current image with a reference background model in a pixel-by-pixel manner.

The background representation is created by averaging several images over time during an initialization period. Subsequently, pixels are classified as foreground if the difference between the input image and the background model is above a learned threshold, which calculation depends on the specific approach. Then, numerous approaches update over time the background model with new images in order to adapt it to dynamic scene changes.

Level	Methods	Shadow Descriptors				Supporters					Decision		
		TM	CIM	PPM	PM	Sp	Te	Ge	Tr	SPK	D	SP	SNP
Pixel	Cucchiara [13]		X								X		
	Horprasert[25]		X										X
	McKena [50]		X			X							X
	Kim [36]		X										X
	Siala [71]			X					X	X			X
	Brisson [47]			X		X			X				X
	Huang [28]			X		X			X				X
Region Adaptive	Xu [87]	X				X	X					X	
	Fung [18]		X			X		X					X
	Huerta [29]		X			X		X	X		X		X
	Toth [76]			X									X
	Nadimi [56]			X			X			X	X		
	Rosin [65]			X									X
	Chang [8]				X					X		X	
	Hsieh [26]				X			X	X			X	
Region Fixed	Amato [1]	X									X		X
	Yuan [91]	X					X				X		
	Grest [19]	X									X		
	Yao [89]	X									X		
	Leone [37]	X							X			X	
	Jacques [7]		X										X
	Yang [88]			X		X	X				X		X
Frame	Liu [42]		X				X					X	
	Stauder [73]			X		X	X				X		

Table 2.1

COMPARISON OF DIFFERENT MOVING SHADOW DETECTION ALGORITHMS. EACH ROW REPRESENTS AN ALGORITHM FROM THE LITERATURE, AND THE COLUMNS REPRESENT A RANGE OF CHARACTERISTICS OF SHADOW DETECTION METHODS.

There are a large number of different algorithms using this background subtraction scheme. Nonetheless, they differ in: (i) the type of cues or structures employed to build the background representation; (ii) the procedure used for detecting the foreground region; and (iii) the updating criteria of the background model.

A naive version of the background subtraction scheme is employed by Heikkila and Silven [22], which classifies an input pixel as foreground if its value is over a predefined threshold when subtracted from the background model. This approach updates the background model in order to guarantee reliable motion detection using a first order recursive filter. However, this method is extremely sensitive to changes of dynamic scenes such as gradual illumination variation or physical changes such

as *ghosts* (i.e., when an object already represented in the background model begins to move). In order to overcome these difficulties, statistical approaches have been applied [81]. These approaches make use of statistical properties of each pixel (or regions), which are updated dynamically during all the process in order to construct the background model.

Haritaoglu et al. in W^4 [20] apply background subtraction by computing for each pixel in the background model, during a training period, three values: its minimum and maximum intensity values, and the maximum intensity difference between consecutive frames. Background model pixels are updated using pixel-based and object-based updating conditions to be adaptive to illumination and physical changes in the scene. However, this approach is rather sensitive to shadows and lighting changes, since the only cue used is intensity.

Alternatively, Wren et al. in Pfunder [85] proposed a framework in which each pixel's value (in YUV space) is represented with a single Gaussian. Then, model parameters are recursively updated. However, a single Gaussian model cannot handle multiple backgrounds, such as waving trees.

Stauffer and Grimson [75, 74] addressed this issue by using a Mixture of Gaussians (MoG) to build a background color model for every pixel.

An improvement of the MoG can be found in Zivkovic et al. [92, 93], where the parameters of a MoG model are constantly updated, while selecting simultaneously the appropriate number of components for each pixel.

Elgammal et al. [15] use a non-parametric Kernel Density Estimation (KDE) to model the background. Their representation samples an intensity values for each pixel to estimate the probability of newly observed intensity values. The background model is also continuously updated to be adaptive to background changes. In addition to color-based information, their system incorporates region-based scene knowledge for matching nearby pixel locations. This approach can successfully handle the problem of small background motion such as tree branches.

Mittal et al. [52] use adaptive KDE for modeling background in motion, and implement optical flow to detect moving regions. In this way, their approach is able to manage complex background; however, the computational cost of this approach is quite high. Chen et al. [9] combine pixel- and block-based approaches to model complex background. Nevertheless, the method is very sensitive to camouflages and shadows.

Cheng et al. in [10] propose an on-line learning method which is able to work in real-time and can be implemented in GPU, which also gives similar results managing complex background. In [4] Barnich and Droogenbroeck also present a really fast method that can cope with background in motion and bootstrapping problems. The method adopts the idea of sampling the spatial neighborhood for refining the per-pixel estimation. The model updating relies on a random process that substitutes old pixel values with new ones. However, it cannot cope with camouflages and shadows. Another solution to bootstrapping problem is presented by Colombari et al. in [11], where a patch-based technique exploits both spatial and temporal consistency of the static background.

Li et al. [39] and Sheikh et al. [70] use Bayesian networks to cope with dynamic backgrounds. Li et al. uses a Bayesian framework that incorporates spectral, spatial,

and temporal features to characterize background appearance. Sheik et al. apply non-parametric density estimation to model the background as a single distribution, thus handling multi-modal spatial uncertainties. Furthermore, they also use temporal information.

The use of layers for image decomposition based on the neighboring pixels is presented in [58]. They state that such approach is robust and efficient to handle dynamic backgrounds. Maddalena et al. [44] use neural networks to overcome the same problem. An improvement of it, using self organizing maps, can be found by Lopez-Rubio et al. [43], which can adapt its color similarity measure to the characteristics of the input video.

Mahadevan et al. in [45] uses a combination of the discriminant center-surround saliency framework with the modeling power of dynamic textures to solve problems with highly dynamic backgrounds and a moving camera. However, this method is not designed for high accurate segmentation.

Toyama et al. [77] in Wallflower use a three-component system to handle many canonical anomalies for background updating. Their work processes input images at various spatial scales, namely pixel, region, and frame levels. Reasonably good foreground detection can be achieved when moving objects or strong illumination changes (for example when turning on/off the light in an indoor scene) are present. However, it fails when modeling small motion in the background or local illumination variations.

Edge cues are also used for motion segmentation. Weiss [84] also extract intrinsic images using edge cues instead of color to obtain the reflectance image. This process requires several frames to determine the reflectance edges of the scene. A reflectance edge is an edge that persists throughout the sequence. Given reflectance edges, the approach re-integrates the information to derive a reflectance image. However, the reflectance image also contains scene illuminations because this approach requires prominent changes in the scene, specifically for the position of shadows.

Jabri et al. [32] use a statistical background modeling which combines color (in RGB space) with edges. The background model is computed in two distinct parts: the color model and the edge model. On the one hand, a color model is represented by two images, the mean and the standard deviation images. On the other hand, an edge model is built by applying the Sobel edge operator to each color channel, thereby yielding horizontal and vertical difference images. Subsequently, background subtraction is performed by subtracting the color and edge channels separately using confidence maps, and then combining the results to get the foreground pixels.

Javed et al. [33] present a method that uses multiple cues, based on color and gradient information. The approach tries to handle different difficulties, such as bootstrapping (initialization with moving objects), repositioning of static background objects, ghost and quick illumination changes using three distinct levels: pixel, region and frame level, inspired from [77].

At the pixel level, two statistical models of gradients and color based on mixture of Gaussians are separately used to classify each pixel as background or foreground. At the region level, foreground pixels obtained from the color model are grouped into regions, and the gradient model is then used to eliminate regions corresponding to highlights or ghosts.

Pixel-based models are updated based on decisions made at the region level. Lastly, the frame level ignores the color based subtraction results if more than 50 percent of the results are considered foreground, thereby using only gradient subtraction results to handle global illumination changes. Nevertheless, *ghosts* cannot be eliminated if the background contains a high number of edges.

Some of the aforementioned motion detection approaches generally obtain good segmentation in indoor and outdoor scenarios, thus some of them have been used in real-time surveillance applications for years. However, most of them are susceptible to both local (such as shadows and highlights) and global illumination changes (like at dawn or dusk, and when the sun is suddenly covered by clouds).

Summary

This chapter presented a comprehensive survey of moving cast shadow algorithms. Although the area of research is relatively young, lot of works have been reported during the last years. The diversity of the information that characterizes moving shadow and the manner that such information is combined or exploited makes very difficult the task of classifying the methods. After a deep review of state-of-the-art taxonomies, this chapter reported a new moving cast shadow detection method taxonomy, where the main layer of this taxonomy describes whether methods operate with a single pixel or with a group of pixels (within a local adaptive region or a fixed region) or using the whole frame information. The second layer taxonomy describes the shadow descriptors used by the methods. The principal shadow descriptors used in the literature are: chromaticity and intensity models, texture models, photometric physical models and projection models. Some methods need to be supported by extra information. Such information can be obtained by a training phase or using some prior-knowledge of the scene or exploiting geometrical, temporal, or spatial cues. Finally the last category is based on the classification's decision of the methods. In addition, in this chapter a topology graph was designed to provide a visual interpretation of all the characteristic involved in the moving cast shadow methods. Such design is intended to facilitate the analysis of the methods by identifying the main features and characteristics used by them, and how they can be combined. Later on, this chapter summarized the most representative characteristics of several moving cast shadow methods. Finally, classical and well-known background subtraction methods were also described.

Chapter 3

Shadows Modeling and Analysis

“The greatest challenge to any thinker is stating the problem in a way that will allow a solution.”

Bertrand Russell

3.1 Introduction

A shadow is a photometric phenomenon that interferes with the scene’s illumination. However, the material properties of the surface, namely the reflectance, can not be obviously modified by shadows. This means that the reflectance of the same non-shadowed and the shadowed pixel of the background and the tested image respectively must be identical. This fact probably represents the strongest feature in order to distinguish shadowed areas from a photometric perspective.

In the literature, specially in the color theory research area, there are huge amounts of works trying to separate illuminat and chrominant components from color images [64, 51, 41, 83]. In essence the chrominant information of the pixels is associated with the reflectance.

Nevertheless, the most effective methods mentioned above, normally exploit global color information of the image that are not very useful to **completely** describe a local phenomenon such as a shadow. Despite of this fact, a significant part of moving cast shadow detection methods e.g., ([25, 50, 13, 1], among other) based on photometric shadow information make use of such separability between illuminat and chrominant.

Following the aforementioned statement, these methods will not be able to properly perform in those cases where there is color blending among objects or where the light sources involved in the scene differ in their chromaticity (chromatic shadows).

In order to describe the limitations of these methods and to individualize a solution, this chapter is organized as follows: the first part describes and discusses a photometric shadow model and the second part introduces an analysis of well-known moving cast shadow detection methods.

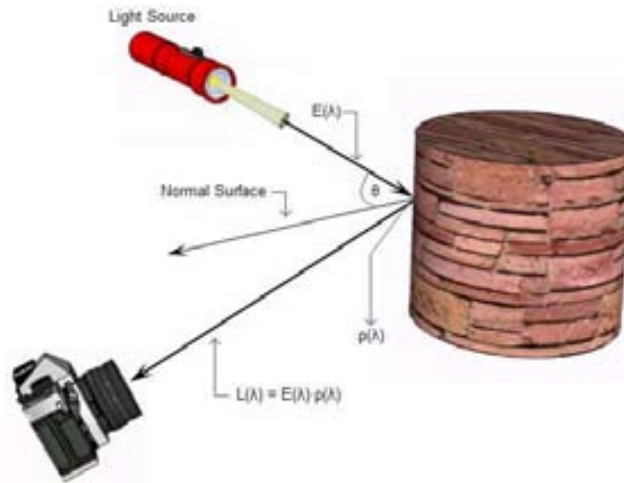


Figure 3.1: The light distribution $L(\lambda)$ that reaches to the camera as function of illuminant $E(\lambda)$ and the surface reflectance $\rho(\lambda)$.

3.2 Shadow Modeling: The Reflection Model

Broadly speaking, a color vision system tries to describe reflectance functions independent of illuminant color. It is obviously useful, since the reflectance is a property of the material that an object is made of. This fact, will be exploited by the proposed moving cast shadow detector method in order to compare the same region with and without shadows.

Let us consider the physics behind the problem of the scene illumination. The light distribution $L(\lambda)$ that reaches to the camera is a multiplication of the illuminant distribution $E(\lambda)$ and the reflectance function of a surface $\rho(\lambda)$, as follows: $L(\lambda) = E(\lambda) \cdot \rho(\lambda)$, (see Fig. 3.1).

The reflectance ρ gives the fraction of light reflected from the object. In other words, the color of an object is defined by the reflectance ρ as a function of wavelength.

A reflection model describes the interaction of light with a surface in terms of the properties of the surface and the nature of the incident light. Normally in such interaction there are a tremendous number of variables and factors which make the modeling a non trivial task. This chaotic lights-surfaces interaction can be simplified with the Phong reflection model [79] (also called Phong illumination or Phong lighting). It is a simple model that attempts to synthesize the way in which light behaves with a surface in the scene.

The Phong reflection model is a shading model commonly used in computer graphics to assign shades to each individual pixel. It can be considered as a local reflection model. For the sake of simplicity it assumes that the scene has only one source of light. However the model can be extended to multiple light sources.

In the Phong reflection model, a surface point is lit by three types of light: Ambient light C_a , Diffuse light C_b and Specular light C_c , as is illustrated in Figure 3.2.

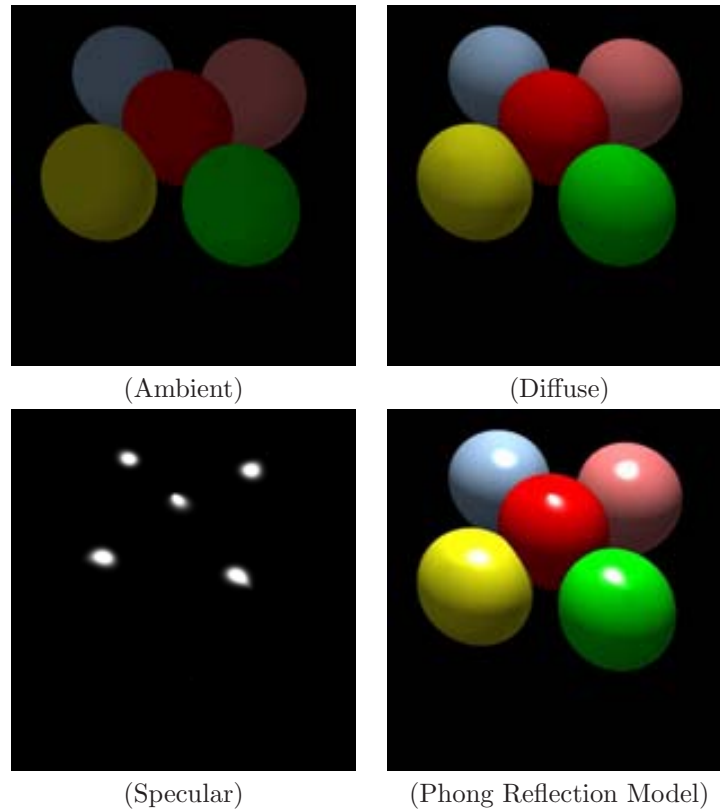


Figure 3.2: Visual illustration of the Phong model: emulated with 3DS Max.

Ambient light is the light coming to a surface as a result of multiple reflections in the scene from all directions, it is modeled as a constant term in the scene. Diffuse light is the light coming from sources such as the sun or light bulbs to the object surface diffused by the object surface to the camera. The specular light is the shiny part, that comes from the light source, reflects on the object and goes directly to the camera. The total light power coming to the camera is described by:

$$\mathbf{L} = (\mathbf{C}_a + \mathbf{C}_b \cos(\theta) + \mathbf{C}_c)\rho. \quad (3.1)$$

where θ is the angle between the light source direction and the surface normal (Fig. 3.1). The specular light is usually small and may be negligible especially when the surface is not very shiny. Therefore, this term for Lambertian surfaces is omitted, then equation (3.1) becomes:

$$\mathbf{L} = (\mathbf{C}_a + \mathbf{C}_b \cos(\theta))\rho. \quad (3.2)$$

The Phong reflection model can be used to express the photometric phenomenon caused by shadows. Cast shadows are the areas projected on a surface due to an

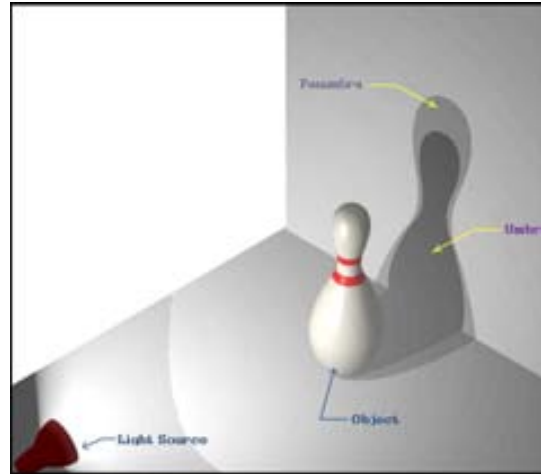


Figure 3.3: Umbra-Penumbra.

object that partially, or totally occludes the direct light source. Inside the shadow, there are two region types: umbra and penumbra, as is illustrated in Fig. 3.3.

The umbra is the darkest part of the shadow. Inside the umbra, diffuse light from the main light source is totally blocked. This region is mainly lit by the ambient light. The penumbra is the region where some part of the diffuse light is blocked, in other words this region is lit by both ambient light and part of the diffuse light.

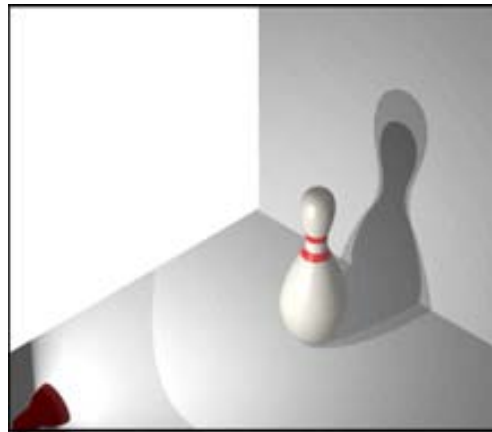
The relative size of the umbra and penumbra regions normally depend on the size and the shape of the light source as well as the distance of the source light respect to the object Fig. 3.4. The penumbra surrounds the umbra and there is always a gradual change in intensity from penumbra to umbra [82]. Inside shadows the light power of both umbra and penumbra regions can be expressed as follows:

$$\mathbf{L}_{SH} = (\mathbf{C}_a + \varsigma \mathbf{C}_b \cos(\theta))\rho. \quad (3.3)$$

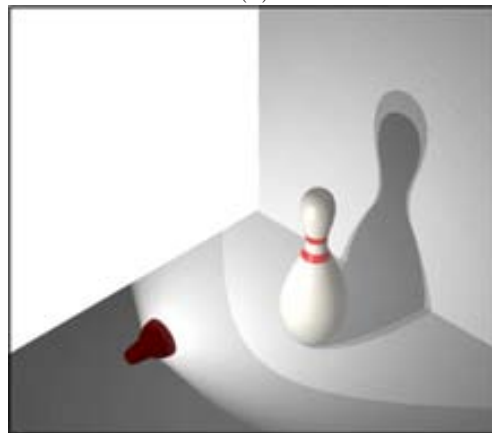
were $\varsigma \in [0, 1]$ is a shadow parameter that represents the transition inside the penumbra.

The shadows' luminance model equation. (3.3) has been directly or indirectly used in many cast shadow detection algorithms, e.g., [73, 50, 25, 13, 1], among others, but with the assumption that ambient light and diffuse light both have a similar chromaticity, which is a correct assumption for achromatic shadows.

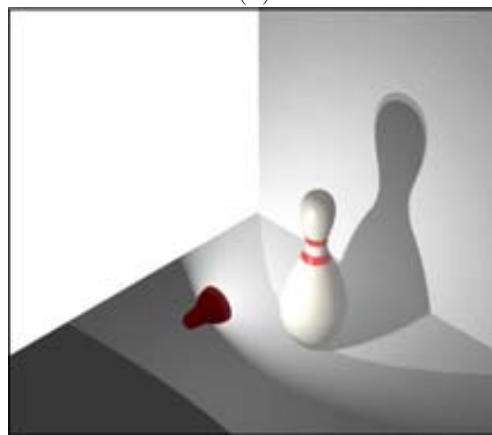
On the contrary, for chromatic shadows, the chromaticity of ambient light is different from the chromaticity of the diffuse light. Chromatic shadows can normally occur, for example in sunny days. In direct sunlight a surface is illuminated by light from the sun (diffuse light) and by light scattered from the sky (ambient light). In shadowed areas, direct sunlight is cut off and the scattered light from the sky is the main illuminat. Scattered light has less long-wavelength light than direct sunlight. Another chromatic shadows scenario can also appear when second order reflections



(a)



(b)



(c)

Figure 3.4: Umbra Penumbra spatial source variation: emulated with 3DS Max.

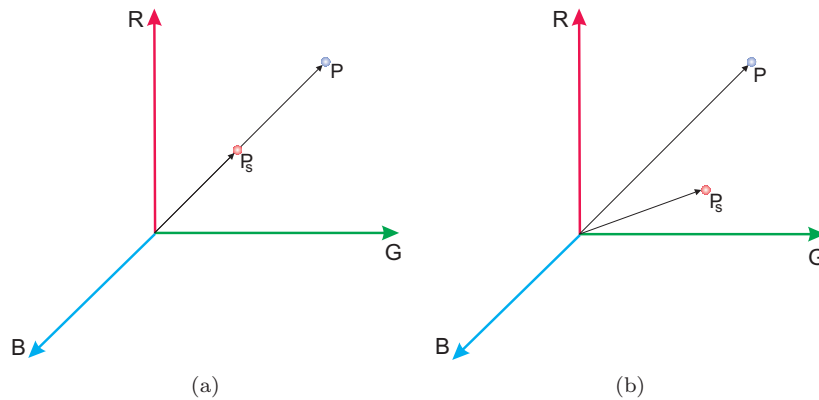


Figure 3.5: Visual illustration of achromatic and chromatic shadow effect over a pixel information in the RGB color space. In (a) the change resides only in the intensity of pixel. In (b) the change is applied in both the intensity as well as the direction of the pixel. P_s stands for a shadowed pixel and P stands for a non shadowed pixel.

cannot be neglected since they have a considerable intensity, for example when an object with non-lambertian surface casts its own shadow (e.g., red brilliant car is casts its shadow on a highway). Therefore, the illuminant's chromaticity of a shadowed pixels is significantly different from the illuminant's chromaticity of the non shadowed pixels. Hence, a variation of intensity as well as 'direction' over a shadowed pixels is induced.

Figure 3.5 illustrates the different effects caused by an achromatic shadow and a chromatic shadow over the intensity and direction of a shadowed and non shadowed pixel. Examples and analysis of these situations will be presented in the next section.

3.3 Moving Shadow Analysis

When an object casts a shadow on a surface it partially or completely blocks the surface of direct illumination from a light source, producing a change in its appearance. The measurement of this change, between a pixel affected by a shadow and the same pixel in the absence of a shadow is one of the main properties used to classify cast shadows in background-foreground segmentation algorithms.

Classical shadow suppression approaches are based on the assumption that the pixel's value of the surface under cast shadows is just a linear scaling back of its brightness component, without significant variations in its chromaticity component. This property has been modeled by several authors over different color spaces.

An analysis of the most used shadows removal methods is presented next. Such analysis includes: (i) a description of the methods; (ii) a qualitative and quantitative evaluation as well as a comparison; and (iii) a discussion of their limitations.

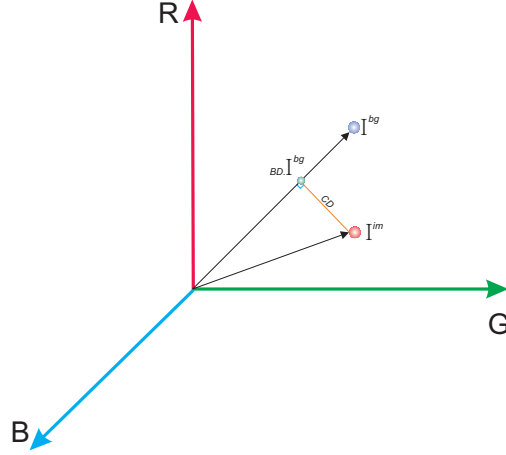


Figure 3.6: Brightness distortion and color distortion in the RGB color space. I^{bg} represents the background pixel, I^{im} is the current image pixel, BD and CD are the brightness distortion and color distortion respectively.

3.3.1 Descriptions

1. Normalized RGB Space for Brightness and Color Distortion

Horprasert et. al. [25] have oriented their work to a color model that separates the brightness and the chromaticity components in RGB color space. They introduce two terms: *brightness distortion* BD and *color distortion* CD . The BD is a scalar value that represents the quantity of change between the module of a background pixel and the module of a pixel of the current image. For shadows, BD is considered to be less than one. CD is defined as the orthogonal distance between a pixel belonging to the current image and the background pixel (see Fig. 3.6).

In order to cope with problems such as camera noise, band unbalancing, etc, the aforementioned terms are normalized by statistical parameters. The statistical values are computed over the training phase during the background modeling process. Brightness distortion and color distortion are defined as follows:

$$BD(\mathbf{x}) = \frac{\sum_{c \in \{R, G, B\}} \frac{I_c^{im}(\mathbf{x}) \cdot \mu_c^{bg}(\mathbf{x})}{(\sigma_c^{bg}(\mathbf{x}))^2}}{\sum_{c \in \{R, G, B\}} \left(\frac{\mu_c^{bg}(\mathbf{x})}{\sigma_c^{bg}(\mathbf{x})} \right)^2} \quad (3.4)$$

$$CD(\mathbf{x}) = \prod_{c \in \{R, G, B\}} \left(\frac{I_c^{im}(\mathbf{x}) - BD(\mathbf{x}) \cdot \mu_c^{bg}(\mathbf{x})}{\sigma_c^{bg}(\mathbf{x})} \right) \quad (3.5)$$

where $\mu^{bg} = (\mu_R^{bg}, \mu_G^{bg}, \mu_B^{bg})$ and $\sigma^{bg} = (\sigma_R^{bg}, \sigma_G^{bg}, \sigma_B^{bg})$ are the arithmetic means and variance respectively. In turn, \mathbf{x} denote a vector position in the discrete

frame coordinate system, $\mathbf{x} \in \{i, j\}$. $I^{im}(\mathbf{x})$ and $I^{bg}(\mathbf{x})$ are the tested image pixel and background image pixel respectively, where $I^{im}(\mathbf{x}) = \{I_R^{im}(\mathbf{x}), I_G^{im}(\mathbf{x}), I_B^{im}(\mathbf{x})\}$ and $I^{bg}(\mathbf{x}) = \{I_R^{bg}(\mathbf{x}), I_G^{bg}(\mathbf{x}), I_B^{bg}(\mathbf{x})\}$.¹

To simplify the thresholding process, BD and CD are normalized as follows:

$$BD'(\mathbf{x}) = \frac{BD(\mathbf{x}) - 1}{\sqrt{\frac{\sum_{i=0}^N (BD(i) - 1)^2}{N}}} \quad (3.6)$$

$$CD'(\mathbf{x}) = \frac{CD(\mathbf{x})}{\sqrt{\frac{\sum_{i=0}^N (CD(i))^2}{N}}} \quad (3.7)$$

where N is the number of training frames. The pixel classification mask (PCM) holds four categories and it performs following the next rules:

$$PCM(\mathbf{x}) = \begin{cases} F : CD'(\mathbf{x}) > \tau_{CD} \vee BD'(\mathbf{x}) < \tau_{Blow} \text{ else} \\ B : BD'(\mathbf{x}) < \tau_{B1} \wedge BD'(\mathbf{x}) > \tau_{B2} \text{ else} \\ S : BD'(\mathbf{x}) < 0 \text{ else} \\ H : \text{ otherwise} \end{cases} \quad (3.8)$$

where F , B , S and H stand for Foreground, Background, Shadows and Highlight respectively. τ_{CD} , τ_{Blow} , τ_{B1} and τ_{B2} are thresholds.

The criterion used in equation (3.8) is depicted in Fig. 3.7(a), a simplification in 2D of the thresholding zone is shown in Fig. 3.7(b);

2. Chromaticity Space

The chromaticity color space was designed in order to restrict the influence of the illumination [86] in the color components. This is done by using a simple components normalization.

$$\begin{aligned} X &= \frac{R}{R+G+B} \\ Y &= \frac{G}{R+G+B} \\ Z &= \frac{B}{R+G+B} = 1 - X - Y \end{aligned} \quad (3.9)$$

Since each of the normalized color is linearly dependent, the XYZ space may be represented by two normalized colors (see Fig. 3.8). Coordinate Z can be calculated for verification. The mentioned space was explored to detect shadows by McKenna et.al [50]. Unlike other approaches, in [50] they also perform the foreground-background classification in the 'chromaticity invariant color space'.

¹From now on, this will be the notation used throughout the thesis.

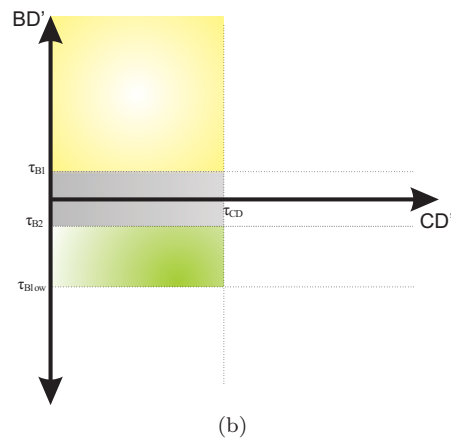
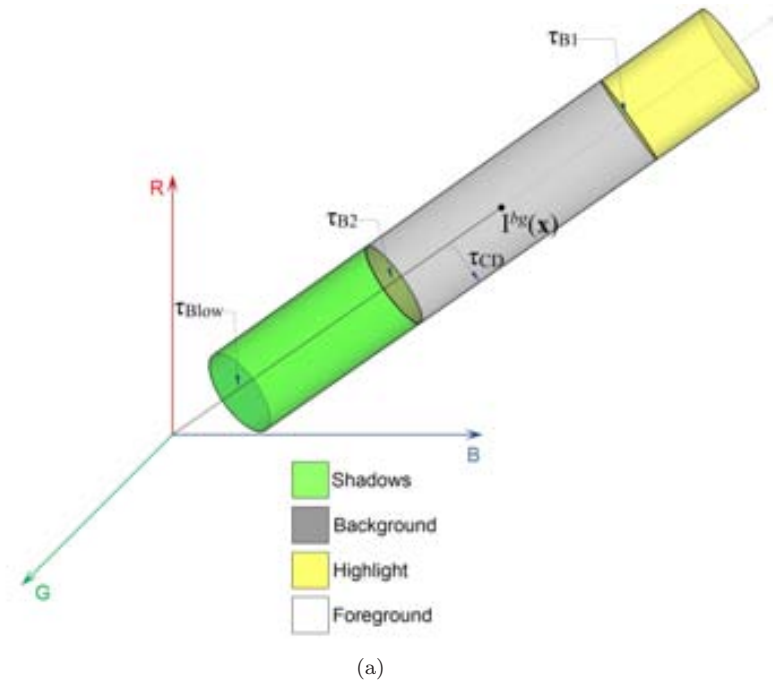


Figure 3.7: Thresholding zone is shown in (a). A simplification in 2D of the decision space is shown in (b).

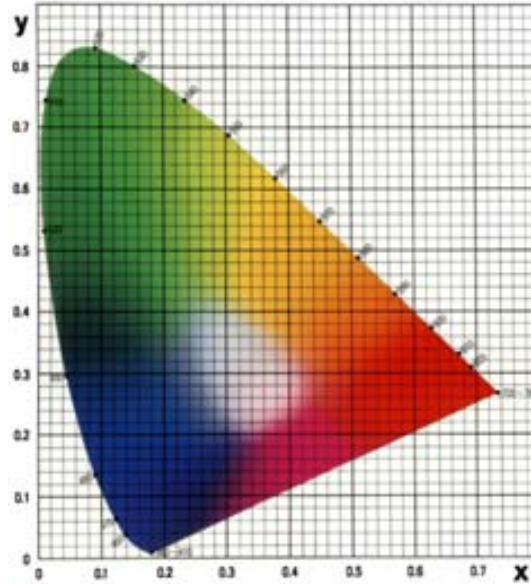


Figure 3.8: Chromaticity space.

This fact can generate problems as foreground camouflage, that is when background and foreground have a similar chroma (e.g. an agent with a dark green coat moves in front of grass or with black trousers crosses a gray concrete path). In this case foreground can be misclassified as background as well as shadow, since the brightness information is ignored. To cope with these limitations a first order image gradient information was included. Therefore, pixel chromaticity is modeled using its mean and variance and the first-order gradient of each background pixel modeled using gradient means and magnitude variance. Moving shadows are then classified if the chromaticity or gradient information supports their classification as such. However, the use of gradient affects in a negative way the classification of those shadows with strong edges, since these types of shadows can be incorrectly classified as foreground. Often these can be the umbra region of the shadows.

3. *HSV Color Space*

In the HSV color space, the brightness information of the pixel is contained in the Value component (V), while Hue component (H) is supposed to intrinsically enclose the chrominance information of the pixel. Moreover, a measure of color purity is characterized by Saturation component (S). Figure 3.9 depicts the space. The precursors to exploit HSV color space for moving cast shadows suppression were Cucchiara et al. [13]. Their statement is based on the fact that HSV color space performs similarly to the human perception of color, therefore a better description of the shadows can be achieved. The proposed method

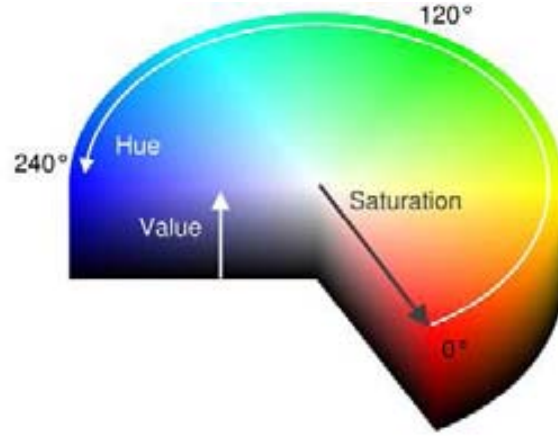


Figure 3.9: Hue Saturation Value color space.

point out that cast shadows darken the background in the V component, while the hue and saturation components change within certain limits (preserving the principal assumption of chromaticity invariant in shadowed and non shadowed pixels). A shadow mask S_M of one pixel at the image position \mathbf{x} is obtained using the following equation:

$$S_M(\mathbf{x}) = \begin{cases} 1 & \text{if } \alpha \leq \frac{I_V^{im}(\mathbf{x})}{I_V^{bg}(\mathbf{x})} \leq \beta \\ & \wedge \left(|I_S^{im}(\mathbf{x}) - I_S^{bg}(\mathbf{x})| \leq \tau_S \right) \\ & \wedge \left(|I_H^{im}(\mathbf{x}) - I_H^{bg}(\mathbf{x})| \leq \tau_H \right) \\ 0 & \text{otherwise} \end{cases} \quad (3.10)$$

were $I_H^{im}(\mathbf{x})$, $I_S^{im}(\mathbf{x})$ and $I_V^{im}(\mathbf{x})$ are the Hue, Saturation and Value components respectively of the current image. Consequently, $I_H^{bg}(\mathbf{x})$, $I_S^{bg}(\mathbf{x})$ and $I_V^{bg}(\mathbf{x})$ are the components of the background image. The parameters α , β , τ_S and τ_H are empirical thresholds.

4. RGB Polar Chromaticity Space

Initially we proposed a shadow descriptor based on chromaticity and intensity patterns [1]. We have observed that the angle between two color vectors in RGB space has been stable against illumination changes, therefore this stability can be seen like a chromaticity invariant property. These patterns are formed by similarity measurement between two sets of RGB color vectors: one belonging to the background image and the other to the current image. So a better shadow discrimination was achieved by including and combining the information of a set of pixels. This fact in turn has benefited the sensibility of background-foreground detection in foreground camouflage zones. The similarity measurements used in the method are described as follows:

- *Angular similarity measurement* $\Delta\theta$ between two color vectors $\mathbf{p}_c(\mathbf{x})$ and $\mathbf{q}_c(\mathbf{x})$ in the RGB color space $c \in \{R, G, B\}$, is defined as follows:

$$\Delta\theta(\mathbf{p}_c(\mathbf{x}), \mathbf{q}_c(\mathbf{x})) = \cos^{-1} \left(\frac{\mathbf{p}_c(\mathbf{x}) \cdot \mathbf{q}_c(\mathbf{x})}{|\mathbf{p}_c(\mathbf{x})| |\mathbf{q}_c(\mathbf{x})|} \right) \quad (3.11)$$

- *Euclidean Distance similarity measurement* ΔI between two color vectors $\mathbf{p}_c(\mathbf{x})$ and $\mathbf{q}_c(\mathbf{x})$ in the RGB color space is defined as follows

$$\Delta I(\mathbf{p}_c(\mathbf{x}), \mathbf{q}_c(\mathbf{x})) = |\mathbf{p}_c(\mathbf{x}) - \mathbf{q}_c(\mathbf{x})| \quad (3.12)$$

For each of the described similarity measurements a threshold function is associated

$$\begin{aligned} T\theta(\Delta\theta, \theta^T) &= \begin{cases} 1 & \text{if } \Delta\theta > \theta^T \\ 0 & \text{otherwise} \end{cases}, \\ TI(\Delta I, I^T) &= \begin{cases} 1 & \text{if } |\Delta I| > I^T \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (3.13)$$

where θ^T and I^T are intrinsic parameters of the threshold functions of the similarity measurements.

To describe a neighborhood similarity measurement let us first characterize the index vector $\mathbf{x} = (\mathbf{n}, \mathbf{m})^t \in \mathbf{\Omega} = \{0, 1, \dots, n, \dots, N; 0, 1, \dots, m, \dots, M\}$, which defines the position of a pixel in the image. Also we need to name the neighborhood radius vector $\mathbf{w} = (\mathbf{i}, \mathbf{j})^t \in \mathbf{W} = \{-W, \dots, 0, 1, \dots, i, \dots, W; -W, \dots, 0, 1, \dots, j, \dots, W\}$, which defines the positions of pixels that belong to the neighborhood relative to current pixel. Indeed, the domain \mathbf{W} is just a square window around a chosen pixel.

- *Angular neighborhood similarity measurement* $\eta\theta$ between two sets of color vectors in the RGB color space $\mathbf{p}_c(\mathbf{x} + \mathbf{w})$ and $\mathbf{q}_c(\mathbf{x} + \mathbf{w})$ can be written as:

$$\eta\theta(\vartheta, \theta^T) = \sum_{\mathbf{w} \in \mathbf{W}} T\theta(\Delta\theta(\vartheta), \theta^T) \quad (3.14)$$

where $T\theta$, θ^T and $\Delta\theta$ are defined in equations (3.13) and (3.11) respectively and ϑ is $(\mathbf{p}(\mathbf{x} + \mathbf{w}), \mathbf{q}(\mathbf{x} + \mathbf{w}))$.

- *Euclidean Distance neighborhood similarity measurement* μI between two sets of color vectors in the RGB color space $\mathbf{p}_c(\mathbf{x} + \mathbf{w})$ and $\mathbf{q}_c(\mathbf{x} + \mathbf{w})$ ($\mathbf{w} \in \mathbf{W}$) can be written as:

$$\mu I(\vartheta, I^T) = \sum_{\mathbf{w} \in \mathbf{W}} TI(\Delta I(\vartheta), I^T) \quad (3.15)$$

where TI , I^T and ΔI are defined in equations (3.13) and (3.12) respectively. With each of the neighborhood similarity measurements we associate a threshold function

$$\begin{aligned} T\eta\theta(\eta\theta(\vartheta), \eta^T) &= \begin{cases} 1 & \text{if } \eta\theta(\vartheta) > \eta^T \\ 0 & \text{otherwise} \end{cases}, \\ T\mu I(\mu I(\vartheta), \mu^T) &= \begin{cases} 1 & \text{if } \mu I(\vartheta) > \mu^T \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (3.16)$$

where η^T and μ^T are intrinsic parameters of the threshold functions of the neighborhood similarity measurements.

As a result, classification between $\mathbf{I}^{bg}(\mathbf{x})$ and $\mathbf{I}^{im}(\mathbf{x})$ is computed as:

$$\begin{aligned} Sh(\mathbf{x}) &= T\mu I(\mu I((\mathbf{I}^{bg}(\mathbf{x} + \mathbf{w}), \mathbf{I}^{im}(\mathbf{x} + \mathbf{w})), \gamma_I T^I(\mathbf{x})), k_F^I) \cap \\ &(|\mathbf{I}^{bg}(\mathbf{x})| > |\mathbf{I}^{im}(\mathbf{x})|) \cap \\ &(1 - T\eta\theta(\eta\theta((\mathbf{I}^{bg}(\mathbf{x} + \mathbf{w}), \mathbf{I}^{im}(\mathbf{x} + \mathbf{w})), \gamma_\theta T^\theta(\mathbf{x})), k_S^\theta)) \cap \\ &(1 - T\mu I(\mu I((\mathbf{I}^{bg}(\mathbf{x} + \mathbf{w}), \mathbf{I}^{im}(\mathbf{x} + \mathbf{w})), \gamma_S T^I(\mathbf{x})), k_S^I)), \end{aligned} \quad (3.17)$$

$$\begin{aligned} Fr(\mathbf{x}) &= T\mu I(\mu I((\mathbf{I}^{bg}(\mathbf{x} + \mathbf{w}), \mathbf{I}^{im}(\mathbf{x} + \mathbf{w})), \gamma_I T^I(\mathbf{x})), k_F^I) \cap \\ &(1 - Sh(\mathbf{x})). \end{aligned} \quad (3.18)$$

The rest of the pixels that are not classified as shadow or foreground pixels are classified as background pixels. Figure 3.10 illustrates the classification regions.

3.3.2 Comparison

This section provides a qualitative and quantitative comparison between the methods mentioned above. All the methods were faithfully implemented. Furthermore, the thresholds of the methods were manually selected to achieve the best overall performance for each video sequence. The comparison was performed over publicly available sequences.

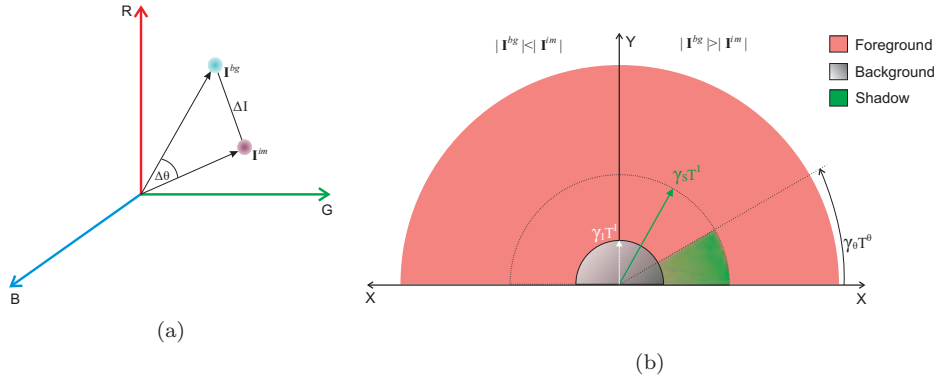


Figure 3.10: (a) Angular similarity measurement $\Delta\theta$ and Euclidean distance similarity measurement ΔI between two color vectors in the RGB color space. (b) Decision space in the Polar RGB space.

1. Quantitative Results

The quantitative comparison is based on two standard metrics for evaluating the performance of cast shadow detection algorithm introduced by Prati et. al [62]: shadow detection rate (η) and shadow discrimination rate (ξ). These two metrics are as follow:

$$\eta = \frac{TP_S}{TP_S + FN_S}; \quad \xi = \frac{\overline{TP}_F}{\overline{TP}_F + FN_F}. \quad (3.19)$$

where TP and FN stand for true positive and false negative pixels detected respect to both shadows S and foreground F . \overline{TP}_F is the number of true positive foreground pixels detected minus the number of points detected as shadows but belonging to the foreground.

The shadow detection rate η is related to the percentage of shadow pixels correctly classified, while the shadow discrimination rate ξ is concerned with foreground pixels correctly classified.

The sequences used for the evaluation were:² Hallway, Pets-2009 View 7 (Pets200 V7) and Football Match (see Table 3.1).

In order to obtain a comparison as fair as possible the thresholds for all the methods were manually selected trying to obtain an optimal response for both metrics (shadow detection rate and shadow discrimination rate) for each sequence. Although the algorithm described in [25] proposes an automatic threshold calculation, the optimal segmentation based on an automatic thresholding is far from be achieved. The comparative results are reported in Fig.3.11

²Note that for a quantitative evaluation a ground truth is necessary, the sequences as well as their ground truth are publicly accessible. More information of the sequences are available in chapter 4

		<i>Sequences</i>		
				
		Hallway	Football Match	Pets 2009 V7
Frames	Number	1800	2699	795
	Hand-labeled	13	13	16
	Size	320x240	320x240	720x576
Scene	Type	Indoor	Outdoor	Outdoor
	Background	Textured	Textured-less	Variable
	Noise	Medium	Medium	Low
Object	Class	People	People	People
	Size	Variable	Small	Variable
Shadows	Size	Variable	Small	Variable
	Visibility	Low	Low	Low
	Direction	Multiple	Multiple horizontal	Single horizontal
	Camouflage	Low	Low	Low
	Chromatic effect	Low	Low	Low

Table 3.1

DESCRIPTION OF THE SEQUENCES USED IN THE COMPARATIVE EVALUATION.

To compare the ability to distinguish shadow in the RGB polar chromaticity space the algorithm proposed in Amato et al. [1] was modified to perform at the pixel level and was also included with the name of PCS in the results (Fig. 3.11). The modification is represented by the next shadow mask:

$$Sh(\mathbf{x}) = \begin{cases} 1 & \text{if } \Delta I(I^{im}(\mathbf{x}), I^{bg}(\mathbf{x})) > TI_1 \\ & \wedge \Delta I(I^{im}(\mathbf{x}), I^{bg}(\mathbf{x})) < TI_2 \\ & \wedge \Delta\theta(I^{im}(\mathbf{x}), I^{bg}(\mathbf{x})) < T\theta \\ & \wedge |I^{im}(\mathbf{x})| < |I^{bg}(\mathbf{x})| \\ 0 & \text{otherwise} \end{cases} \quad (3.20)$$

where ΔI and $\Delta\theta$ were defined in equations (3.12) and (3.11) respectively. TI_1 , TI_2 and $T\theta$ are thresholds.

The methods used in the comparison illustrated in Fig. 3.11 were: Amato et al. [1]; PCS; Cucchiara et al. [13]; McKenna et al. [50] and Horprasert et al. [25].

In general, for these tested sequences the methods have reported good and similar performance. The better shadow discrimination capability was achieved by the region-based approach [1]. The Football Match sequence shown the biggest variability between the shadow detection rate (less than 0.7) and shadow discrimination rate (more than 0.9) for all the methods. The shadows in this sequence are very soft in terms of intensity and very small. The lowest shadow

discrimination rates were found in the Hallway sequence, because some foreground pixels are not detected due to foreground camouflage (i.e., strong similarity between background and foreground pixels exists).

2. Qualitative Results

Visual comparison of methods’s performance is provided in Fig. 3.12. The first column is referred to the Hallway sequence, the second to the Football Match sequence and the third to the Pets2009 V7 sequence. The testing image frames are shown in the first row. Next rows depict the classification results of the previously described methods. Note that the segmented images of the method reported in [1] (row (a)) are less noisy, this noise reduction is due to the region based methodology (neither pre-processing nor post-processing filter were applied).

Parts of foreground objects are misclassified as background of the Hallway sequence (see Fig. 3.12 (I)) ; this fact is reflected in the decrease of the shadow discrimination rate.

3.3.3 Limitations

The approaches described in [25] and [50] make the foreground-shadow classification by thresholding statistical information learned from previous frames. In the technique proposed in [13], the classification is based on a set of scene-constant thresholds. The algorithm developed in [1] uses a combination of scene-constant thresholds and statistical thresholding process.

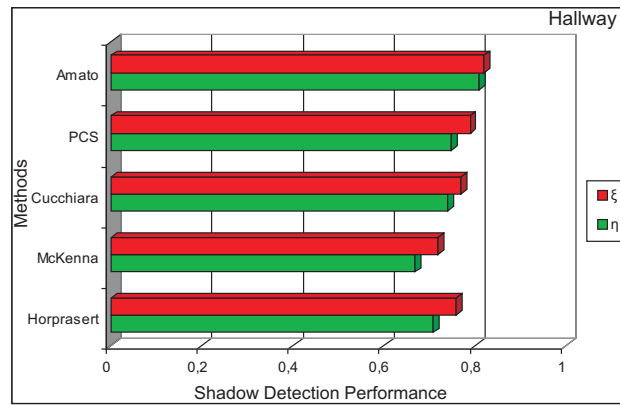
Although in [25, 50] as well as [1] the classification decision is associated with smoothing terms (statistical inference of the data), still they often require explicit tuning parameters for each new scene to achieve optimal results. Therefore, all the previously described methods suffer from non-completely automatic tuning thresholds that attempt the portability of the methods.

On the other hand, in some specific cases these approaches do not work properly. To illustrate the cases where the methods can satisfactorily be applied and the cases where the methods severely decrease their accuracy, an experimental analysis based on ground truth distribution is performed.

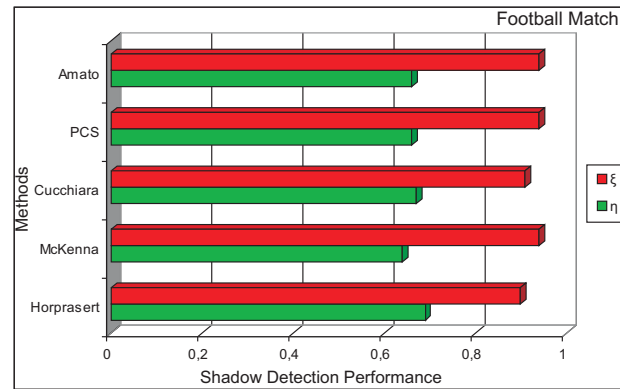
The experiment aims to visualize the shadow-foreground separability in the features space (delta color spaces). Thus, a better classification will obviously be achieved for those cases where less shadow-foreground overlapped pixels exist. To obtain the ground truth distribution of foreground and shadow pixels in the feature spaces ((a-d) for: Fig.3.13, Fig.3.14 and Fig.3.15) a hand-labeled segmentation image was used.³

Three sequences with different characteristics were chosen. In Fig. 3.13, a frame with achromatic shadows is shown. Fig. 3.14 illustrates a frame where there are

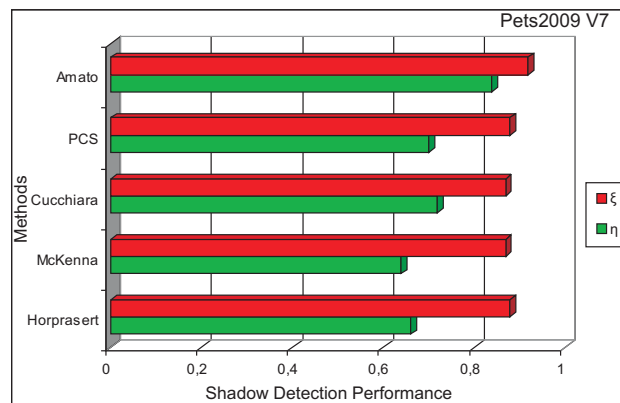
³Note that the hand-labeled segmentations are deliberately conservative in labeling pixels either foreground or shadow. This is done in order to minimize the chance of falsely labeling pixels in the ambiguous region at the boundary between object and shadow



(a)



(b)



(c)

Figure 3.11: Comparison of moving cast shadow detection methods in different sequences: (a) Hallway; (b) Football Match; and (c) Pets2009 V7.

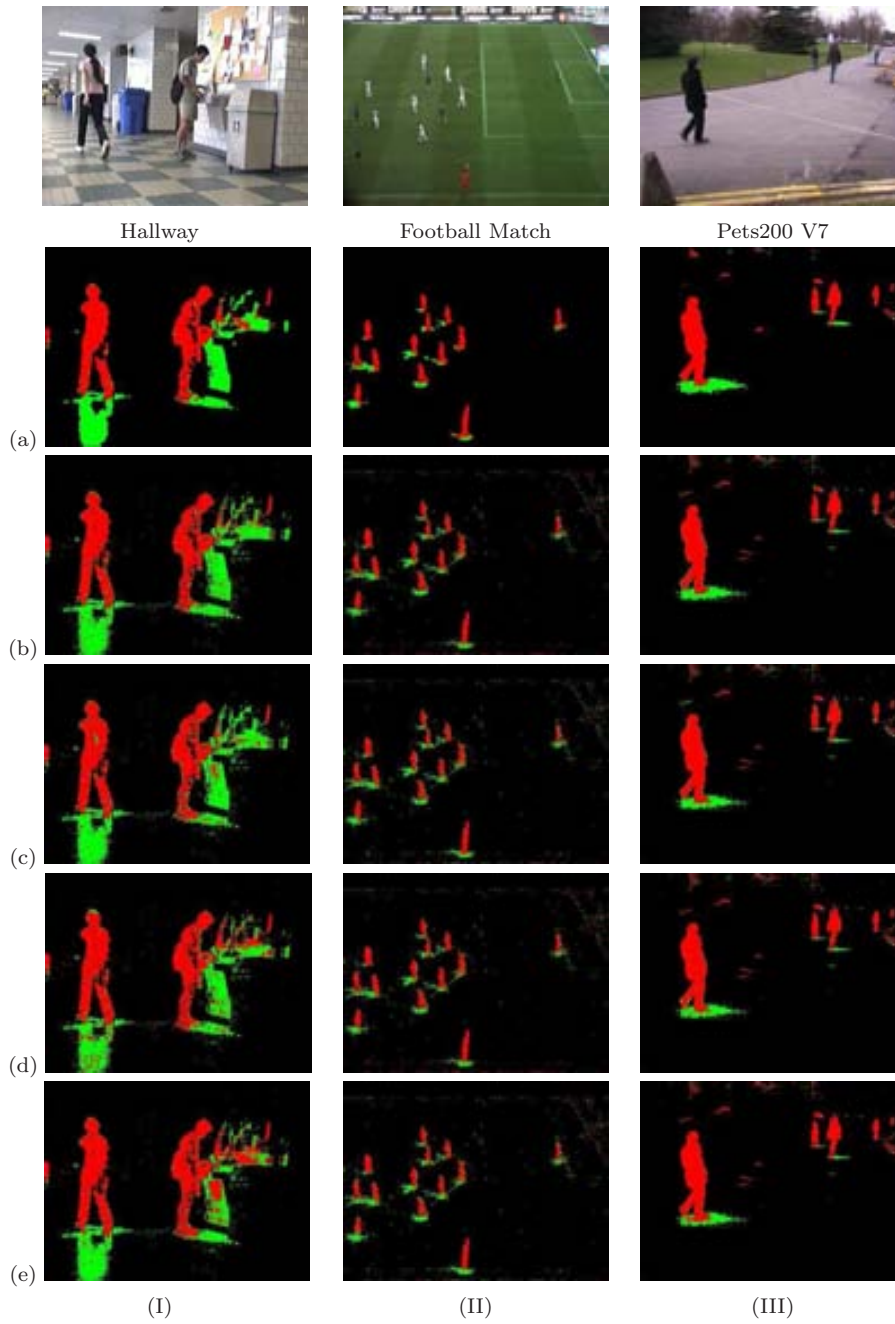


Figure 3.12: Comparison of moving cast shadow detection methods((a) Amato, (b) PCS, (c) Cucchiara, (d) McKenna and (e) Horprasert) in different sequences. (I) Hallway (frame #974), (II) Football Match (frame#004) and (III) Pets2009 V7 (frame#234). (Red color stands for pixels classified as foreground and green color stands for pixels classified as shadows).

regions affected by shadow camouflage. Finally, Fig. 3.15 shows a frame where chromatic shadows occur. The different delta color spaces are computed as follow:

- *Delta Brightness and Color Distortion Space:*

$$DCD'(\mathbf{x}) = |I^{im}(\mathbf{x})| \cdot \sin \left(\cos^{-1} \left(\frac{I^{bg}(\mathbf{x}) \cdot I^{im}(\mathbf{x})}{|I^{bg}(\mathbf{x})| |I^{im}(\mathbf{x})|} \right) \right)$$

$$DBD'(\mathbf{x}) = \frac{|I^{im}(\mathbf{x})|}{|I^{bg}(\mathbf{x})|} \cdot \cos \left(\cos^{-1} \left(\frac{I^{bg}(\mathbf{x}) \cdot I^{im}(\mathbf{x})}{|I^{bg}(\mathbf{x})| |I^{im}(\mathbf{x})|} \right) \right) - 1$$

- *Delta Chromaticity Space:*

$$X(\mathbf{x}) = \left| \frac{I_R^{bg}(\mathbf{x})}{I_R^{bg}(\mathbf{x}) + I_G^{bg}(\mathbf{x}) + I_B^{bg}(\mathbf{x})} - \frac{I_R^{im}(\mathbf{x})}{I_R^{im}(\mathbf{x}) + I_G^{im}(\mathbf{x}) + I_B^{im}(\mathbf{x})} \right|$$

$$Y(\mathbf{x}) = \left| \frac{I_G^{bg}(\mathbf{x})}{I_R^{bg}(\mathbf{x}) + I_G^{bg}(\mathbf{x}) + I_B^{bg}(\mathbf{x})} - \frac{I_G^{im}(\mathbf{x})}{I_R^{im}(\mathbf{x}) + I_G^{im}(\mathbf{x}) + I_B^{im}(\mathbf{x})} \right|$$

- *Delta HSV Space:*

$$\frac{V_{im}(\mathbf{x})}{V_{bg}(\mathbf{x})} = \frac{I_V^{im}(\mathbf{x})}{I_V^{bg}(\mathbf{x})}$$

$$|S_{im} - S_{bg}|(\mathbf{x}) = \left| I_S^{im}(\mathbf{x}) - I_S^{bg}(\mathbf{x}) \right|$$

$$|H_{im} - H_{bg}|(\mathbf{x}) = \left| I_H^{im}(\mathbf{x}) - I_H^{bg}(\mathbf{x}) \right|$$

In this space, a current frame pixel and a background pixel respectively are defined as: $I^{im}(\mathbf{x}) = \{I_H^{im}(\mathbf{x}), I_S^{im}(\mathbf{x}), I_V^{im}(\mathbf{x})\}$ and $I^{bg}(\mathbf{x}) = \{I_H^{bg}(\mathbf{x}), I_S^{bg}(\mathbf{x}), I_V^{bg}(\mathbf{x})\}$

- *Delta Polar RGB Space:*

$$\bar{X}(\mathbf{x}) = |I^{bg}(\mathbf{x}) - I^{im}(\mathbf{x})| \cos \left(\cos^{-1} \left(\frac{I^{bg}(\mathbf{x}) \cdot I^{im}(\mathbf{x})}{|I^{bg}(\mathbf{x})| |I^{im}(\mathbf{x})|} \right) \right)$$

$$\bar{Y}(\mathbf{x}) = |I^{bg}(\mathbf{x}) - I^{im}(\mathbf{x})| \sin \left(\cos^{-1} \left(\frac{I^{bg}(\mathbf{x}) \cdot I^{im}(\mathbf{x})}{|I^{bg}(\mathbf{x})| |I^{im}(\mathbf{x})|} \right) \right)$$

It can be observed in Fig. 3.13 that for the achromatic shadows foreground pixels are scattered in the feature space. Obviously the pixels' location as well as they density

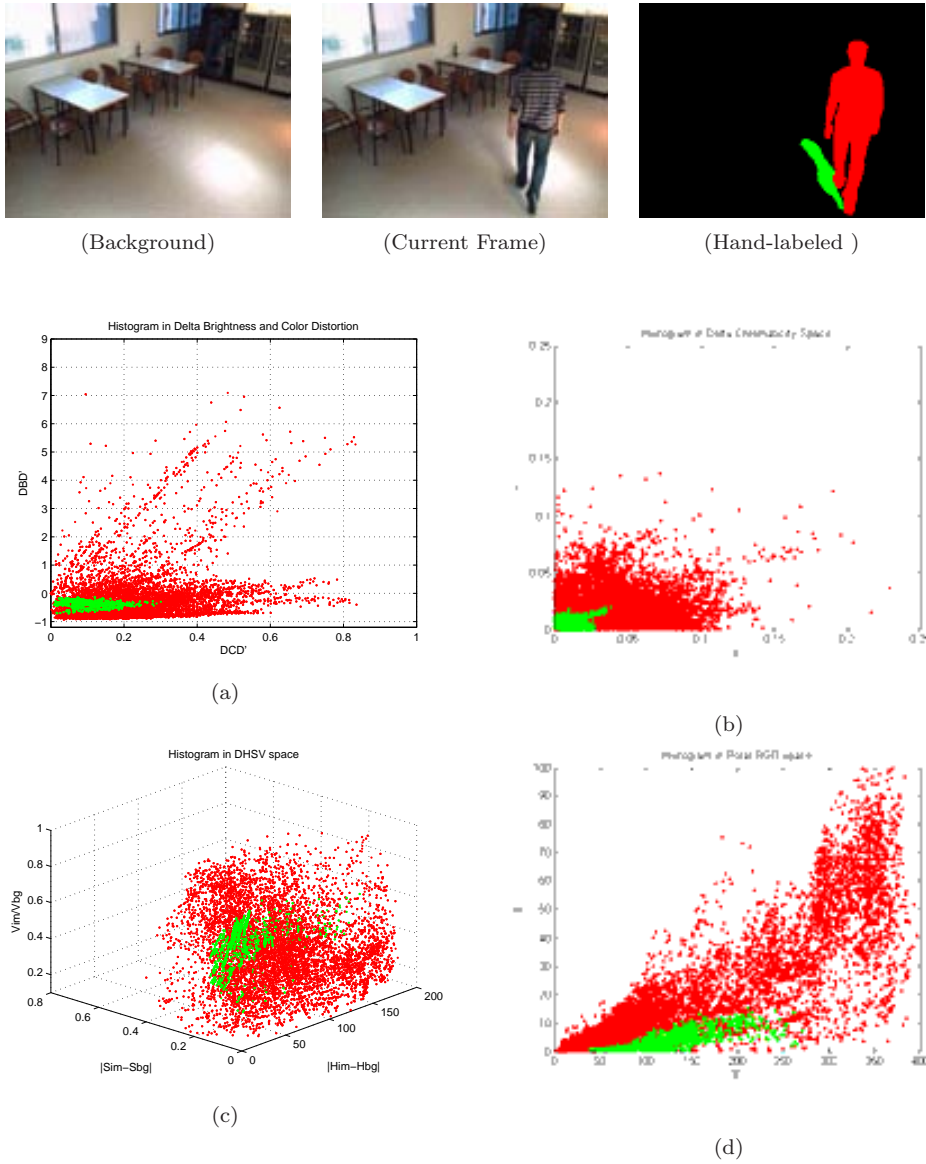


Figure 3.13: Frame with achromatic shadows. Real distribution of the foreground and shadow pixels in: (a) Delta Brightness and Color Distortion space; (b) Delta Chromaticity Space; (c) DHSV space; and (d) Polar RGB space. *Source: Hermes Indoor*

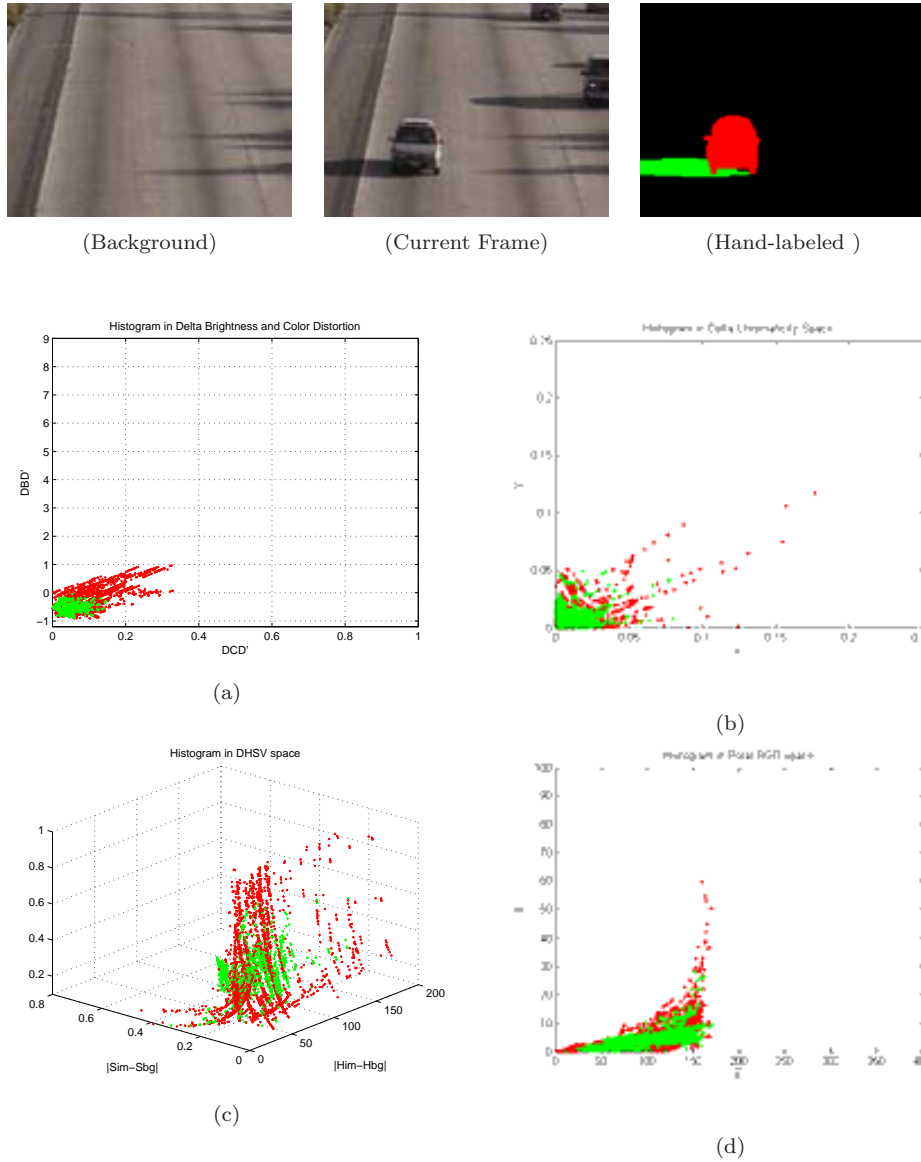


Figure 3.14: Frame with shadow camouflage. Real distribution of the foreground and shadow pixels in: (a) Delta Brightness and Color Distortion space; (b) Delta Chromaticity Space; (c) DHSV space; and (d) Polar RGB space. *Source: HWI*

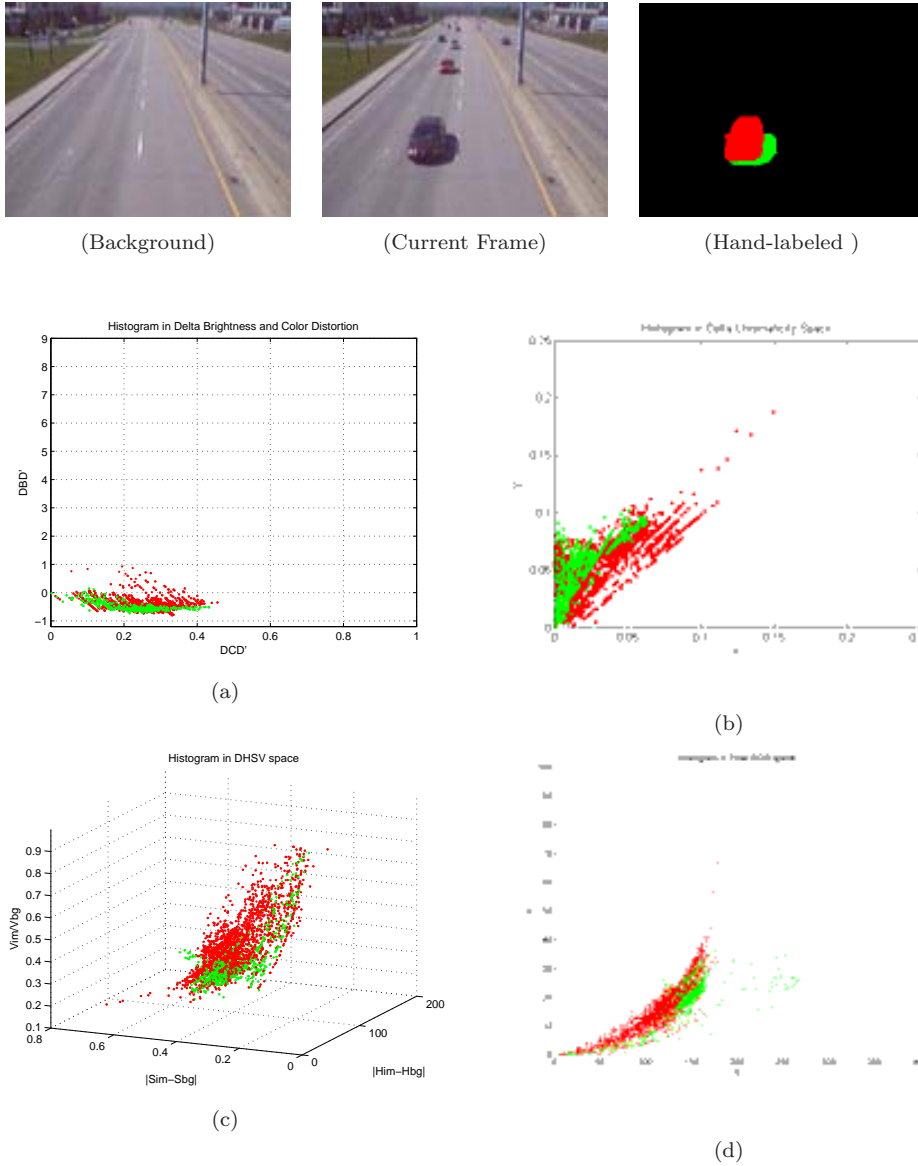


Figure 3.15: Frame with chromatic shadows. Real distribution of the foreground and shadow pixels in: (a) Delta Brightness and Color Distortion space; (b) Delta Chromaticity Space; (c) DHSV space; and (d) Polar RGB space. *Source: HWII*

depend on the relationship between the information of the background image and the information of the tested image. In the contrary, shadow pixels are concentrated in a specific location of the features space. Such location corresponds to the assumptions done, that characterize shadows, for the previously described methods. Therefore, in this case satisfactory results can be expected.

In the example done for shadow camouflage (see Fig. 3.14), the shadow pixels behave similarly to the previous example. The difference in this example relies in the foreground pixels. They are mainly located and mixed in the shadowed region. Thus, doing less reliable the classification.

Last example aims at chromatic shadows (see Fig. 3.15). In this case the problem arises since the shadow's pixels are shifted in a such a way that they are invading the typical foreground zone. Furthermore, such shift does not follow any constant pattern; it depends on the illuminations of the scene as well as the multiple reflections among the objects in the scene. Therefore, in this case satisfactory results cannot be achieved. Results of foreground-shadow segmentation of the entire frames from the above selected images are shown in Fig. 3.16.

3.4 Discussion

Moving cast shadow detection methods that only exploit chromaticity invariant property are not intrinsically prepared to cope with 'chromatic shadows'. In turn, methods that perform at the 'pixel level' highly decrease their performance in those cases where 'shadow camouflage' and 'chromatic shadows' occur, since the information of a single pixel is not enough to discriminate between shadow and foreground due to the ambiguity in their pixels' values.

In comparison with methods that perform at the pixel-level, the region-based method reported in [1] makes the detection more robust against noise and more efficient in those cases where ambiguity in the pixel's information occurs (see Fig. 3.11). However, this method also suffer from the chromatic shadow effect. Furthermore, an intrinsic difficulty of the fixed region-based methods resides in the criterion of the region's size that is used. Thus a strong dependency between the size of the region and the success of the method exists. Several factors are involved in the choice of the region's size, for example: size of the object, textural composition of the background as well as of the object, etc. Consequently, an optimal region's size is highly depending on the scene; moreover, an optimal size can change in different frames of the same scene or even the optimal region's size can change within the frame.

Although only four methods have been analyzed in this chapter, above mentioned limitations can be extended to similar approaches.

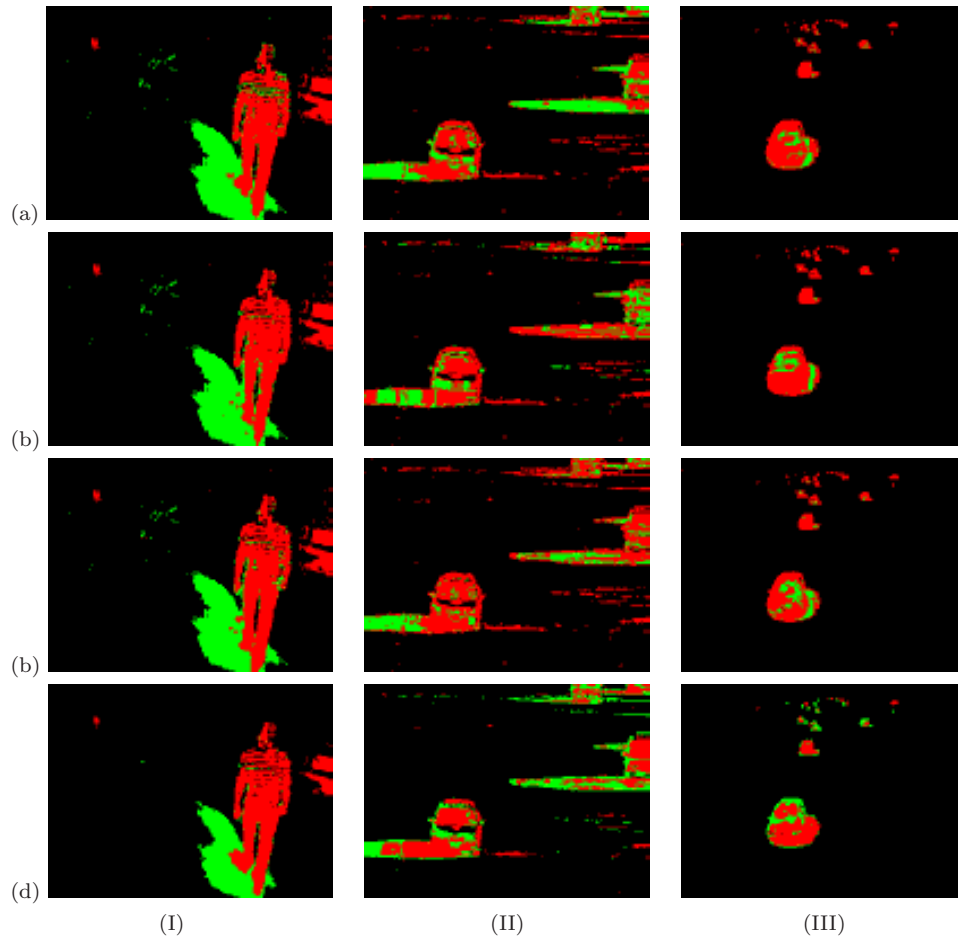


Figure 3.16: Foreground-shadow segmentation on: (I) Hermes Indoor (achromatic shadow), (II) HWI (shadow camouflage) and (III) HWII (chromatic shadow). The row (a), (b), (c) and (d) stands for: Horprasert et. al [25], Mckenna et. al [50], Cucciara et. al [13] and PCS from Amato et. al [1] respectively.

Summary

This chapter contains two parts. In the first part, a shadow luminance model is derived from the Phong reflection model. The shadow luminance model is crucial to describe: *(i)* the effect inside the shadows (umbra-penumbra); and *(ii)* the phenomenon caused by chromatic shadows. Consequently, the shadow luminance model establishes the basis of the proposed method. In the second part a case analysis of different moving cast shadow algorithms is presented. The analyzed algorithms perform at the pixel-wise and at the fixed region-wise; exploiting different color spaces.

The methods that only exploit chromaticity invariant property are intrinsically not prepared to cope with 'chromatic shadows'. In turn, the methods that perform at the 'pixel level' severely decrease their performance in those cases where 'shadow camouflage' and 'chromatic shadows' occur, since the information of a single pixel is not enough to discriminate between shadow and foreground due to the ambiguity in their pixels' values. On the other hand, methods that analyze 'fixed regions' are strongly dependent on the size of the selected region and their performance. Furthermore the textural composition of the background as well as of the object is crucial to obtain a good segmentation.

The final remarks of this chapter are summarized next:

- Chromaticity invariant is a powerful shadow descriptor but cannot be used for all sequences.
- Pixel-based methods are often easy to implement. Furthermore, they normally are suitable for real-time applications due to they low computational complexity as well as low space complexity, but are restricted to some scene types.
- Fixed Region-based methods suffer for the scene/object composition in terms of (texture and colors).

Chapter 4

Moving Cast Shadows Suppression

*“It is my ambition to say in few sentences
what others say in a whole book.”*

Friedrich Nietzsche

The framework of the proposed moving shadow detector is presented in this chapter. The goal of this approach is to detect and suppress moving cast shadow regions for most possible scenarios that occur in real video sequences, under the following conditions: (i) not make use of prior knowledge about the scene; (ii) not be restricted to specific scene structures; (iii) be automatically adaptive to different sequences; and (iv) perform in real-time.

The method is based on the assumption that in the luminance ratio space a low gradient constancy exists in all shadowed regions, due to a local color constancy effect caused by reflectance suppression. Such a color constancy effect, over a shadow region, is achieved through dividing the values of the background image by the values of the current image to form a new image in the division space. In this space, segments with low gradients correspond to all shadow regions, as opposed to foreground regions which, in most cases, exhibit higher gradients. The segments are formed by using an algorithm specifically designed for grouping shadowed pixels. The method makes use of intrinsic shadows features in order to classify each segment as shadow or foreground.

The whole description of the proposed framework is presented in this chapter. It is organized in the following sections: Section 4.1 presents the pipeline of the framework. Section 4.2 introduces the algorithm used for motion detection. Reflectance Suppression for shadowed pixel is described in section 4.3. Local constancy in shadowed areas is demonstrated in section 4.4. Section 4.5 describes a novel gradient-based segmentation algorithm. The final shadow-foreground classification is reported in section 4.6. Finally, experimental results and discussions are given in section 4.7 and section 4.8 respectively.

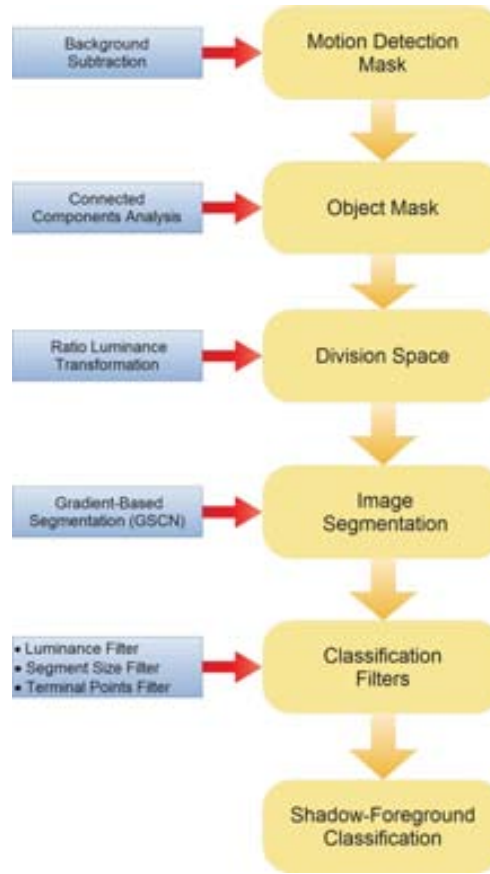


Figure 4.1: Framework Pipeline.

4.1 Framework Pipeline

A brief description of the proposed framework architecture is presented in this section. Since the method aims to detect only moving cast shadows, in this case all the image pixels must be previously classified into background pixels and motion pixels (foreground and shadow). Thus, first an initial change detection mask containing moving objects and cast shadows is obtained using a background subtraction technique. Then, the objects' mask is computed by using connected component analysis. Next, each object area is partitioned into a set of low gradient segments. To do this, luminance values of the background image are divided by the corresponding luminance values of the current frame, thus suppressing the reflectance component in shadowed areas; then the segmentation is performed by using a novel gradient-based segmentation algorithm. Finally, these segments are classified as foreground or shadow by analyzing their intrinsic parameters. A block diagram of the proposed framework is reported in Fig. 4.1.

4.2 Motion and Object Mask Formation

A binary mask of moving regions is obtained by using a background subtraction technique. A simple and common background subtraction procedure involves subtracting each new image frame from a static model of the scene. As a result a binary mask with two labels (foreground and background) is formed for each pixel in the image plane. Although the proposed shadow detection method is not limited to any specific background subtraction algorithm, in this work we choose a simplified version of the method reported in [1]. The criterion to decide this approach was based on a trade-off among robustness, accuracy and computational complexity.

Broadly speaking, a background subtraction technique can be separated in three stages, one stage deals with scene modeling, the other one with motion detection process and the last one with model updating. The scene modeling stage represents a crucial part in the background subtraction technique [77, 15, 52].

Usually a simple unimodal approach uses statistical parameters such as mean and standard deviation values [24, 49, 36]. Such statistical parameters are obtained during a training period and then they are dynamically updated. In the background modeling process the statistical values depend on both the low and high frequency changes of the camera signal. Normally, low frequency variations correspond to global illumination changes in the scene. If the standard deviations of the low and high frequency components of the signal are comparable, methods based on such statistical parameters exhibit robust discriminability. When the standard deviation of the high frequency change is significantly smaller than the low frequency change, then the background model can be improved to make the discriminative sensitivity much higher.

The work reported in [1] proposes to build a model more insensitive to low frequency changes. The main idea is to estimate only the high frequency change per each pixel value as one inter-frame interval. The general background model in this case can be explained as the subtraction between the current frame and the previous frame, which is supposed to be the background image. Two values for each pixel in the image are computed to model background changes during the training period: the maximum difference in angular and Euclidean distances between the color vectors of the consecutive image frames. The angular difference is also used because it can be considered as a photometric invariant of color measurement and it has been proved to be an acceptable cue to detect highlights.

Below the three steps of the method are depicted.

1. Background Scene Modeling.

The scene modeling step consist of two parts:

Similarity Measurements:

Two similarity measurements are used to compare a background image with the current frame. They are: *angular similarity measurement* $\Delta\theta$ and *Euclidean distance similarity measurement* ΔI . (see equations (3.11) and (3.12)).

Scene Modeling:

The background model (BG) will be represented with two classes of components namely running components (RC) and training components (TC). The RC is

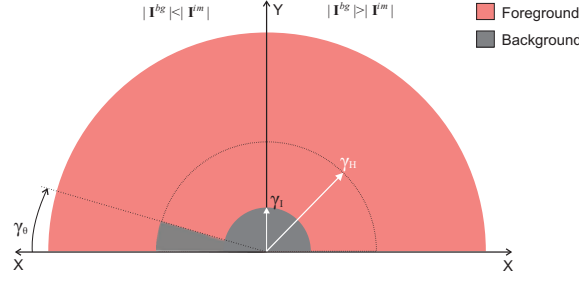


Figure 4.2: Difference in angle and magnitude in 2D "Polar RGB Color Space". The axes are computed as: $x = \Delta I \cdot \cos(\Delta\theta)$ and $y = \Delta I \cdot \sin(\Delta\theta)$

a color vector in the RGB color space, which is only updated in the running process. The TC is a set of fixed threshold values obtained during the training. The background model is represented by:

$$BG(\mathbf{x}) = \{ \{ \mathbf{I}^{bg}(\mathbf{x}) \}, \{ T^\theta(\mathbf{x}), T^I(\mathbf{x}) \} \} \quad (4.1)$$

where $T^\theta(\mathbf{x})$ is the maximum of the chromaticity variation and $T^I(\mathbf{x})$ is the maximum of the intensity variation computed during the training period as follows:

$$\begin{aligned} T^\theta(\mathbf{x}) &= \max_{f \in \{1, 2, \dots, F\}} \{ \Delta\theta(\mathbf{I}^{f-1}(\mathbf{x}), \mathbf{I}^f(\mathbf{x})) \}, \\ T^I(\mathbf{x}) &= \max_{f \in \{1, 2, \dots, F\}} \{ \Delta I(\mathbf{I}^{f-1}(\mathbf{x}), \mathbf{I}^f(\mathbf{x})) \}, \end{aligned} \quad (4.2)$$

where F is the number of frames in the training period, and RC is initialized as follows:

$$RC = \{ \mathbf{I}^F(\mathbf{x}) \}$$

where \mathbf{I}^F represents the last image frame of the training period.

2. Classification Process.

All those pixels that satisfy the condition reported in equation (4.3) will be classified as background, the rest will be considered as foreground. Note that the second, third and fourth rows of the condition (equation (4.3)) represent the criterion of the highlight regions. Figure 4.2 illustrates the classification regions.

$$\begin{aligned} &(\Delta I(\mathbf{I}^{bg}(\mathbf{x}), \mathbf{I}^{im}(\mathbf{x})) < \gamma_I(\mathbf{x})) \vee \\ &((\Delta I(\mathbf{I}^{bg}(\mathbf{x}), \mathbf{I}^{im}(\mathbf{x})) < \gamma_H(\mathbf{x})) \wedge \\ &(\Delta I(\mathbf{I}^{bg}(\mathbf{x}), \mathbf{I}^{im}(\mathbf{x})) < \gamma_\theta(\mathbf{x})) \wedge \\ &(|\mathbf{I}^{bg}(\mathbf{x})| < |\mathbf{I}^{im}(\mathbf{x})|)), \end{aligned} \quad (4.3)$$



Figure 4.3: (a) Background image, (b) Current frame, (c) Binary moving pixel mask.

where \mathbf{I}^{bg} and \mathbf{I}^{im} represent the pixel values of the background and the current frame respectively; $\gamma_I(\mathbf{x}) = (T^I(\mathbf{x}) + kb_I)$, $\gamma_\theta(\mathbf{x}) = (T^\theta(\mathbf{x}) + kb_\theta)$, (T^I and T^θ were defined in equation (4.2)). kb_I, kb_θ and γ_H are thresholds.

3. Model Updating.

In order to maintain the stability of the background model through the time, the model needs to be dynamically updated. As it was explained before, only the RC is updated in running phase. The update process is done at every frame, but only for those pixels that have been classified as background. The model is updated as follows:

$$\mathbf{I}^{bg}(\mathbf{x}) = \beta \cdot \mathbf{I}^{bg-1}(\mathbf{x}) + (1 - \beta) \cdot \mathbf{I}^{im}(\mathbf{x}). \quad (4.4)$$

where ($0 < \beta < 1$) is an update rate.

An indicator function $M(\mathbf{x})$ of moving pixels is a simple binary mask, where:

$$M(\mathbf{x}) = \begin{cases} 1 & \text{foreground} \\ 0 & \text{background} \end{cases} \quad (4.5)$$

The result of computing a moving mask image M between a background image and a frame of test sequence is shown in Fig.4.3.

Despite of the fact that the components defined in the model of the background (RC and TC) were enough to achieve the motion detection mask, the model will also include one term related with shadow components (SC). These SC are composed by the mean and the variance of each pixel frame. SC are computed in the training phase and updated in running stage, SC will be used to compute some shadow parameters. More details related to the use of SC are given in section 4.7. In equation (4.6) the new model of the background is shown. The update process of the mean value and the variance are respectively reported in equations (4.7) and (4.8).

$$BG(\mathbf{x}) = \{ \{ \mathbf{I}^{bg}(\mathbf{x}) \}, \{ T^\theta(\mathbf{x}), T^I(\mathbf{x}) \}, \{ \mu^{bg}(\mathbf{x}), (\sigma^{bg}(\mathbf{x}))^2 \} \}, \quad (4.6)$$

$$\mu^{bg}(\mathbf{x}) = \beta \cdot \mu^{bg-1}(\mathbf{x}) + (1 - \beta) \mathbf{I}^{im}(\mathbf{x}), \quad (4.7)$$

$$(\sigma^{bg}(\mathbf{x}))^2 = \beta \cdot ((\sigma^{bg-1}(\mathbf{x}))^2 + (\mu^{bg}(\mathbf{x}) - \mu^{bg-1}(\mathbf{x}))^2) + (1 - \beta) (\mathbf{I}^{bg}(\mathbf{x}) - \mu^{bg}(\mathbf{x}))^2. \quad (4.8)$$

The parameters (γ_H and β) are often scene-dependent, while (kb_I and kb_θ) are more independent since they behave as an offset of pre-estimated data. However, a precise selection of all these parameters is not crucial in the performance of the proposed detector. Normally, in a background subtraction context a accurate thresholds setting plays an important role in order to enhance the motion segmentation mask. Nevertheless, for the proposed region-based approach the motion segmentation mask is transformed in the motion object mask using a connected component analysis [12], which indirectly copes with the outliers of the motion segmentation (working as a post-processing filter). To illustrate this fact, Fig. 4.4 shows three image frames from different sequences, where the parameters were severely altered. In Fig. 4.4(a) β was manipulated in such a way to generate the motion detection more unstable, since this sequence suffer from illumination changes problems. Results of motion detection mask and object mask are reported in Fig. 4.4(d) and (g) respectively. In the case of Fig. 4.4(b) and (c) the thresholds were manipulated in order to obtain low sensibility detection (Fig. 4.4(b)) and high sensibility detection (Fig. 4.4(c)). It is observed that in Fig. 4.4(e) many true foreground pixels are not correctly classified, while in Fig. 4.4(f) the motion detection became very noisy. However, the object masks for all the examples (see Fig. 4.4(g-i)) are properly formed.

To obtain the objects' masks starting with the binary motion mask where naturally and obviously is formed by several connected regions, the following steps are needed: (i) assign a unique identifier for each connected region; (ii) discard small (relative to the image size) group of moving pixels; and (iii) merge an empty region with the area that is surrounded by. Using depth-first search algorithm, this can handily be computed.

Therefore, the initial binary mask is transformed into a set of independent regions $\Phi = \{o_1, o_2, \dots, o_k\}$, where K is the number of motion regions in the current frame. Figure 4.5 depicts the process and Fig. 4.6 reports several object mask formation examples.

4.3 Reflectance Suppression

An image obtained from a scene by a standard RGB camera, assuming Lambertian reflectance can be analyzed using a simple luminance model [73]:

$$\mathbf{L}(\mathbf{x}) = \mathbf{E}(\mathbf{x}) \boldsymbol{\rho}(\mathbf{x}).$$

where $\mathbf{L}(\mathbf{x}) = [L_R(\mathbf{x}), L_G(\mathbf{x}), L_B(\mathbf{x})]^T$ is the luminance vector of the RGB color space in the image plane $\mathbf{x} \in \mathbf{X}$, $\mathbf{E}(\mathbf{x}) = [E_R(\mathbf{x}), E_G(\mathbf{x}), E_B(\mathbf{x})]^T$ is the irradiance

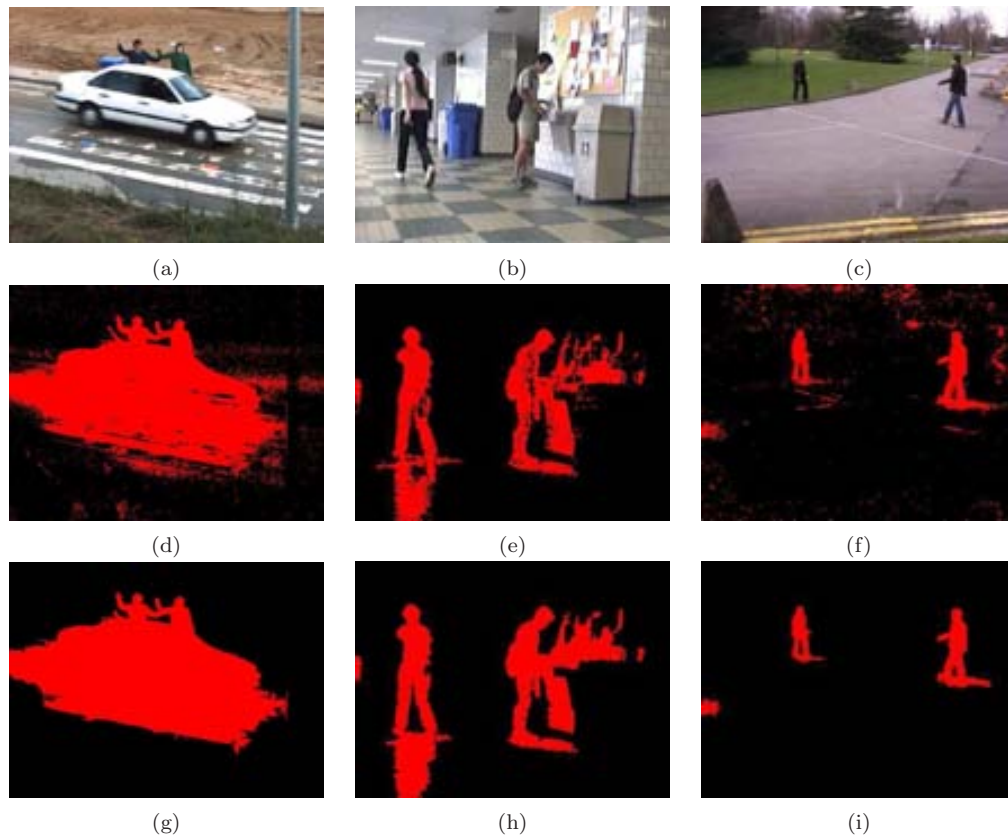


Figure 4.4: Thresholds evaluations. (a-c) test images; (d-f) motion detection masks; (g-i) objects masks.

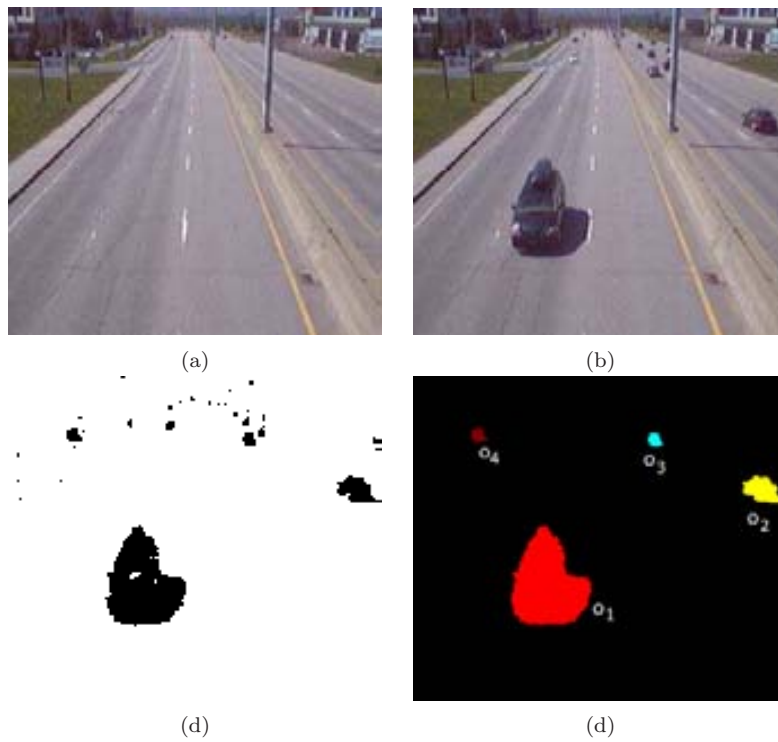


Figure 4.5: (a) Background image; (b) current frame; (c) binary motion pixels mask and (d) object mask.

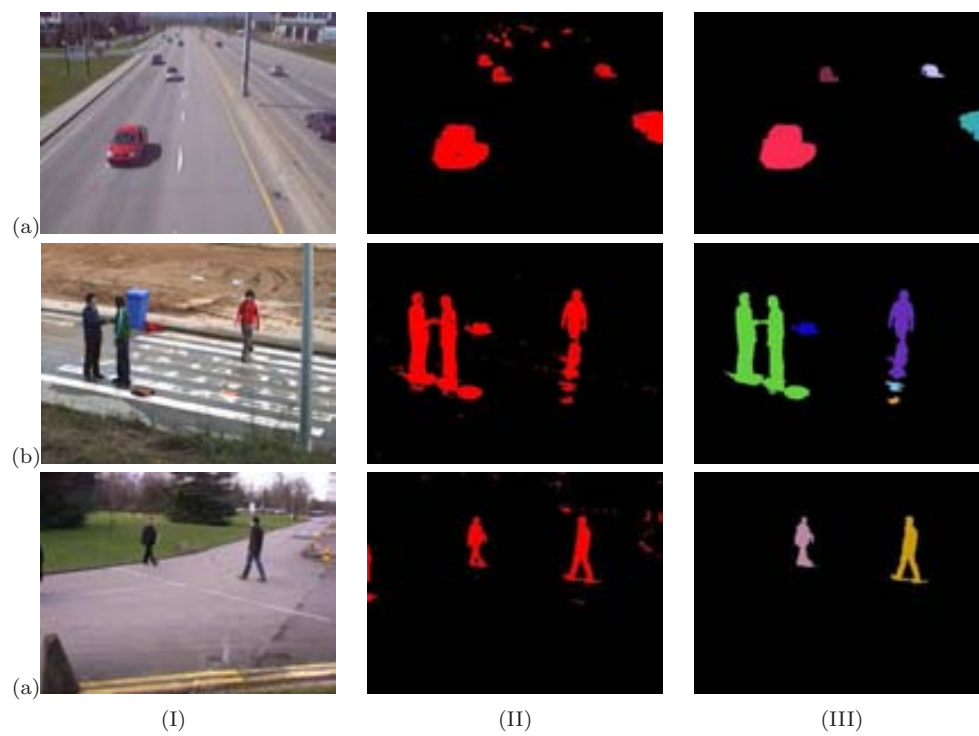


Figure 4.6: Object Mask Examples. (a) HWIII #198; (b) HERMES Outdoor Cam3 #1097; and (c) Pets 2009 V7 #9. Columns (I), (II) and (III) represent current image, motion mask and objects mask respectively.

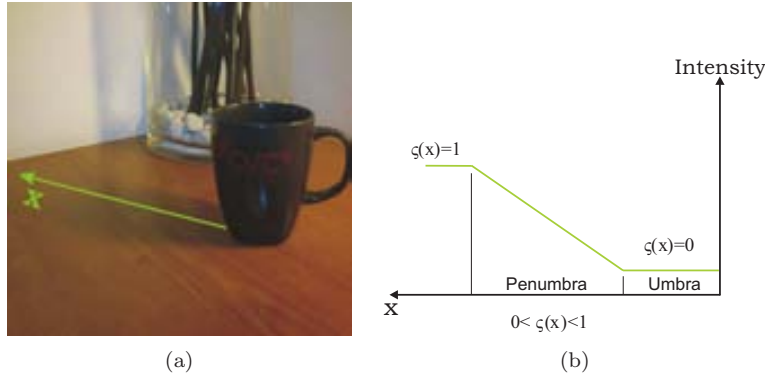


Figure 4.7: Umbra-penumbra transition in terms of intensity and spatial distance from the object. (a) test image; (b) illustration of the shadow parameter.

vector of the input signal, and $\boldsymbol{\rho}(\mathbf{x}) = [\rho_R(\mathbf{x}), \rho_G(\mathbf{x}), \rho_B(\mathbf{x})]^T$ is the reflectance vector of the object surface reflected at the pixel \mathbf{x} . Though vectors \mathbf{E} and $\boldsymbol{\rho}$ should be properly expressed by diagonal matrices, notation throughout the text is simplified. The product of two vectors indicates simple component-wise multiplication $\mathbf{a}\mathbf{b} = a_i b_i$. Division is similarly indicated.

The irradiance component of the input signal for one light source in shadows areas can now be expressed (see chapter 3.2) as:

$$\mathbf{E}(\mathbf{x}) = \mathbf{C}_a + \mathbf{C}_b \cos(\theta(\mathbf{x}))\zeta(\mathbf{x}). \quad (4.9)$$

where \mathbf{C}_a , \mathbf{C}_b and $\theta(\mathbf{x})$ are the intensity of ambient light, the intensity of the light source and the angle between the light source direction and a surface normal respectively; $\zeta(\mathbf{x}) \in [0, 1]$ is a shadow parameter that represents the transition inside the penumbra, which depends on the light source and scene geometry. Generally, it is characterized by slow spatial variation [82].

When $\zeta(\mathbf{x})$ is equal to 0, the quantity of light reflected at pixel \mathbf{x} is from ambient light alone and it belongs to the umbra region. For $0 < \zeta(\mathbf{x}) < 1$, the pixel is located in the penumbra and when $\zeta(\mathbf{x}) = 1$ the pixel is outside of the shadow region (see Fig. 4.7).

The appearance of an arbitrary pixel \mathbf{x} in an image sequence will vary according to illumination conditions and the configuration of objects that may cast shadows. Let $\mathbf{L}^{im}(\mathbf{x})$ be the pixels belonging to cast shadows in the current frame and $\mathbf{L}^{bg}(\mathbf{x})$ those that do not. The luminance ratio of these pixels can be written as:

$$\mathbf{D}(\mathbf{x}) = \frac{\mathbf{L}^{bg}(\mathbf{x})}{\mathbf{L}^{im}(\mathbf{x})} = \frac{\mathbf{E}^{bg}(\mathbf{x})\boldsymbol{\rho}^{bg}(\mathbf{x})}{\mathbf{E}^{im}(\mathbf{x})\boldsymbol{\rho}^{im}(\mathbf{x})}. \quad (4.10)$$

If the point \mathbf{x} belongs to the shadow region R^{sh} of the current image $\mathbf{L}^{im}(\mathbf{x})$ then the two reflectances in equation (4.10) are equal because the reflectance $\boldsymbol{\rho}(\mathbf{x})$ of the projected surface point \mathbf{x} does not changes with time. Therefore, the result of the luminance ratio \mathbf{D} is reduced to:

$$\mathbf{D}(\mathbf{x}) = \frac{\mathbf{E}^{bg}(\mathbf{x})}{\mathbf{E}^{im}(\mathbf{x})}, \forall \mathbf{x} \in R^{sh}. \quad (4.11)$$

After substituting equation (4.9) in (4.11) \mathbf{D} becomes:

$$\mathbf{D}(\mathbf{x}) = \frac{\mathbf{C}_a^{bg} + \mathbf{C}_b^{bg} \cos(\theta(\mathbf{x}))}{\mathbf{C}_a^{im} + \mathbf{C}_b^{im} \cos(\theta(\mathbf{x}))\zeta(\mathbf{x})}. \quad (4.12)$$

Let $\Delta\mathbf{x}$ be the distance between two neighboring pixels in the image plane. The difference between two luminance ratios can be written as:

$$\begin{aligned} \mathbf{D}(\mathbf{x}) - \mathbf{D}(\mathbf{x} + \Delta\mathbf{x}) &= \frac{\mathbf{C}_a^{bg} + \mathbf{C}_b^{bg} \cos(\theta(\mathbf{x}))}{\mathbf{C}_a^{im} + \mathbf{C}_b^{im} \cos(\theta(\mathbf{x}))\zeta(\mathbf{x})} - \\ &\frac{\mathbf{C}_a^{bg} + \mathbf{C}_b^{bg} \cos(\theta(\mathbf{x} + \Delta\mathbf{x}))}{\mathbf{C}_a^{im} + \mathbf{C}_b^{im} \cos(\theta(\mathbf{x} + \Delta\mathbf{x}))\zeta(\mathbf{x} + \Delta\mathbf{x})}. \end{aligned} \quad (4.13)$$

Assuming that the scale factor ζ and the angle θ are slowly varying functions: $\zeta(\mathbf{x}) \approx \zeta(\mathbf{x} + \Delta\mathbf{x})$ and $\theta(\mathbf{x}) \approx \theta(\mathbf{x} + \Delta\mathbf{x})$, we obtain:

$$\mathbf{D}(\mathbf{x}) - \mathbf{D}(\mathbf{x} + \Delta\mathbf{x}) \approx 0. \quad (4.14)$$

which means that a local constancy exists for any pair of pixels belonging to the shadow region. In contrast, local constancy in equation (4.14) does not hold for foreground pixels because of the inequality of the reflectance components in equation (4.10).

The local color constancy condition in equation (4.14) is derived using a single light source model. However, the used assumption will also hold when $\mathbf{E}(\mathbf{x})$ in equation (4.9) is formed by a linear combination of multiple light sources. The commonly used irradiance model of equation (4.9) from [73] assumes that the intensity of ambient light \mathbf{C}_a is constant and that the intensity of the light source \mathbf{C}_b is proportional to $(\frac{1}{r^2})$, where r is the distance between the object and light source [72]. These assumptions are valid for a broad range of imaging conditions, and for deriving the proposed local color constancy criterion it is reasonable to use a model where both \mathbf{C}_a and \mathbf{C}_b are constants. This simplified lighting model has been used in the derivation of several other shadow suppression models [73, 76, 88].

Using the local constancy effect that exists in shadow regions, the proposed algorithm distinguishes between shadows and foreground regions. Note that there is not any assumption with the chromaticity of the light sources, thus local color constancy will exist for achromatic shadows as well as chromatic shadows.

4.4 Shadow Region with Local Color Constancy

The Luminance ratio image is firstly calculated in order to detect regions with a local color constancy. The luminance ratio for a single pixel is written as:

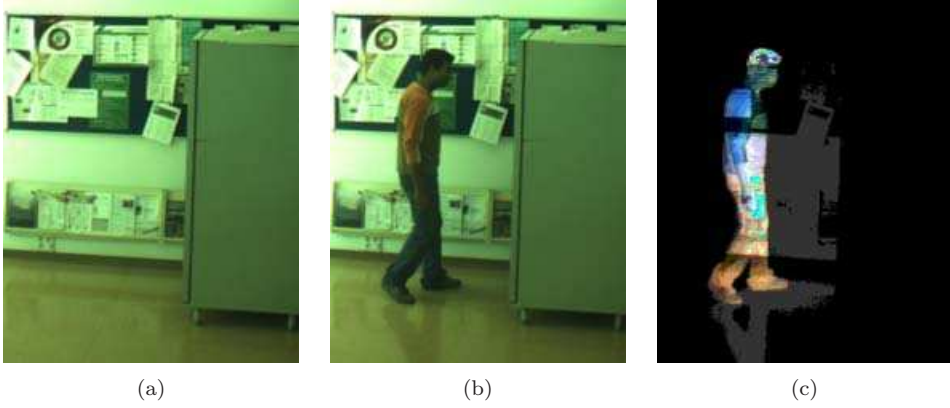


Figure 4.8: (a) Background image; (b) current image; (c) luminance ratio image in the RGB color space.

$$\mathbf{D}(\mathbf{x}) = \frac{\mathbf{L}^{bg}(\mathbf{x}) + \nu}{\mathbf{L}^{im}(\mathbf{x}) + \nu}, \quad (4.15)$$

where ν is a quantization constant, which is chosen to be unity for the standard eight bit input signal. It is very important to note that for our shadow detection algorithm it is more essential to make fine segmentation in shadow-like pixels than in foreground pixels (shadow-like pixels are referred to those current image pixels with lower luminance than their corresponding background image pixels). When the background image is divided by the current image, the luminance ratio image \mathbf{D} is segmented into two types of regions: (i) foreground regions where $2^{-8} \leq \mathbf{D}(\mathbf{x}) \leq 1$ and (ii) shadow like regions where $1 \leq \mathbf{D}(\mathbf{x}) \leq 2^8$. It is easy to see that in the shadow-like regions, measurement is more precise.

An example of the luminance ratio image in the RGB color space is illustrated in Fig. 4.8(c), where Fig. 4.8(a) shows the background model and Fig. 4.8(b) is a frame of the test sequence.

Let us analyze the values of the luminance ratio \mathbf{D} inside each motion areas. As we explained in equation (4.11), the value of the function $\mathbf{D}(\mathbf{x})$ in the shadowed area only depends on the irradiance ratios between the background and current image, and form regions (usually one shadows region per motion segment) that are characterized by smooth spatial changes in equation (4.14), which we call local color constancy. We stress that color constancy does not assume, in general, color constancy in a full shadow region and a value of the luminance ratio $\mathbf{D}(\mathbf{x}_1)$ can be considerably different from a value $\mathbf{D}(\mathbf{x}_2)$ if the distance between two pixels $|\mathbf{x}_1 - \mathbf{x}_2|$ is significant.

In contrast, in the foreground areas the value of the function $\mathbf{D}(\mathbf{x})$ depends both on the irradiance ratios and on reflectance ratios (equation (4.10)) and in general, form small regions with local color constancy. In this way, local color constancy detection allows us to distinguish shadow from foreground.

In some cases, when this assumption in the foreground areas does not hold (i.e.

poor texture or texture-less in both foreground and background), we use additional features to solve the problem (external terminal points as will be explained later).

Thus, the main goal of our algorithm is to detect local color constancy regions, which is a classical segmentation task. Segmentation problems can usually be solved by local neighborhood analysis. Local neighborhood analysis is a widely used technique in image processing. Most of them utilize sliding windows with a fixed shape and size or with a locally adaptive shape.

Other approaches work with neighborhoods that are the result of an initial segmentation of the entire region of interest, for example watershed [19] or mean shift segmentation [20]. But in our case, to detect shadow segments we have to overcome the problem of gradual change. Therefore, we apply a gradient-based segmentation technique. More precisely, our algorithm forms a set of gradient space connected neighborhoods (GSCN) or a set of nonoverlapping segments. In the next section we explain this technique.

4.5 Gradient Space Connected Neighborhoods Segmentation

A gradient-space connected neighborhood is defined as a set of pixels in an image in which any two pixels are gradient connected. That is, there exists a (four- or eight-connected) path between any pair of pixels in the neighborhood. All pixels of the path satisfy the following condition: $|D(\mathbf{p}) - D(\mathbf{q})| \leq \partial$, where ∂ is a given threshold, $\{D(\mathbf{p}), D(\mathbf{q})\}$ are values of any pair of adjacent pixels $\{\mathbf{p}, \mathbf{q}\} \in \mathbf{X}$ on the path.

Figure 4.9(a) illustrates two paths between pixels $(0, 2) - (2, 0)$ and $(0, 5) - (7, 0)$ of gradient-space connected neighborhoods. Figure 4.9(b) shows two formed gradient adaptive neighborhoods. The edges that separate these pixels, or cells, may be represented as partitions or thin dams (see Fig. 4.9(b)). The heights of these partitions are proportional to the difference between adjacent pixels. When every pixel of the image is flooded by letting water rise to a fixed level ∂ then several noncommunicating pools are formed, each with a unique levels of water.

To form GSCNs we use a standard graph-based technique. An undirected image cover graph is defined as $G = (V, E)$, where each pixel \mathbf{x} of a segmented object $o_k \in \Phi$ has a corresponding node $v(\mathbf{x})$, and to each pair of neighboring pixels $\mathbf{x}_i, \mathbf{x}_j \in o_k$ corresponds one edge $e(\mathbf{x}_i, \mathbf{x}_j)$ with weight:

$$w(e(\mathbf{x}_i, \mathbf{x}_j)) = \prod_{c \in \{R, G, B\}} H(\partial - |D_c(\mathbf{x}_i) - D_c(\mathbf{x}_j)|), \quad (4.16)$$

where H is the Heaviside step function:

$$H(y) = \begin{cases} 0 & \text{if } y < 0 \\ 1 & \text{otherwise} \end{cases} .$$

Then a new graph is formed $\tilde{G} = (V, \tilde{E})$ such that $\tilde{G} \subset G, \forall e \in E \text{ s.t. } |e| \neq 0 \Rightarrow e \in \tilde{G}$. In other words, this new graph inherits all vertices of the graph G and only edges with

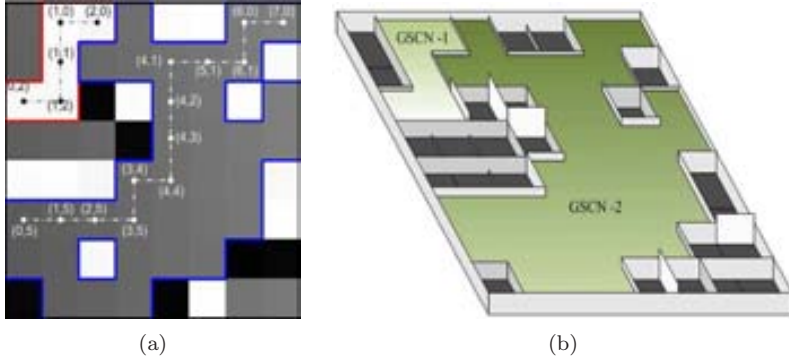


Figure 4.9: (a) Illustration of the GSCN concept, (b) formed GSCNs.

nonzero weight $|e|$. To reach a new subdivision of an object o_k , a forest of graph trees is formed with the aid of depth first or breadth first search algorithms [12]. Finally, all vertices (pixels) of a tree form a sub-segment inside a considered object such that:

$$\bigcup_{l=1}^{L_k} s_l^k = o_k, \bigcap_{l=1}^{L_k} s_l^k = \emptyset, \bigcup_{k=1}^K o_k = \Phi. \quad (4.17)$$

where L_k is the number of trees, i.e. the number of sub-segments s_l^k of object o_k , of the graph \tilde{G} .

Such segmentation is unique and only depends on a given threshold ∂ . Further, it permits us to classify as shadow or foreground all pixels inside a sub-segment s_l . In contrast, approaches based on sliding windows must form a neighborhood for each pixel separately, and consequently have much higher computational complexity. However, we choose gradient-space connected neighborhoods not only because of the low computational complexity relative to similar segmentation techniques, but also because for our method is very important to obtain a unique shadow segment for every shadow region. The main reason for this requirement is that sometimes shadows are formed by a large penumbra, and therefore segmentation algorithms based on global pixel analysis, such as mean shift and watershed segmentations, can fail due to over- and/or under-segmentation. This is illustrated in Fig. 4.10. For example, if a considered region must include only the pixels with values in some fixed interval ∂ then the segmentation process usually splits shadows region or merges shadow and foreground pixels into one segment. This situation is shown in Fig. 4.10(b), where ∂ is a small interval value that results in three segments $\{s_1, s_2, s_3\}$, while $\tilde{\partial}$ is a big interval value that under-segments by merging $SR \subset \tilde{s}$. If there is no penumbra region Fig. 4.10(a), over-segmentation does not occur.

Figures 4.11 illustrates two segmentation examples. The segments in the images are represented by random colors (white color is assigned to those pixels that cannot form a group, in other words, segment with only one pixel); it is observed that in general segments in the shadowed areas are bigger than segments in the foreground regions.

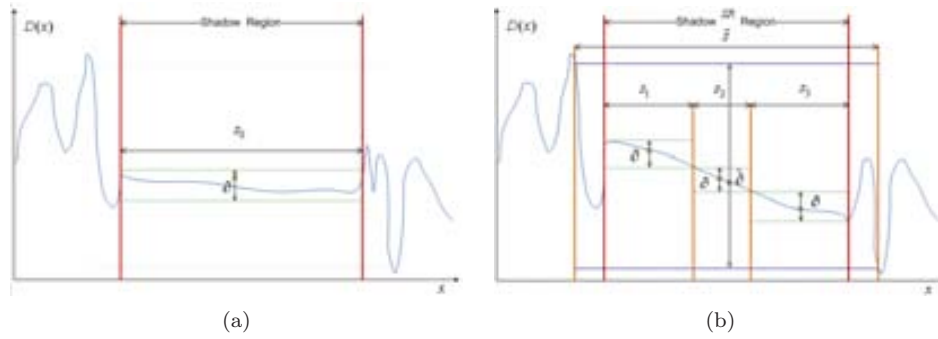


Figure 4.10: Illustration of shadow splitting and over-segmentation effects. (a) Signal without penumbra effect. (b) Signal with penumbra that causes splitting and over-segmentation.

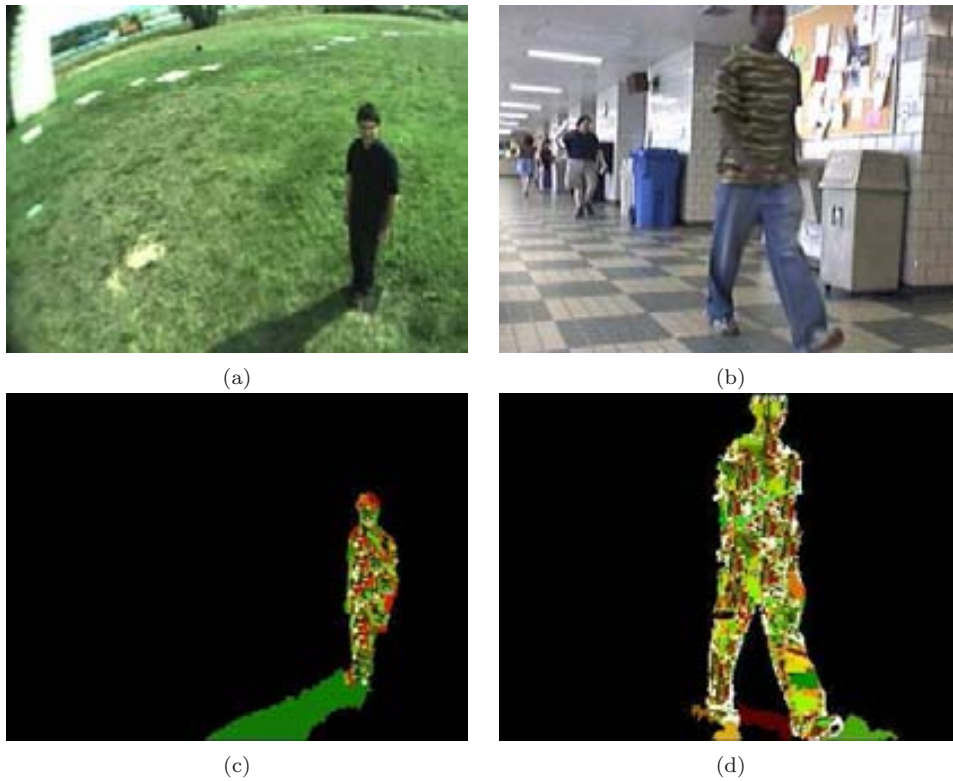


Figure 4.11: (a) and (b) test image frames. (c) and (d) GSCN segmentation results in the luminance ratio space. The segments in the images are represented by random colors.

4.6 Classification Process

Three classification criteria are used to classify every sub-segment s_l^k of an object region o_k as shadow. They exploit local features of the regions. The features used to perform this local classification are:

$\{\mu_l^k, |s_l^k|, \tau_l^k\}$, where:

- μ_l^k : Mean value in the region s_l^k .

$$\mu_l^k = |s_l^k|^{-1} \sum_{\mathbf{x} \in s_l^k} \mathbf{D}(\mathbf{x}).$$

- $|s_l^k|$: Number of pixels that belong to a segment s_l^k .

$$|s_l^k| = \sum_{\mathbf{x} \in o_k} M_{\mathbf{x} \in s_l^k}(\mathbf{x}).$$

- τ_l^k : Terminal pixel weight, where $\hat{t}_l^k(s_l^k)$ is the number of external terminal pixels of a sub-segment s_l^k . This means, the number of edges of sub-segment s_l^k that are neighbor to non-object parts of the image. Finally, $t_l^k(s_l^k)$ is the number of all terminal pixels of the sub-segment s_l^k . This includes all internal and external pixels.

$$\tau_l^k = \frac{\hat{t}_l^k(s_l^k)}{t_l^k(s_l^k)}.$$

Sub-segments are classified based on the combination of three decision rules:

1. Luminance difference criterion.

The first classification rule is based on the assumption that in shadow regions the luminance of each RGB pixel component in the background image is greater than the luminance of each RGB pixel component in the current frame. Following this concept we introduce the shadow-like indicator function:

$$Sh_\mu(\mu_l^k) = \prod_{c \in \{R, G, B\}} H(\mu_c(s_l^k) - 1). \quad (4.18)$$

where $Sh(s_l^k) = 1$ when the sub-segment s_l^k can belong to the shadow class. When $Sh(s_l^k) = 0$, the sub-segment s_l^k can be directly classified as foreground.

2. Segment size criterion.

The size classification rule is based on the assumption that for each shadow region of an object o_k , there is just one shadow sub-segment (the ideal case), or

that there are a few relatively large shadow sub-segments. In contrast, the object o_k is generally formed by many foreground sub-segments, but each of these sub-segments contains few pixels as a result of superposition of two topological structures: background and foreground. Thus, the segment size criterion is represented by the indicator function:

$$Sh_{|s|}(|s_l^k|) = \begin{cases} 1, & \text{if } (|s_l^k| > |o_k| \lambda) \\ 0, & \text{otherwise,} \end{cases} \quad (4.19)$$

where λ is the relative size of the smallest sub-segment in an object o_k that can be shadow.

3. *External terminal point weight criterion.*

This rule is based on the spatial topology of shadows. Shadow regions are usually located around foreground regions. Therefore any shadow sub-segment of an object o_k contains a considerable amount of external terminal points of the region (see Fig. 4.12), relative to the total number of terminal points in the region. In contrast, foreground sub-segments typically have the weight of the extrinsic terminal points equal to zero or have less amount of such points. Therefore, the external terminal point weight criterion is:

$$Sh_{\tau}(\tau_l^k) = \begin{cases} 1, & \text{if } (\tau_l^k > \tau_0) \\ 0, & \text{otherwise,} \end{cases} \quad (4.20)$$

where τ_0 is an experimentally determined threshold.

However, when motion regions $o_k \in \{\Phi\}$ are detected, they might include outliers pixels in the object mask that form a bright narrow fringe between shadow and background regions in the image plane. This edge effect can result from JPEG or similar compression techniques and can also be the result of other signal transmission artifacts. This outlier fringe can completely spoil the result of classification based on the external terminal point weight criterion. To overcome the negative consequences of the 'bright edge effect' we shrink each individual motion region o_k mask by a small morphological erosion. Only then we start the classification process. When the classification is finished we add all unclassified pixels that belong to the edge region to the nearest shadow or foreground sub-region of the motion region o_k .

Joint classification rule

The final shadow classification rule is based on the superposition of all previously described criteria equations: (4.18), (4.19) and (4.20):

$$s_l^k = \begin{cases} \mathbf{Shadow}, & \text{if } (Sh_{\mu}(\mu_l^k) \cap Sh_{|s|}(|s_l^k|) \cap Sh_{\tau}(\tau_l^k)) \\ \mathbf{Foreground}, & \text{otherwise} \end{cases} \quad (4.21)$$

The result of the proposed method applied in three sequences frames is reported in Fig. 4.13. Rows (a-i) in Fig. 4.13 illustrate all steps:

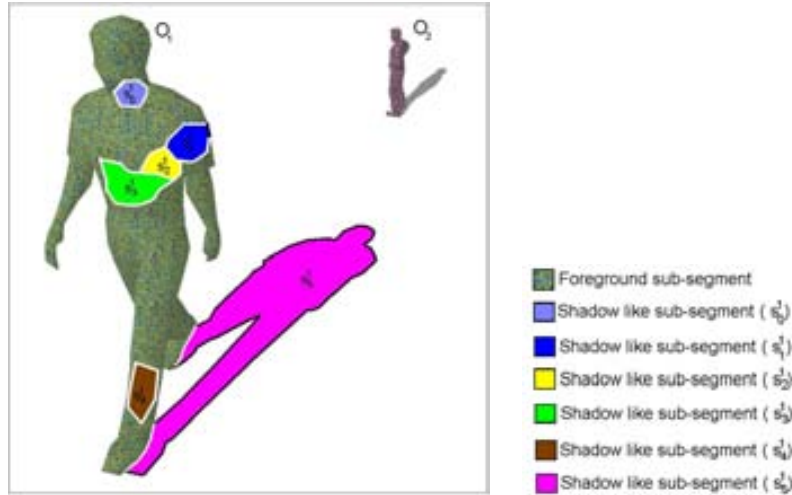


Figure 4.12: Point-wise border representation: white borderlines represent end points of the segment spatially connected with another object's segments; black borderlines represent external terminal points.

- (a) The image being segmented.
- (b) The motion detection binary mask.
- (c) Object masks.
- (d) Ratio luminance space.
- (e) Result of GSCN segmentation.
- (f) Edge noise correction.
- (g) Classification based on the luminance difference and segment size criteria.
- (h) Classification based on the terminal point weight criterion.
- (i) Final segmentation.

The three different columns of Fig. 4.13 represent three different scenes taken from sequences: (I) - Grass field #184, (II) - Highway II #157 and (III) Highway II #801. For scene (I) the classification was completely done based on the segment size criterion (I-g), in spite of that the scene has an irregular background. The classification could be done directly with segment size criterion because in this case the background is rich in term of texture and the foreground is darker than the background. In the scene (II) the luminance difference criterion also plays an important role since the big part of the foreground is brighter than the background, but the final decision was done by terminal point weight criterion. The classification of the scene (III) can be done only with all three proposed criteria.

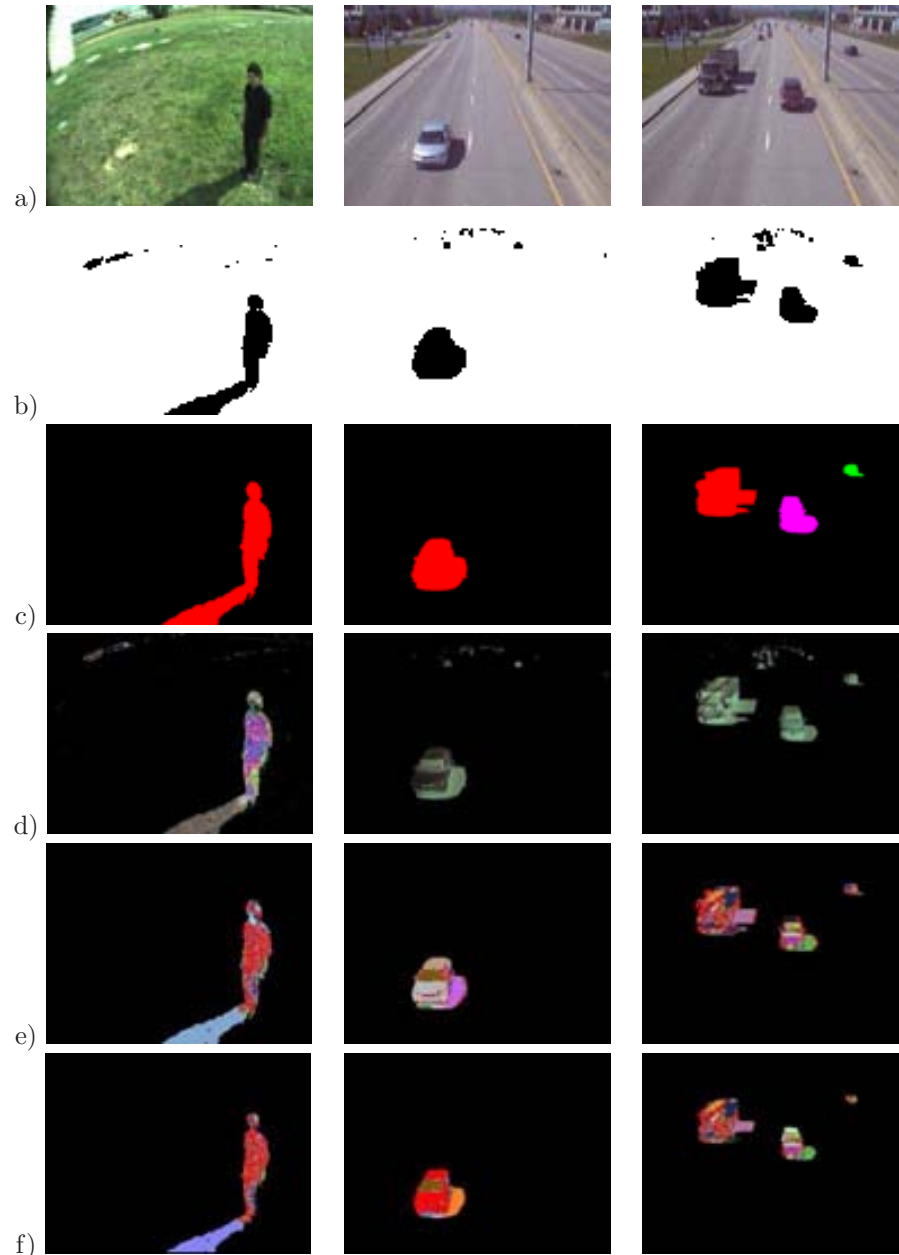


Figure 4.13: Results of different steps of the method (red color means foreground, green color means shadow): (a) image being segmented; (b) motion detection mask; (c) object masks; (d) image difference plane; (e) result of GSCN segmentation; and (f) edge noise correction. *(figure continue on next page).*

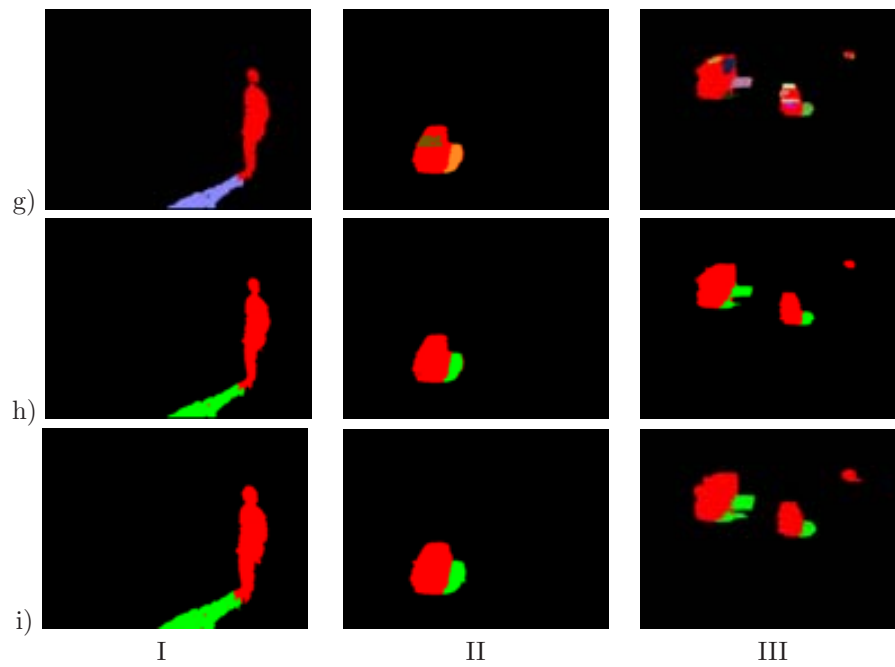


Figure 4.13: (*continued*) (g) classification based on the luminance difference and segment size criteria; (h) classification based on the terminal point weight criterion; and (i) final segmentation. The used sequences are: (I) Grass field #184; (II) Highway III #157; and (III) Highway III #801.

4.7 Experimental Results

This section first discusses the parameters that are involved in the method as well as their estimation, and next, explores the performance of the proposed approach from a qualitative and quantitative viewpoint.

Parameter Analysis

The proposed method relies on several parameters that must be set. Each parameter and how they may each be estimated directly from data, is described next.

1. *Minimum gradient threshold* ∂ .

For every motion object segment o_k a specific ∂_k must be computed:

$$\partial_k = \alpha |o_k|^{-1} \left(\sum_{\mathbf{x} \in o_k} |(\sigma^{bg}(\mathbf{x}))^2| \right), \quad (4.22)$$

where $(\sigma^{bg}(\mathbf{x}))^2$ is the variance of the background image model described in Section 4.2, and $|o_k|$ represents the number of pixels in region k . Because this parameter is a vectors in the RGB color space, $|(\sigma^{bg}(\mathbf{x}))^2|$ is the magnitude or grayscale value of a color vector. So, the threshold ∂_k is proportional to the mean values of the variance over an entire motion segment o_k that must be sub-segmented into a set $s_i^k \in \{o_k\}$. Although the experimental parameter α of equation (4.22) could be optimized for each scene, it was robust for all tested sequences ($\alpha = 0.045$).

2. *Relative size threshold* λ .

The value of the relative size threshold λ has been calculated on the basis of the optimization of two criteria: true positive foreground (TPf) and false positive foreground (FPf). The value of this threshold could be optimized for each scene. However, a scene independent value can be used, since the final result does not show a big variation relative to optimal (less than 0.5%) with λ equal to 0.04. Note that this parameter can be decreased if image or scene conditions warrant. The segment size criteria uses λ to quickly discard sub-segments formed by a small number of pixels. In the case of extreme situations where shadows are formed by few pixels, the value of λ can be safely decreased at the cost of increasing the number of sub-segments to be classified by the 'External terminal point wight criterion'.

3. *External terminal point weight threshold* τ_0 .

The value of the threshold τ_0 defined in equation (4.20) is calculated using the same optimization process as in the previous for λ . At least for all tested sequences the value of $\tau_0 = 0.30$ was optimal.

Performance Evaluation

This section presents quantitative and qualitative results in order to demonstrate the validity of the proposed approach. The method was tested over a wide range of scenes

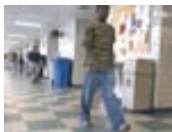
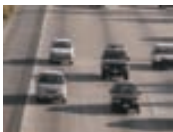

		<i>Sequences</i>		
				
		Hallway	HWI	HWIII
Frames	Number	1800	440	2227
	Hand-labeled	13	8	7
	Size	320x240	320x240	320x240
Scene	Type	Indoor	Outdoor	Outdoor
	Background	Textured	Textured-less	Texture-less
	Noise	Medium	Medium	Medium
Object	Class	People	Vehicles	Vehicles
	Size	Variable	Large	Variable (small)
Shadows	Size	Variable	Large	Variable
	Visibility	Low	High	High
	Direction	Multiple	Single horizontal	Single horizontal
	Camouflage	Low	High	High
	Chromatic effect	Low	Low	High

Table 4.1

DESCRIPTION OF THE SEQUENCES HALLWAY, HW I AND HW III.

with variations in the type and size of objects as well as shadows. In turn, this section also presents comparisons with most classical and sophisticated moving cast shadow detection methods.

Quantitative Results:

The quantitative evaluation is based on two standard metrics for evaluating the performance of cast shadow detection algorithms introduced by Prati et al. [62]. They are: shadow detection rate (η) and shadow discrimination rate (ξ) (previously defined in section 3.3.2).

The quantitative comparison was done with six methods ([13, 25, 50, 1, 47, 28]). The Cucchiara et al. [13], the Horprasert et al. [25] and the McKenna et al. [50] methods were chosen since they are the most cited works in moving cast shadow detection area. While the Martel-Brisson et al. method [47] as well as the Jia-Bin Huang et al. method [28] were selected since these methods perform best among current state-of-the-art techniques.

Comparative results are shown on the standard benchmark sequences¹: Hallway, Highway I (HW I) and Highway III (HW III).(see Table 4.1)

We have ground-truthed three additional sequences for the purpose of bench-

¹Note that the ground-truth used for these sequences was provided by Martel-Brisson et al. [47] <http://vision.gel.ulaval.ca/~CastShadows/>. In the work of Jia-Bin Huang et al. [28] the authors make use of the same sequences as well as the ground truth.




		<i>Sequences</i>		
				
		CVC Outdoor	Football Match	Pets 2009 V7
Frames	Number	800	2699	795
	Hand-labeled	12	13	16
	Size	320x240	320x240	720x576
Scene	Type	Outdoor	Outdoor	Outdoor
	Background	Textured	Textured-less	Variable
	Noise	Low	Medium	Low
Object	Class	People	People	People
	Size	Large	Small	Variable
Shadows	Size	Large	Small	Variable
	Visibility	High	Low	Low
	Direction	Single horizontal	Multiple horizontal	Single horizontal
	Camouflage	Low	Low	Low
	Chromatic effect	Medium	Low	Low

Table 4.2

DESCRIPTION OF THE SEQUENCES CVC OUTDOOR, FOOTBALL MATCH AND PETS 2009 V7.

marking shadow suppression algorithms and in order to provide further evidence of the performance of the proposed approach. These sequences and ground truth are publicly available and were selected to contain a variety of imaging scenarios and challenging background conditions². The new sequences are: CVC-Outdoor, Football Match and Pets-2009 View 7 (see Table 4.2).

The results of methods' performance are reported in Table 4.3. Best performances are highlighted using bold text. Note that there is no publicly available source code or executables for the methods reported in [47, 28], for this reason the performance evaluation of these methods over CVC Outdoor, Football Match and Pets 2009 V7 sequences are missing. Consequently the results reported in Table 4.3 of these methods over Hallyway, HW I and HW III sequences have been obtained directly from [47, 28]. The rest of the methods ([13, 25, 50, 1]) were faithfully implemented. Furthermore, the thresholds of these methods were manually selected to achieve the best overall performance for every video sequence. Note that in the work of Prati et al. [62] different results of the methods from [13] ($\eta = 69$ and $\xi = 76$) and [25] ($\eta = 81$ and $\xi = 63$) over HWI sequence were reported. The difference arises since different test frames were chosen to perform the comparison. The selected frames used in this thesis for the quantitative comparison in (Hallway, HW I and HW III) were the same that have been proposed by Martel-Brisson et al. [47].

²http://www.cvc.uab.es/~aamato/Shadows_Detection/

Sequences Methods	Hallyway		HWI		HWIII		CVC Outdoor		Football Match		Pets 2009 V7	
	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$
McKenna	67	72	56	49	28	62	80	61	64	94	64	87
Horprasert	71	76	71*	51*	31	60	76	64	69	90	66	88
Cucchiara	74	77	61*	64*	39	55	74	68	67	91	72	87
Amato	81	82	59	51	26	41	79	66	66	94	84	92
Brisson	72	86	70	84	68	71						
Huang	82	90	70	82	76	74						
Proposed	84	91	81	85	72	75	91	96	80	95	96	95

Table 4.3

QUANTITATIVE RESULTS FOR DIFFERENT SEQUENCES.

The comparison shows that the proposed technique surpasses the overall performance of the other methods. The lowest performance was found in the HW III sequence. In this sequence the size of moving objects (vehicles) are variable due to the relative position with respect to the camera, when objects are far from the camera they became very small, thus inducing the misclassification.

Qualitative Results:

Visual results of the proposed method over different sequences are shown in Fig. 4.14 and Fig. 4.15. The selected sequences for Fig. 4.14 were: Hallway (#450, #279 and #367), HWI (#18, #46 and #147) and HWIII (#152, #205 and #294). For the Fig. 4.15: CVC Outdoor (#189, #456 and #691), Football Match (#256, #263 and #1211) and Pets 2009 V7(#52, #99 and #316).

Examples of method's performance in different challenging scenarios are illustrated in Fig. 4.16. In the figure five frames taken from five different scenes are included: (I) - Hallway frame #163; (II) - Auto frame #1143; (III) - Highway III frame #253; (IV) - Highway I frame #353; and (V) - CVC outdoor #509. The columns in Fig. 9 (a), (b) and (c) represent:

- (a) The image being segmented.
- (b) Motion object mask.
- (c) Final segmentation.

Scene (I) is an indoor scenario where shadows are projected on the floor and on the wall being a two disconnected shadows patch. Scene (II) is also an indoor scenario, but the environment contains multiple overlapping light sources and a large penumbra region. Scene (III) shows an outdoor scene with flat gray background that is affected by severe shadow camouflage and with chromatic shadow. Scene (IV) is another outdoor scenario where multiple moving objects are combined in a single object mask.

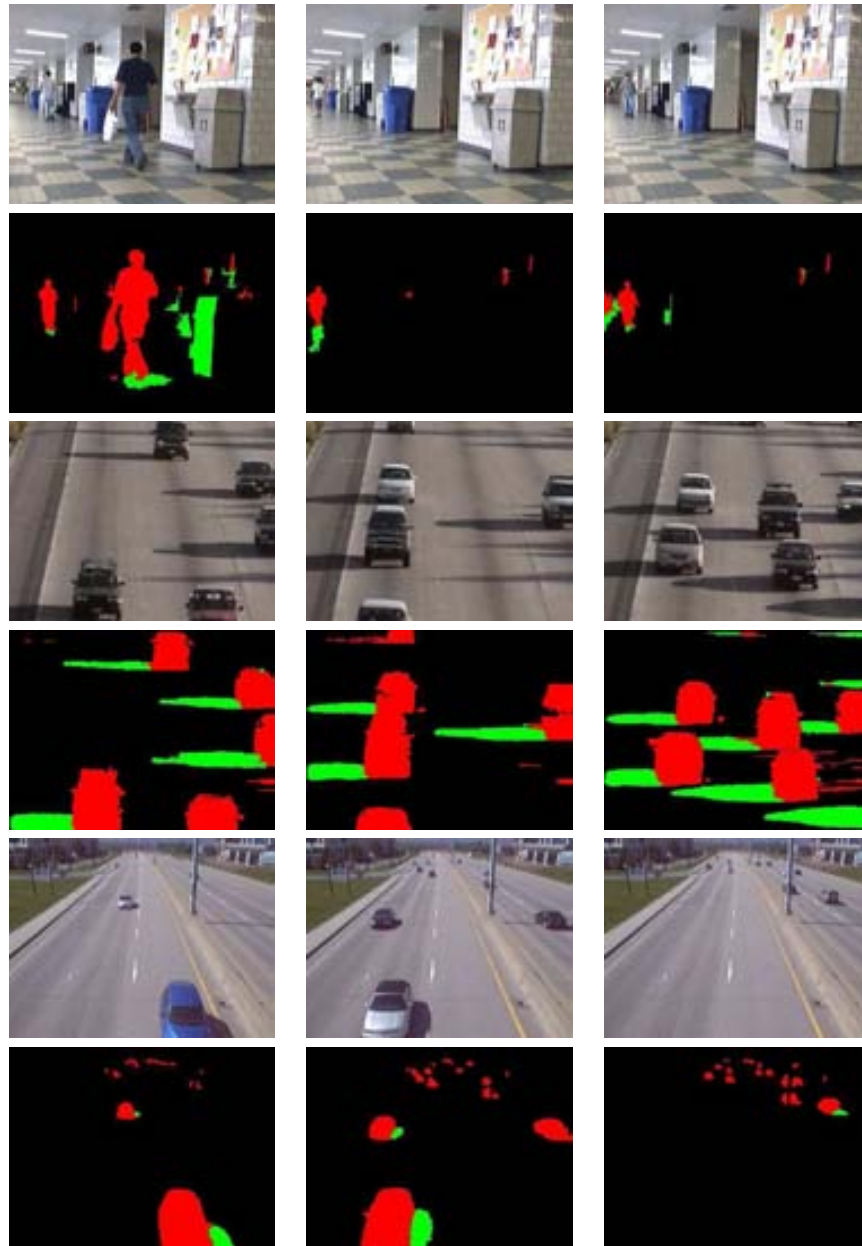


Figure 4.14: Results from the proposed method in different sequences: Hallway; HWI; HWIII.

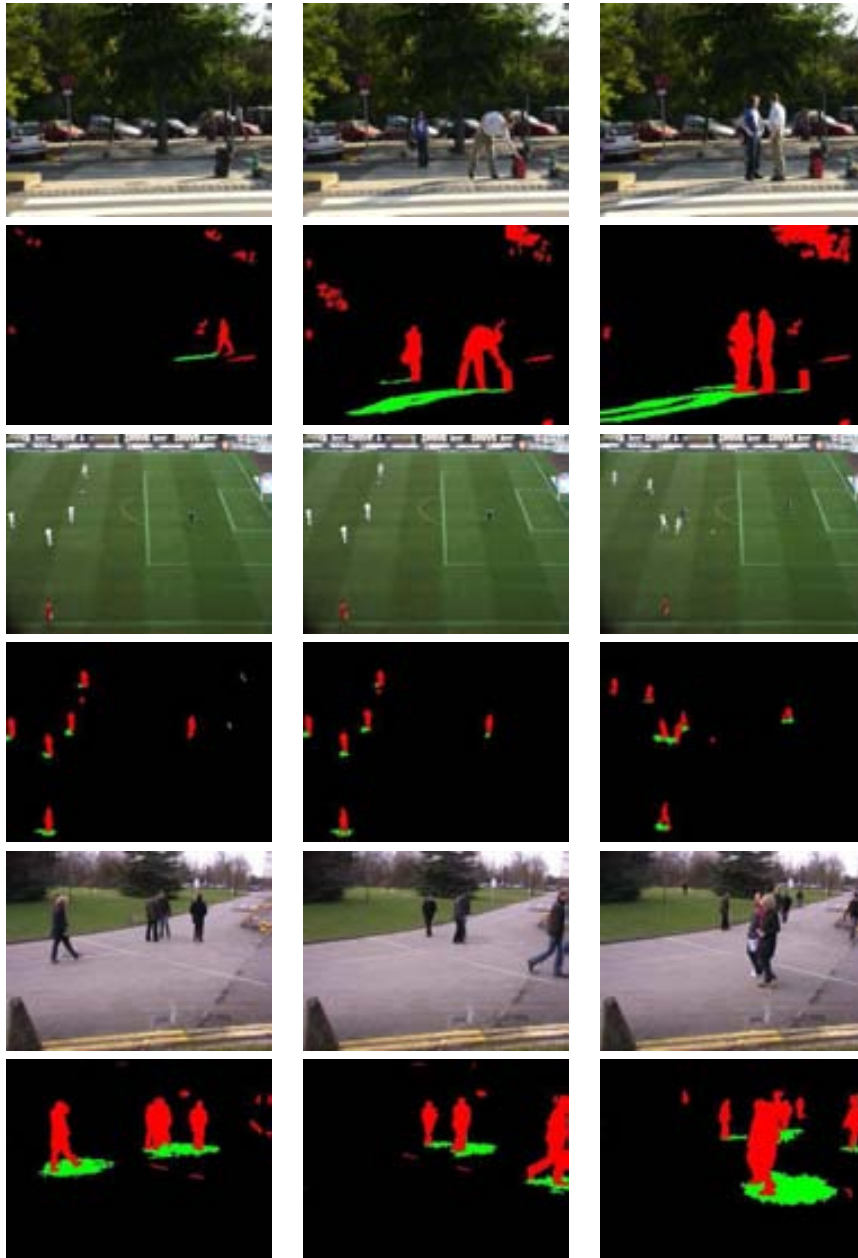


Figure 4.15: Results from the proposed method in different sequences: CVC outdoor; Football Match; Pets 2009 V7.

METHODS	Chromatic Shadows	Shadow Camouflage	Surface Topology
Cucchiara et al. [13]	<i>High</i>	<i>High</i>	<i>Low</i>
Horprasert et al. [25]	<i>High</i>	<i>High</i>	<i>Low</i>
McKena et al. [50]	<i>High</i>	<i>High</i>	<i>High</i>
Kim et al. [36]	<i>High</i>	<i>High</i>	<i>Low</i>
Siala et al. [71]	<i>Low</i>	<i>High</i>	<i>Low</i>
M.-Brisson et al. [47]	<i>High</i>	<i>High</i>	<i>Medium</i>
Huang et al. [28]	<i>High</i>	<i>High</i>	<i>Medium</i>
Fung et al. [18]	<i>High</i>	<i>High</i>	<i>High</i>
Huerta et al. [29]	<i>Low</i>	<i>High</i>	<i>High</i>
Toth et al. [76]	<i>Low</i>	<i>Medium</i>	<i>High</i>
Nadimi et al. [56]	<i>Low</i>	<i>High</i>	<i>Medium</i>
Amato et al. [1]	<i>High</i>	<i>Medium</i>	<i>High</i>
Yuan et al. [91]	<i>Low</i>	<i>Medium</i>	<i>High</i>
Grest et al. [19]	<i>High</i>	<i>Medium</i>	<i>High</i>
Yao et al. [89]	<i>High</i>	<i>Medium</i>	<i>High</i>
Leone et al. [37]	<i>Low</i>	<i>High</i>	<i>High</i>
Jacques et al. [7]	<i>Low</i>	<i>High</i>	<i>High</i>
Yang et al. [88]	<i>Low</i>	<i>High</i>	<i>Medium</i>
Proposed	<i>Low</i>	<i>Low</i>	<i>Medium</i>

Table 4.4

QUALITATIVE EVALUATION FOR DIFFERENT METHODS. THE TABLE VALUATES THE NEGATIVE EFFECT DEGREE WITH: *Low*, *Medium* and *High*.

Finally scene (V) contains a long chromatic shadow cast on an irregular background surface.

Additionally, Table 4.4 presents a qualitative comparison among several moving cast shadow detection algorithms. It reports the negative impact that chromatic shadow and shadow camouflage might cause over the performance of the methods. The table valuates the negative effect degree with: *Low*, *Medium* and *High*. In turn, the table also shows the dependency of algorithms' performance respect to surface topology (namely texture or texture-less). The degree of this dependency is similarly classified.

4.8 Discussion

The robustness of the proposed method to perform under chromatic shadow is due to that in the luminance ratio space a local constant region exists due to the reflectance suppression. Such a local color constancy region is completely independent of the chrominance of the light sources involved in the scene. Moreover, the method can properly classify shadow-foreground pixels even when there is a strong similarity be-

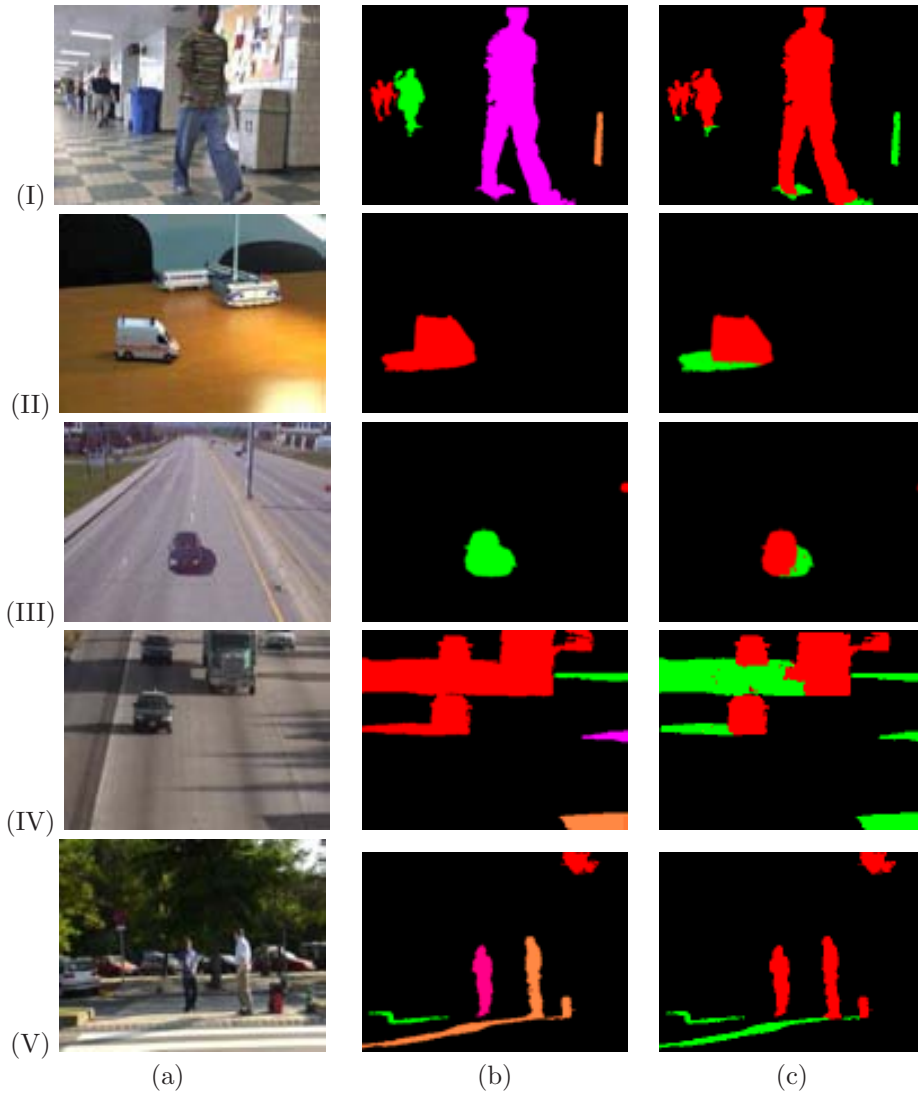


Figure 4.16: Results from the proposed method in different challenging scenarios: (I) Hallway #163; (II) Auto #1143; (III) Highway III #253; (IV) Highway I #353; and (V) CVC outdoor #509. The meaning of each column here is: (a) current image; (b) motion object mask; and (c) final classification.

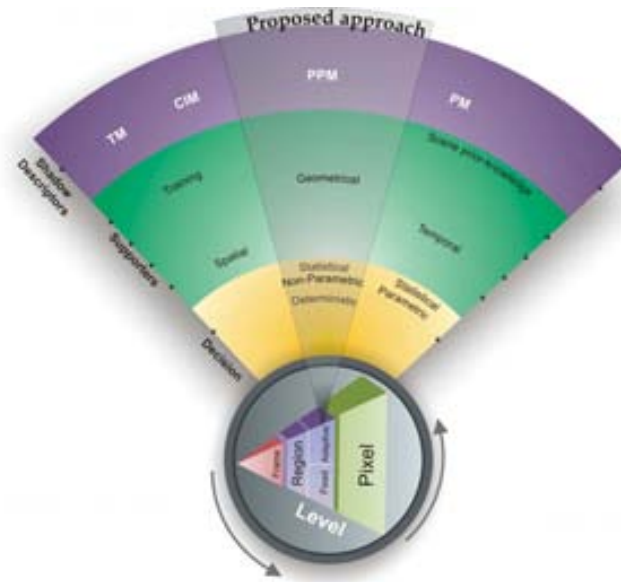


Figure 4.17: Characteristic of the proposed approach.

tween their information, namely chroma and brightness (shadow camouflage), since the method performs at the region level, making use of more than the information of a single pixel.

The computational complexity of the method is lineal to the number of pixel in the frame, thus allowing the detection to be performed in Real-time.

The proposed method has demonstrated high success to distinguish between foreground and moving shadow in different scene conditions. However, the best performance was found in those cases where either background or foreground, were rich in terms of texture. It is because the method can easily discard small segment formed in the foreground region (as a result of superposition of two different topological structures: background and foreground). Following the hypothetical assumption that either background or foreground are always rich in terms of texture the luminance difference filter and the segment size filter could be enough to efficiently distinguish shadow from foreground. Nevertheless, to maintain the generality of the method, thus performing in different scene structures, the terminal point filter was designed. The external terminal point filter exploits geometrical topology of shadows. Comparing among geometrical supporters of shadow, terminal points criterion is the most general, since this does not need any prior knowledge of scene or objects, furthermore it does not need any prior knowledge of sources light.

According to the taxonomy proposed in chapter 2 the presented method follows the configuration illustrated in Fig. 4.17.

Summary

This chapter presents a novel approach to detect and remove moving cast shadows. The method exploits local color constancy due to reflectance suppression over shadowed regions. Such a color constancy effect, over a shadow region, is achieved through dividing the values of the background image by the values of the current image to form a new image in the division space. In this space, segments with low gradients correspond to all shadow regions, as opposed to foreground regions which, in most cases, exhibit higher gradients. The segments are formed by using an algorithm specifically designed for grouping shadowed pixels. To classify each segment as shadow or foreground the method makes use of intrinsic shadows features.

All the parameters in the presented method can be directly estimated from the data, and it deal efficiently and accurately with the most difficult issues in the field, including umbra-penumbra detection, chromatic shadows, and shadow camouflage. The presented method is fast and the computational complexity is linear in the number of pixels in the frame. Furthermore the effectiveness of the proposed method was validated by the higher recognition rates achieved over a collection of publicly available sequences.

Chapter 5

Conclusion

And ne forhtedon na

The final remarks of this dissertation are presented in this chapter. It is organized as follows. Section 5.1 briefly reviews the topics discussed in the different sections of this work and summarizes the main contributions made throughout the development of this thesis. In section 5.2 future lines of research are addressed.

5.1 Summary and Contributions

This thesis has addressed the problem of distinguishing moving cast shadows from the moving objects. The presented work has been motivated by the goal of developing a method that can properly detect moving cast shadows for most possible scenarios occurring in real video sequences.

In chapter 2 the literature has been explored. Different methodologies were found trying to deal with the problem of detecting moving cast shadows. Despite of the fact that the area of research is relatively young, it has engaged the interest of many researchers highlighting the importance in the field.

Moving cast shadow methods mainly identify shadows by using some shadow descriptors. These descriptors basically model shadows by exploiting properties such as: chromaticity invariant, textural patterns, photometric physical models, or even by analyzing the projected areas in terms of size, shape and direction. Diverse information that characterizes moving shadows is exploited and in many cases such information is combined or used in different ways.

With the purpose to better individualize the capability of methods to perform under different situation we have proposed a new taxonomy to represent moving cast shadow detection methods. The main layer of the proposed taxonomy describes whether methods operate with a single pixel or with a group of pixels (within a local adaptive region or a fixed region) or using the whole frame information. The second layer taxonomy describes the shadow descriptors used by the methods. The principal shadow descriptors used in the literature are: chromaticity and intensity models, texture models, photometric physical models and projection models. Some methods

need to be supported by extra information. Such information can be obtained by a training phase or using some prior-knowledge of the scene or exploiting geometrical, temporal, or spatial cues. Finally the last category is based on the classification's decision of the methods.

In chapter 3 first, a known reflection model to describe shadows from a photometric viewpoint has been used. A reflection model basically describes the interaction of light with a surface in terms of the properties of the surface and the nature of the incident light. Normally, in such interaction there are too many number of variables and factors, which make the modeling a non trivial task. In this thesis we have used the Phong reflection model to derivate the shadow luminance model. The shadow luminance model was crucial to: *(i)* describe the effect inside of the shadows (umbra-penumbra); *(ii)* describe the phenomenon caused by chromatic shadows; and *(iii)* establish the basis of the proposed method.

Subsequently a case analysis of different moving cast shadows algorithms is also presented. We have analyzed algorithms that perform in different color spaces based on two kinds of methodologies: *(i)* pixel-wise; and *(ii)* fixed region-wise.

We have observed that methods that only exploit chromaticity invariant property are intrinsically not prepared to cope with 'chromatic shadows'. Methods that perform at the 'pixel level' severely decrease their performance in those cases where 'shadow camouflage' and 'chromatic shadows' occur, since the information of a single pixel is not enough to discriminate between shadow and foreground due to the ambiguity in their pixels' values.

On the other hand, the performance of methods that analyze 'fixed regions' is strongly dependent to the size of the region. Furthermore the textural composition of the background as well as of the object is crucial to obtain a good segmentation.

The detector proposed in this thesis is described in chapter 4. First an initial change detection mask containing moving objects and moving shadows is obtained using a background subtraction technique. Then, objects' masks are computed by using connected component analysis. Based on the shadow luminance model, we state that in the luminance ratio space, a low gradient constancy exists in all shadowed regions, as opposed to foreground regions which, in most cases, exhibit higher gradients. To exploit these foreground-shadow characteristics, we designed a gradient-based segmentation algorithm to partition each object area into a set of low gradient segments (object's sub-segments). The object's sub-segments are classified as shadow or foreground, following three criteria: *(i)* luminance difference criterion; *(ii)* segment size criterion; and *(iii)* extrinsic terminal point weight criterion.

The effectiveness of the proposed method has been validated by the higher recognition rates achieved over a collection of publicly available sequences.

The outcome of this work ends with a new moving cast shadow detector, which is able to properly performs in most possible scenarios occurring in real video sequences. Furthermore, the proposed detector outperforms the state-of-art [2].

Table 5.1 summarizes an account of the specific contributions achieved throughout the development of this thesis and the chapters in which such contributions appear.

LOCATION	CONTRIBUTIONS
Chapter 2	A comprehensive review of the literature. A new taxonomy of moving cast shadow methods.
Chapter 3	A deep analysis of shadow modeling. An exhaustive examination of relevant methods, which includes description, comparison and limitations.
Chapter 4	A motion detection algorithm based on angular and euclidean distance similarity measurement. A shadow feature based on low gradient region in the luminance ratio space. A gradient-based segmentation approach. A geometrical shadow supporter based on external terminal points.

Table 5.1
THESIS CONTRIBUTIONS

5.2 Future Lines of Research

Future directions will be split into two paths.

The first one, aims to evaluate the possibility of using temporal information (temporal supporter) in order to improve the general performance of the method. The purpose behind this idea resides in to cope with wrong sub-segment classifications that can appear in specific location of the scene by using sub-segment confident detection ranks obtained from previous frames.

The second path opens other line of research, which aims to design a structural-color based signature of the moving object that can highly benefit tracking algorithms as well as re-identification algorithms. To this end, a preliminary idea consists first in performing a color segmentation of the object. Then, the object will be represented in a graph where each node of the graph is associated with a color segment of the object and the edges are the spatial connections between the color segments. Consequently the tracking and/or the re-identification will be performed by the matching between two graphs (model graph and target graph).

Appendix A

Publications

Journal Articles:

- Ivan Huerta, Ariel Amato, F. Xavier Roca, Jordi Gonzàlez, “Multiple Cues Fusion for Robust Motion Segmentation using Background Subtraction”. *Neurocomputing*, (Elsevier). (in press 2011)
- Ariel Amato, Mikhail G. Mozerov, Andrew D. Bagdanov, Jordi Gonzàlez, “Accurate Moving Cast Shadow Suppression based on Local Color Constancy Detection”. *IEEE Transactions on Image Processing (TIP)* vol.20, no.10, pp. 2954-2966, October 2011.
- Ariel Amato, Mikhail G. Mozerov, F. Xavier Roca, Jordi Gonzàlez, “Robust Real-Time Background Subtraction based on Local Neighborhood Patterns”. *EURASIP Journal on Advances in Signal Processing 2010*.
- Mikhail G. Mozerov, Ariel Amato, F. Xavier Roca, Jordi Gonzàlez, “Solving the Multi-Object Occlusion Problem in a Multiple Camera Tracking System”. *Pattern Recognition and Image Analysis*, volume 19, number 1, pp. 165–171. March 2009.
- Mikhail G. Mozerov, Ariel Amato, F.Xavier Roca, Jordi Gonzàlez, “Trajectory Occlusion Handling with Multiple View Distance Minimization Clustering”. *Optical Engineering*, vol. 47(4) pp- 047202-1/047202-9. April 2008 SPIE.

Conferences

- Mikhail Mozerov, Ariel Amato, Xavier Roca, ”Occlusion Handling in Trinocular Stereo using Composite Disparity Space Image”. In *19th International Conference on Computer on Graphics and Vision (GraphiCon 2009)*, Moscow, Russia, October, 2009.

- Ariel Amato, Mikhail Mozerov, Ivan Huerta, Jordi González, Juan J. Villanueva, "Background Subtraction Technique Based on Chromaticity and Intensity Patterns", In *In 19th International Conference on Pattern Recognition (ICPR 2008)*. Tampa (Florida), USA. Pattern Recognition 2008 pp.1-4.
- Mikhail Mozerov, Ariel Amato, F. Xavier Roca, Jordi González, "Trajectory Extrapolation for Multiple Views Tracking" In *8th International Conference on Pattern Recognition and Image Analysis (PRIA 2008)*, Yoshkar-Ola, Russia, October, 2007.
- Ariel Amato, Murad Al Haj, Mikhail Mozerov, Jordi González, "Trajectory Fusion for Multiple Camera Tracking" In *5th International Conference on Computer Recognition Systems (CORES'2007)*, Wroclaw, Poland, 2007, vol 45, pp 19–26.
- Mikhail Mozerov, Ariel Amato, Murad Al Haj, Jordi González, "A Simple Method of Multiple Camera Calibration for the Joint Top View Projection" In *5th International Conference on Computer Recognition Systems (CORES'2007)*, Wroclaw, Poland, 2007, vol 45, pp 164–170.
- Murad Al Haj, Ariel Amato, F. Xavier Roca, Jordi González, "Face Detection in Color Images using Primitive Shape Features" In *5th International Conference on Computer Recognition Systems (CORES'2007)*, Wroclaw, Poland, 2007, vol 45, pp 179–186.

Workshops

- Ariel Amato, Murad Al Haj, Josep Llados, Jordi González, "Computationally Efficient Graph Matching via Energy Vector Extraction" In *International Workshop on Advances in Pattern Recognition (IWAPR 2007)*, Plymouth, UK, 2007, pp 47–53.
- Murad Al Haj, Ariel Amato, Gemma Sánchez, Jordi González, "On-line One Stroke Character Recognition Using Directional Features" In *International Workshop on Advances in Pattern Recognition (IWAPR 2007)*, Plymouth, UK, 2007, pp 145–151.
- Ariel Amato, Murad Al Haj, Mikhail Mozerov, Jordi González, "Trajectory Reconstructions with Multiple View". In *2nd CVC Workshop: Progress of Research and Development (CVCRD'2007)*. Cerdanyola del Vallès, Barcelona, Spain, October 2007.
- Murad Al Haj, Javier Orozco, Ariel Amato, F. Xavier Roca, Jordi González. "Finding Faces in Colour Images through Primitive Shape Features" In *2nd CVC Workshop: Progress of Research and Development (CVCRD'2007)*. Cerdanyola del Vallès, Barcelona, Spain, October 2007.

- Pau Baiget, Carles Fernández, Ariel Amato, F. Xavier Roca, Jordi González. “Constructing a Path Database for Scene Categorization”. In *2nd CVC Workshop: Progress of Research and Development (CVCRD'2007)*. Cerdanyola del Vallès, Barcelona, Spain, October 2007.

Technical Reports

- Ariel Amato, “Multiple Camera Calibration for Trajectories Tracking”. CVC Technical Report 112, CVC (UAB), July 2007.

References

- [1] A. Amato, M. Mozerov, X. Roca, and J. Gonzàlez. Robust real-time background subtraction based on local neighborhood patterns. *EURASIP Journal on Advances in Signal Processing*, pages 1–7, June 2010. [Pages **30**, **31**, **35**, **45**, **48**, **55**, **59**, **60**, **67**, **68**, **73**, **92** and **93**]
- [2] A. Amato, M.G. Mozerov, A.D. Bagdanov, and J. Gonzàlez. Accurate moving cast shadow suppression based on local color constancy detection. *Image Processing, IEEE Transactions on*, 20(10):2954–2966, oct. 2011. [Pages **24** and **102**]
- [3] Ariel Amato, Murad Al Haj, Mikhail G. Mozerov, and Jordi Gonzàlez. Trajectory fusion for multiple camera tracking. In *5th International Conference on Computer Recognition Systems (CORES'2007), Wroclaw, Poland*, volume 45, pages 19–26, march 2007. [Page **11**]
- [4] O. Barnich and M. Van Droogenbroeck. Vibe: A universal background subtraction algorithm for video sequences. *IEEE TIP*, 20(6):1709–1724, June 2011. [Page **40**]
- [5] S. Brutzer, B. Hoferlin, and G. Heidemann. Evaluation of background subtraction techniques for video surveillance. In *IEEE CVPR'11*, pages 1937–1944, June 2011. [Page **38**]
- [6] Anthony Caputo. *Digital Video Surveillance and Security*. Butterworth-Heinemann, 2010. [Page **10**]
- [7] J. Cezar Silveira Jacques, C. Rosito Jung, and S.R. Musse. A background subtraction model adapted to illumination changes. In *Image Processing, 2006 IEEE International Conference on*, pages 1817–1820, October 2006. [Page **34**]
- [8] Chia-Jung Chang, Wen-Fong Hu, Jun-Wei Hsieh, and Yung-Sheng Chen. Shadow elimination for effective moving object detection with gaussian models. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 2, pages 540–543, 2002. [Pages **31** and **37**]
- [9] Y. Chen, C. Chen, C. Huang, and Y. Hung. Efficient hierarchical method for background subtraction. *Pattern Recognition*, 40(10):2706–2715, October 2007. [Page **40**]

- [10] Li Cheng, M. Gong, D. Schuurmans, and T. Caelli. Real-time discriminative background subtraction. *IEEE TIP*, 20(5):1401–1414, 2011. [Page 40]
- [11] A. Colombari, A. Fusiello, and V. Murino. Patch-based background initialization in heavily cluttered video. *IEEE TIP*, 19(4):926–933, April 2010. [Page 40]
- [12] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms, Second Edition*. The MIT Press and McGraw-Hill Book Company, 2001. [Pages 76 and 84]
- [13] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti. Improving shadow suppression in moving object detection with hsv color information. In *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*, pages 334–339, 2001. [Pages 29, 30, 33, 45, 48, 54, 59, 60, 68, 92 and 93]
- [14] Weiyao Lin (Ed.). *Video Surveillance*. InTech, 2011. [Page 10]
- [15] A. Elgammal, D. Harwood, and L. S. Davis. Nonparametric background model for background subtraction. In *ECCV'00*, pages 751–767, Dublin, 2000. [Pages 40 and 73]
- [16] G. Finlayson, DHordley, Steven D., and Cheng Lu. On the removal of shadows from images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(1):59–68, 2006. [Page 25]
- [17] D. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, August 2002. [Pages 9 and 17]
- [18] G.S.K. Fung, N.H.C. Yung, G.K.H. Pang, and A.H.S. Lai. Effective moving cast shadow detection for monocular color image sequences. In *Image Analysis and Processing, 2001. Proceedings. 11th International Conference on*, pages 404–409, sep 2001. [Page 36]
- [19] Daniel Grest, Jan michael Frahm, and Reinhard Koch. A color similarity measure for robust shadow removal in real time. In *In Vision, Modeling and Visualization*, pages 253–260, 2003. [Pages 31 and 34]
- [20] I. Haritaoglu, D. Harwood, and L.S. Davis. W4: Real-time surveillance of people and their activities. *IEEE TPAMI*, 22(8):809–830, 2000. [Page 40]
- [21] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, Second Edition, 2003. [Page 11]
- [22] J. Heikkila and O. Silven. A real-time system for monitoring of cyclists and pedestrians. In *Proceedings of the Second IEEE Workshop on Visual Surveillance*, pages 74–81, Washington, DC, USA, 1999. IEEE Computer Society. [Page 39]
- [23] M. Heikkila and M. Pietikainen. A texture-based method for modeling the background and detecting moving objects. *IEEE TPAMI*, 28(4):657–662, 2006. [Page 31]

- [24] T. Horprasert, D. Harwood, and L.S. Davis. A statistical approach for real-time robust background subtraction and shadow detection. In *IEEE Frame-Rate Applications Workshop*, Kerkyra, Greece, 1999. [Page **73**]
- [25] Thanarat Horprasert, David Harwood, and Larry S. Davis. A statistical approach for real-time robust background subtraction and shadow detection. In *ICCV Frame-Rate WS*. IEEE, 1999. [Pages **29, 30, 33, 45, 48, 51, 58, 59, 60, 68, 92** and **93**]
- [26] Jun-Wei Hsieh, Shih-Hao Yu, Yung-Sheng Chen, and Wen-Fong Hu. A shadow elimination method for vehicle analysis. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, pages 372–375, aug. 2004. [Pages **31** and **37**]
- [27] Jun-Wei Hsieh, Shih-Hao Yu, Yung-Sheng Chen, and Wen-Fong Hu. Automatic traffic surveillance system for vehicle tracking and classification. *Intelligent Transportation Systems, IEEE Transactions on*, 7(2):175–187, june 2006. [Page **31**]
- [28] Jia-Bin Huang and Chu-Song Chen. Moving cast shadow detection using physics-based features. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:2310–2317, 2009. [Pages **31, 33, 92** and **93**]
- [29] I. Huerta, M. Holte, T.B. Moeslund, and J. González. Detection and removal of chromatic moving shadows in surveillance scenarios. In *ICCV2009*, Kyoto, Japan, 2009. [Pages **30** and **37**]
- [30] Ivan Huerta, Ariel Amato, F. Xavier Roca, and Jordi González. Multiple cues fusion for robust motion segmentation using background subtraction. *Neurocomputing, Elsevier, in press*, 2011. [Page **17**]
- [31] A. Ilyas, M. Scuturici, and S. Miguet. Inter-camera color calibration for object re-identification and tracking. In *Soft Computing and Pattern Recognition (SoC-Par), 2010 International Conference of*, pages 188–193, dec. 2010. [Page **11**]
- [32] H.W.S. Jabri, Z. Duric, and A. Rosenfeld. Detection and location of people in video images using adaptive fusion of color and edge information. In *15th ICPR*, volume 4, pages 627–630, Barcelona, Spain, September 2000. [Page **41**]
- [33] O. Javed, K. Shafique, and M. Shah. A hierarchical approach to robust background subtraction using color and gradient information. In *Proc. of the Workshop on Motion and Video Computing (MOTION'02)*, page 22, Orlando, 2002. [Page **41**]
- [34] Li Jin-chao, Tang Hui-ming, and Lu Chao. Noise estimation in video surveillance systems. In *Computer Science and Information Engineering, 2009 WRI World Congress on*, volume 6, pages 578–582, april 2009. [Page **11**]
- [35] M. Karaman, L. Goldmann, D. Yu, and T. Sikora. Comparison of static background segmentation methods. In *VCIP '05*, Beijing, China, July 2005. [Page **38**]

- [36] K. Kim, T.H. Chalidabhongse, D. Harwood, and L.S. Davis. Real-time foreground-background segmentation using codebook model. *Real-Time Imaging*, 11(3):172–185, June 2005. [Pages **33** and **73**]
- [37] A. Leone and C. Distanto. Shadow detection for moving objects based on texture analysis. *Pattern Recognition*, 40(4):1222–1233, April 2007. [Page **35**]
- [38] Kun Li, Qionghai Dai, and Wenli Xu. High quality color calibration for multi-camera systems with an omnidirectional color checker. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 1026–1029, march 2010. [Page **11**]
- [39] L. Li, W. Huang, I. Yu-Hua Gu, and Qi Tian. Statistical modeling of complex backgrounds for foreground object detection. *IEEE TIP*, 13(11):1459–1472, November 2004. [Page **40**]
- [40] T. M. Lillesand and R. W. Kiefer. *Remote Sensing and Image Interpretation, 4th Edition*. Wiley, 2000. [Page **25**]
- [41] Chengjun Liu and Jian Yang. Ica color space for pattern recognition. *Neural Networks, IEEE Transactions on*, 20(2):248–257, February 2009. [Page **45**]
- [42] Zhou Liu, Kaiqi Huang, Tieniu Tan, and Liangsheng Wang. Cast shadow removal combining local and global features. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8, june 2007. [Page **38**]
- [43] E. Lopez-Rubio, R.M. Luque-Baena, and E. Dominguez. Foreground detection in video sequences with probabilistic self-organizing maps. *International Journal of Neural Systems*, 21(3):225–246, 2011. [Page **41**]
- [44] L. Maddalena and A. Petrosino. A self-organizing approach to background subtraction for visual surveillance applications. *IEEE TIP*, 17(7):1168–1177, July 2008. [Page **41**]
- [45] V. Mahadevan and N. Vasconcelos. Spatiotemporal saliency in dynamic scenes. *IEEE TPAMI*, 32(1):171–177, 2010. [Page **41**]
- [46] N. Martel-Brisson and A. Zaccarin. Learning and removing cast shadows through a multidistribution approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(7):1133–1146, 2007. [Pages **31** and **33**]
- [47] Nicolas Martel-Brisson and André Zaccarin. Kernel-based learning of cast shadows from a physical model of light sources and surfaces for low-level segmentation. In *CVPR08*, pages 1–8, 2008. [Pages **31**, **33**, **92** and **93**]
- [48] A. McIvor. Background subtraction techniques. In *In Proc. of Image and Vision Computing*, Auckland, New Zealand, 2000. [Page **38**]
- [49] S. J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler. Tracking groups of people. *CVIU*, 80(1):42–56, 2000. [Page **73**]

- [50] Stephen J. McKenna, Sumer Jabri, Zoran Duric, Azriel Rosenfeld, and Harry Wechsler. Tracking groups of people. *Computer Vision and Image Understanding: CVIU*, 80(1):42–56, 2000. [Pages **29**, **33**, **45**, **48**, **52**, **59**, **60**, **68**, **92** and **93**]
- [51] H.B. Mitchell. *Image Fusion: Theories, Techniques and Applications*. Springer-Verlag, February 2010. [Page **45**]
- [52] A. Mittal and N. Paragios. Motion-based background subtraction using adaptive kernel density estimation. In *Proc. CVPR'04*, volume 2, pages 302–309, Washington DC, USA, July 2004. [Pages **40** and **73**]
- [53] Mikhail G. Mozerov, Ariel Amato, Murad Al Haj, and Jordi Gonzàlez. A simple method of multiple camera calibration for the joint top view projection. In *5th International Conference on Computer Recognition Systems (CORES'2007)*, Wroclaw, Poland, volume 45, pages 164–170, march 2007. [Page **11**]
- [54] Mikhail G. Mozerov, Ariel Amato, Xavier F. Roca, and Jordi Gonzàlez. Trajectory occlusion handling with multiple view distance minimization clustering. *Optical Engineering*, 47(4):047202–1/047202–9, April 2008. [Page **11**]
- [55] Mikhail G. Mozerov, Ariel Amato, Xavier F. Roca, and Jordi Gonzàlez. Solving the multi-object occlusion problem in a multiple camera tracking system. *Pattern Recognition and Image Analysis*, 19(1):165–171, March 2009. [Page **11**]
- [56] S. Nadimi and B. Bhanu. Physical models for moving shadow and object detection in video. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(8):1079–1087, aug. 2004. [Pages **31** and **37**]
- [57] Goro Obinata and Ashish Dutta. *Vision Systems: Segmentation and Pattern Recognition*. I-Tech Education and Publishing, 2007. [Pages **9** and **17**]
- [58] K. A. Patwardhan, G. Sapiro, and V. Morellas. Robust foreground detection in video using pixel layers. *IEEE TPAMI*, 30(4):746–751, April 2008. [Page **41**]
- [59] M. Piccardi. Background subtraction techniques: a review. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 4, pages 3099 – 3104, The Hague, Netherlands, 2004. [Page **38**]
- [60] A. M. Polidorio, F. C. Flores, N. N. Imai, A. M. G. Tommaselli, and C. Franco. Automatic shadow segmentation in aerial color images. In *XVI Brazilian Symp. Computer Graphics and Image Processing*, pages 270–277, 2003. [Page **25**]
- [61] F. Porikli and J. Thornton. Shadow flow: a recursive method to learn moving cast shadows. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 891–898 Vol. 1, 2005. [Page **33**]
- [62] Andrea Prati, Ivana Mikic, Mohan M. Trivedi, and Rita Cucchiara. Detecting moving shadows: Algorithms and evaluation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(7):918–923, 2003. [Pages **9**, **26**, **29**, **32**, **58**, **92** and **93**]

- [63] Ravi Ramamoorthi, Melissa L. Koudelka, and Peter N. Belhumeur. A fourier theory for cast shadows. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(2):288–295, 2005. [Page **25**]
- [64] Reinhard, Erik, Khan, Erum Arif, Akyz, Ahmet Oguz, and Garrett M. Johnson. *Color Imaging: Fundamentals and Applications*. A. K. Peters, Ltd., 2008. [Page **45**]
- [65] Paul Rosin and Tim Ellis. Image difference threshold strategies and shadow detection. In *in Proc. British Machine Vision Conf*, pages 347–356. BMVA Press, 1995. [Page **36**]
- [66] Elena Salvador, Andrea Cavallaro, and Touradj Ebrahimi. Shadow identification and classification using invariant color models. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, volume 3, pages 1545–1548, 2001. [Pages **29** and **35**]
- [67] Elena Salvador, Andrea Cavallaro, and Touradj Ebrahimi. Spatio-temporal shadow segmentation and tracking. In *Proc. of Visual Communications and Image Processing*, pages 389–400, 2003. [Pages **26**, **29** and **35**]
- [68] Elena Salvador, Andrea Cavallaro, and Touradj Ebrahimi. Cast shadow segmentation using invariant color features. *Computer Vision and Image Understanding*, 95(2):238–259, 2004. [Page **35**]
- [69] J.C. SanMiguel and J.M. Martinez. On the evaluation of background subtraction algorithms without ground-truth. In *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, pages 180–187, sept. 2010. [Page **38**]
- [70] Y. Sheikh and M. Shah. Bayesian modeling of dynamic scenes for object detection. *IEEE TPAMI*, 27(11):1778–1792, November 2005. [Page **40**]
- [71] K. Siala, M. Chakchouk, F. Chaieb, and O. Besbes. Moving shadow detection with support vector domain description in the color ratios space. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, pages 384–387, aug. 2004. [Page **33**]
- [72] Jürgen Stauder. Estimation of point light source parameters for object-based coding. *Signal Processing: Image Communication*, 7(4-6):355 – 379, 1995. [Page **81**]
- [73] Jürgen Stauder, Roland Mech, and Jörn Ostermann. Detection of moving cast shadows for object segmentation. *IEEE Transactions on Multimedia*, 1(1):65–76, 1999. [Pages **38**, **48**, **76** and **81**]
- [74] C. Stauffer, W. Eric, and L. Grimson. Learning patterns of activity using real-time tracking. *IEEE TPAMI*, 22(8):747–757, 2000. [Page **40**]
- [75] C. Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. In *IEEE CVPR'99*, volume 1, pages 22–29, Ft. Collins, CO, USA, 1999. [Page **40**]

- [76] Daniel Toth, Ingo Stuke, Andreas Wagner, and Til Aach. Detection of moving shadows using mean shift clustering and a significance test. In *International Conference on Pattern Recognition (ICPR 2004)*, volume 4, pages 260–263, 2004. [Pages **36** and **81**]
- [77] K. Toyama, J.Krumm, B.Brumitt, and B.Meyers. Wallflower: Principles and practice of background maintenance. In *Proc. ICCV'99*, volume 1, pages 255–261, Kerkyra, Greece, 1999. [Pages **41** and **73**]
- [78] Victor J.D. Tsai. A comparative study on shadow compensation of color aerial images in invariant color models. *IEEE Transactions on Geoscience and Remote Sensing*, 44(6):1661 – 1671, 2004. [Page **25**]
- [79] Phong Bui Tuong. Illumination for computer generated pictures. *Commun. ACM*, 18:311–317, June 1975. [Page **46**]
- [80] Habib Ullah, Mohib Ullah, Muhammad Uzair, and Fasih ur Rehman. Comparative study: The evaluation of shadow detection methods. *INTERNATIONAL JOURNAL OF VIDEO & IMAGE PROCESSING AND NETWORK SECURITY (IJVIPNS)*, 10(2):1–7, April 2010. [Pages **9** and **26**]
- [81] L. Wang, W. Hu, and T. Tan. Recent developments in human motion analysis. *Pattern Recognition*, 36(3):585–601, 2003. [Page **40**]
- [82] Alan Watt. *3D Computer Graphics*. Addison-Wesley, 1993. [Pages **48** and **80**]
- [83] J. Van De Weijer, Th. Gevers, and A.D. Bagdanov. Boosting color saliency in image feature detection. *IEEE TRANS. PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 28:150–156, 2005. [Page **45**]
- [84] Yair Weiss. Deriving intrinsic images from image sequences. In *Proc. ICCV'01*, volume 02, pages 68–75, Vancouver, Canada, 2001. [Page **41**]
- [85] C.R. Wren, A. Azarbayejani, T. Darrell, and A.P. Pentland. Pfnder: Real-time tracking of the human body. *IEEE TPAMI*, 19(7):780–785, 1997. [Page **40**]
- [86] G. Wyszecki and W.S. Stiles. *Color science: Concepts and methods, quantitative data and formulae*. John Wiley, Second Edition, February 1982. [Page **52**]
- [87] Dong Xu, Xuelong Li, Zhengkai Liu, and Yuan Yuan. Cast shadow detection in video segmentation. *Pattern Recognition Letters*, 26(1):91–99, 2005. [Page **36**]
- [88] Yang, Lo, Chinag, and Tai. Moving cast shadow detection by exploiting multiple cues. *Image Processing, IET*, 2(2):95–104, 2008. [Pages **31**, **35** and **81**]
- [89] J. Yao and J.M Odobez. Multi-layer background subtraction based on color and texture. In *IEEE CVPR'07*, pages 17–22, Minneapolis, Minnesota, USA, June 2007. [Pages **31** and **35**]

- [90] A. Yoneyama, C.H. Yeh, and C.-C.J. Kuo. Moving cast shadow elimination for robust vehicle extraction based on 2d joint vehicle/shadow models. In *Proceedings. IEEE Conference on Advanced Video and Signal Based Surveillance, 2003.*, pages 229–236, july 2003. [Page **31**]
- [91] Chao Yuan, Chenhui Yang, and Zhiming Xu. Simple vehicle detection with shadow removal at intersection. In *Proceedings of the 2010 Second International Conference on Multi-Media and Information Technology*, volume 02 of *MMIT '10*, pages 188–191. IEEE Computer Society, 2010. [Pages **31** and **34**]
- [92] Z. Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *Proc. ICPR'04*, volume 2, pages 23–26, August 2004. [Page **40**]
- [93] Z. Zivkovic and F. Heijden. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognition Letters*, 27(7):773–780, May 2006. [Page **40**]