Foreground Object Segmentation and Shadow Detection for Video Sequences in Uncontrolled Environments

A dissertation submitted by **Ivan Huerta Casado** at Universitat Autònoma de Barcelona to fulfil the degree of **Doctor en Informàtica**.

Bellaterra, June 14, 2010

Director: **Dr. Jordi Gonzàlez i Sabaté**
Dept. de Ciències de la Computació, Universitat Autònoma de Barcelona.
Computer Vision Center, Universitat Autònoma de Barcelona.
Co-director: **Dr. Xavier Roca i Marba**
Dept. de Ciències de la Computació, Universitat Autònoma de Barcelona.
Computer Vision Center, Universitat Autònoma de Barcelona.

This document was typeset by the author using LaTeX 2_ε.

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

ISBN 978-84-937261-3-3

Printed by Ediciones Gráficas Rey, S.L.

Dedicado a mis padres, Angelines y Fermin,
y a mi hermano Mario.

Acknowledgments

Mirando hacía atrás puedo ver como ha pasado el tiempo, mi infancia, mi escuela, mi instituto, incluso como fue cuando pise por primera vez esta universidad, han pasado muchos muchos años, he conocido mucha gente, y he compartido muchas cosas. He acabado una carrera y ahora estoy apunto de acabar un doctorado, por el medio muchas cosas han pasado, pero creo que lo más importante a sido disfrutar de este viaje.

Lo curioso de recordar es que siempre recuerdas las cosas buenas, por este motivo pienso que estos cinco años que he estado aquí haciendo el doctorado me han servido de mucho más que para tener un título, he aprendido mucho de los papers, de los libros, de los journals, pero puedo considerar sin lugar a dudas que de quien he aprendido más es de la gente que me rodea, es por eso que quisiera agradecer a todo el mundo que ha estado aquí conmigo, que ha compartido algo conmigo, que me ha enseñado algo, o que simplemente hemos tenido una breve charla en un pasillo, por que que es la vida sino un aprendizaje continuo. Entre estas cuatro paredes, he aprendido, me he divertido y alguna vez incluso he trabajado ;) Fuera bromas, la verdad que en el fondo me sabe mal el haber acabado, por el simple hecho de que dejo atrás una parte de mi, que se que no volverá, pero que es la vida sino el acabar cosas para comenzar otras. Por que aunque la verdad veo incierto mi futuro por primera vez desde que comencé a tener uso de razón (y no me refiero a mi infancia :) se que seguro que en el fondo me van a esperar cosas buenas.

Es curioso ahora leyendo los agradecimientos que hice en mi tesina, comentaba que tenía todavía dos años por delante para acabar la tesis, como pasa el tiempo. Han pasado tres años y aun recuerdo perfectamente cuando hablé por primera vez con Dani Rowe, fue gracias a el sin lugar a dudas que comencé a hacer el proyecto de fin de carrera y que después se convirtió en el doctorado. Aun recuerdo cuando me lo encontré por primera vez cuando me lo presento Juanjo y me dio un paper para q me lo fuera leyendo, todavía lo recuerdo por que fue ni más ni menos que un survey de 34 paginas en ingles, Wang 'Recent developments in Human Motion Analysis' esta en la bibliografía por si alguien se lo quiere leer, jeje. Por eso Dani muchas gracias, gracias por enseñarme, por dirigirme, y sobretodo por que aun sin saberlo me has mostrado un camino el cual no pensaba que pudiera existir. Me da pena que no estuvieras aquí para hacer todo el doctorado por que sinceramente me hubiera gustado seguir discutiendo y aprendiendo contigo.

Quiero dar las gracias a Juanjo quien me dio la oportunidad de poder hacer un doctorado, aunque pueda ser que después se haya arrepentido de ello, jeje. Muchas

gracias por que es una oportunidad que creo que el lector puede comprobar por si mismo si he aprovechado. Gracias también a Xavi, por que es la persona que cuando tienes un problema se las arregla todavía no se como para solucionarlo, creo que debe tener una legión de enanos que trabajan para el de escondidas. Poal no tiene esos enanos, por que siempre esta ocupado, yo creo que necesitaría dos vidas para llegar a poder hacer las cosas que quiere o que tiene que hacer. Pero sin lugar a dudas es un jefe que todo el mundo le gustaría tener, eso si tener más tiempo, jajjaja. Fuera bromas, muchas gracias Poal, por todo el tiempo que me has dedicado aun sin tenerlo disponible, por todos esos ratos hablando, por esas birras compartidas, y sobretodo por lo que me has enseñado. Se que sin tu ayuda y no tan solo en el sentido académico sino también en el político no hubiera podido llegar a escribir estos agradecimientos. Por eso mis más sinceros agradecimientos son para ti, por que se que este doctorado no se podría haber acabado sin tu ayuda.

I would like to also thank to Dr. Thomas Moeslund for accepting and driving my stage in Aalborg, Denmark, thanks for giving good advices. Many thank Michael Holte because it is a pleasure work with you.

Aunque mis compañeros de pupitre han sido muchos en estos 5 años, quisiera hacer una mención especial a todos aquellos con los que he disfrutado de su compañía. Especialmente, a Pau que creo que es con la persona que he pasado más tiempo y he compartido más cosas, también a Carles por que siempre esta allí dispuesto a ayudarte cuando lo necesitas, a Bhaskar con quien hablando encuentro puntos diferentes de vista que agrandan mi pequeño mundo. También agradecer la ayuda y compañía del resto del grupo ISE como Marco, Murad, Pep, Miguel, Ariel, Marc, Joan, Ignasi entre otros. También me gustaría agradecer a la gente del trabajo, con los que he compartido muchos ratos, muchas risas, y espero que más cervezas Edu, Marçal, Alicia, Sergio, Xevi, Debora, Joan, Pierluigi, ... A mis compañeros de despacho Partha, David & David, Jose, Agnes, entre otros. A esas nuevas incorporaciones que vienen pisando fuerte Camp, Monica, Toni... a los compañeros del labo, Enric, Raul, ... y en general a toda la gente del CVC, que se que son muchos y siempre te dejas alguno. Quiero agradecer especialmente también a todas las chicas de administración, de dirección, de marketing, de secretaria, que te ayudan siempre con una sonrisa de oreja a oreja, a Montse por que siempre me solucionas todos los problemas aun con el poco tiempo que te doy, muchas gracias. A Gigi, a Helena, a Ana, a Mari, a Mireia, a Raquel, a todas, muchas gracias. Y a todos los demás muchas gracias también.

Y como no, agradecer a mis amigos de toda la vida, los nuevos, los viejos, a los q ya no veo. Por estar siempre allí alrededor, ayudándome y apoyándome, algunos ya no los veo, pero igualmente muchas gracias Jordi F., Oscar, Dani H, Magi, Vanesa, Victor, Dani R. A mis amigos de la uni como Jordi, Victor, Gabri, Lluis entre otros. Todos no estáis aquí porque no tengo espacio para ponerlos, pero a todos ellos os agradezco vuestra compañía. Agradecer a Iria que fue de gran apoyo en mis primeros años de doctorado cuando hice la tesina, muchas gracias.

Y finalmente, lo más importante para mi, a las personas que quiero agradecer y dedicar este doctorado. A mi madre Angelines, por que madre solo hay una y yo creo que he tenido mucha suerte por que me he debido llevar la mejor, muchas gracias por ayudarme, por animarme, por consolarme, por cuidarme, por darme todo tu cariño y amor, muchas gracias por ser como eres. A mi padre Fermin, por que siempre esta allí

cuando lo necesitas, por que siempre puedes contar con el, por su amor, su cariño, por eso y mil razones más muchas gracias. A mi hermano Mario, por que aunque muchas veces nos peleemos y nos enfademos siempre esta ahí y se que siempre estará, por que el hablar contigo y estar contigo es un placer, muchas gracias. A toda mi familia que siempre me esta dándome todo su amor y cariño, a mi tías Vicenta, Milagros, Mari, a mis tíos Segundo, Antonio, Agustin, Narciso. A mis primos Miriam, David, Anabel, Yolanda, Aitor, Juanjo, Francis, y a las nuevas generaciones que han venido :)

Muchas gracias a todos !!

Abstract

This Thesis is mainly divided in two parts. The first one presents a study of motion segmentation problems. Based on this study, a novel algorithm for mobile-object segmentation from a static background scene is also presented. This approach is demonstrated robust and accurate under most of the common problems in motion segmentation. The second one tackles the problem of shadows in depth. Firstly, a bottom-up approach based on a chromatic shadow detector is presented to deal with umbra shadows. Secondly, a top-down approach based on a tracking system has been developed in order to enhance the chromatic shadow detection.

In our first contribution, a case analysis of motion segmentation problems is presented by taking into account the problems associated with different cues, namely colour, edge and intensity. Our second contribution is a hybrid architecture which handles the main problems observed in such a case analysis, by fusing (i) the knowledge from these three cues and (ii) a temporal difference algorithm. On the one hand, we enhance the colour and edge models to solve both global/local illumination changes (shadows and highlights) and camouflage in intensity. In addition, local information is exploited to cope with a very challenging problem such as the camouflage in chroma. On the other hand, the intensity cue is also applied when colour and edge cues are not available, such as when beyond the dynamic range. Additionally, temporal difference is included to segment motion when these three cues are not available, such as that background not visible during the training period. Lastly, the approach is enhanced for allowing ghost detection. As a result, our approach obtains very accurate and robust motion segmentation in both indoor and outdoor scenarios, as quantitatively and qualitatively demonstrated in the experimental results, by comparing our approach with most best-known state-of-the-art approaches.

Motion Segmentation has to deal with shadows to avoid distortions when detecting moving objects. Most segmentation approaches dealing with shadow detection are typically restricted to penumbra shadows. Therefore, such techniques cannot cope well with umbra shadows. Consequently, umbra shadows are usually detected as part of moving objects.

Firstly, a bottom-up approach for detection and removal of chromatic moving shadows in surveillance scenarios is proposed. Secondly, a top-down approach based on kalman filters to detect and track shadows has been developed in order to enhance the chromatic shadow detection. In the Bottom-up part, the shadow detection approach applies a novel technique based on gradient and colour models for separating chromatic moving shadows from moving objects.

Well-known colour and gradient models are extended and improved into an invariant colour cone model and an invariant gradient model, respectively, to perform automatic segmentation while detecting potential shadows. Hereafter, the regions corresponding to potential shadows are grouped by considering "a bluish effect" and an edge partitioning. Lastly, (i) temporal similarities between local gradient structures and (ii) spatial similarities between chrominance angle and brightness distortions are analysed for all potential shadow regions in order to finally identify umbra shadows.

In the top-down process, after detection of objects and shadows both are tracked using Kalman filters, in order to enhance the chromatic shadow detection, when it fails to detect a shadow. Firstly, this implies a data association between the blobs (foreground and shadow) and Kalman filters. Secondly, an event analysis of the different data association cases is performed, and occlusion handling is managed by a Probabilistic Appearance Model (PAM). Based on this association, temporal consistency is looked for the association between foregrounds and shadows and their respective Kalman Filters. From this association several cases are studied, as a result lost chromatic shadows are correctly detected. Finally, the tracking results are used as feedback to improve the shadow and object detection.

Unlike other approaches, our method does not make any a-priori assumptions about camera location, surface geometries, surface textures, shapes and types of shadows, objects, and background. Experimental results show the performance and accuracy of our approach in different shadowed materials and illumination conditions.

Resum

Aquesta tesis esta dividida en dos parts principalment. A la primera, es presenta un estudi dels problemes que es poden trobar en la segmentació per moviment, basant-se en aquest estudi es presenta un algoritme genèric el qual es capaç de solucionar d'una forma acurada la majoria dels problemes que es poden trobar en aquest tipus de segmentació. En la segona part, es tracta el tema de les ombres en profunditat. Primer, es presenta un algoritme bottom-up basat en un detector de ombres cromàtiques el qual es capaç no només de solucionar les ombres que es troben a la penombra, sinó també les ombres que podem trobar a l'umbra. Segon, es presenta un sistema top-down basat en un sistema de tracking per tal de trackejar les ombres i d'aquesta manera millorar la detecció de les ombres cromàtiques.

En la nostra primera contribució, presentem un anàlisis del possibles problemes que trobem en la segmentació per moviment quan utilitzem el color, els gradients, o la intensitat. La nostra segona aportació es una arquitectura hibrida la qual pot solucionar els principals problemes observats en l'anàlisi, mitjançant la fusió de (i) la informació obtinguda per aquestes tres cues, i (ii) un algoritme de diferencia temporal. Per un costat, em aconseguit millorat els models de color i de gradients per que puguin solucionar tant el problemes amb els canvis de il.luminació global y local (com les ombres no cromàtiques) i els camuflatges en intensitat. A més a més, la informació local es explotada per tal de solucionar el problema dels camuflatges en croma. Per una altra banda, la intensitat es aplicada quan el color i els gradients no estan disponibles degut a problemes en la obtenció d'aquests (es troben fora del rang dinàmic). Addicionalment, la diferencia temporal es inclosa en la segmentació per moviment en el moment en que cap de les cues estudiades no estan disponibles, com per exemple quan el fons de la imatge no es visible en el període de entrenament. Per últim en aquesta primera part, el nostre algoritme també es capaç de solucionar el problema de les segmentacions fantasma. Com a resultat, el nostre algoritme obté una segmentació robusta i acurada tant en escenaris d'interior com d'exterior, tal i com s'ha demostrat tant quantitativament com qualitativament en els resultats experimentals, mitjançant la comparació del nostre algoritme amb els més coneguts algoritmes de l'estat de l'art.

La segmentació en moviment té que tenir en compte el problema de les ombres per tal de evitar distorsions quan intentem segmentar els objectes en moviment. Però molts dels algoritmes que son capaços de detectar les ombres solament son capaços de detectar les ombres a la penombra. En conseqüència, aquestes tècniques no son capaces de detectar les ombres a l'umbra les quals son normalment detectades com

part dels objectes en moviment.

En aquesta tesis presentem primer una innovadora tècnica que es basa en els models de gradients i de color per tal de separar aquestes ombres cromàtiques dels objectes en moviment. Primerament, construïm tant un model de color en forma de con, com també un model de gradient els quals son invariant a les cromaticitats per tal d'aconseguir fer una segmentació automàtica a la vegada que totes les possibles ombres son detectades. En un segon pas, les regions que poden ser ombres son agrupades considerant "l'efecte blau" i les particions obtingudes mitjançant els gradients. Finalment, analitzem (i) les similituds temporals entre els les estructures locals dels gradients i (ii) les similituds espacials entre els angles cromàtics i les distorsions de la lluminositat de totes les ombres potencials per tal d'identificar les ombres a la umbra.

Segon, en el procés top-down després de la detecció dels objectes i les ombres els dos son seguits usant un filtre de Kalman, per d'aquesta manera millorar la detecció de les ombres cromàtiques. Primerament, l'algoritme fa una associació entre els blobs (foreground i ombres) i els filtres de Kalman. Segon, es realitza un anàlisis dels possibles casos entre las associacions obtingudes anteriorment, i a més a més es tracten les oclusions mitjançant un Model Probabilístic d'Aparença. Basant-se en aquesta associació es busca la consistència temporal entre els foregrounds, les ombres, i els seus respectius filtres de Kalman. A partir d'aquesta nova associació son estudiats diferents casos, com a resultat les ombres cromàtiques que s'havien perdut son detectades. Finalment, els resultats son utilitzats com a feedback per millorar la detecció de la ombra i del objecte.

Pel contrari que altres algoritmes el nostre mètode no fa cap assumpció a priori sobre la localització de la càmera, les geometries o les textures de les superfícies, les formes o els possibles tipus de ombres, objectes o de fons de la imatge. Els resultats experimentals mostren la performance i la precisió del nostre algoritme en la detecció de les ombres cromàtiques en diferents materials i amb diferents condicions de il.luminació.

Contents

List of Tables

List of Figures

List of Algorithms

Chapter 1

Introduction

Human beings have been trying to emulate the perception of the motion within Computer Science during the last three decades. Hence, important research efforts in computer vision have been focused on developing theories, methods and systems applied to the description of human movements in image sequences. The ultimate aim of it is to interpret people behaviour.

The evaluation of human motion in image sequences involves different tasks, such as acquisition, detection (motion segmentation and target classification), tracking, action recognition, behaviour reasoning and natural language modelling. However, the basis for high-level interpretation of observed patterns of human motion still relies on *when* and *where* motion is being detected in the image. Therefore, motion segmentation constitutes the basic and the most critical step towards more complex tasks such as Human Sequence Evaluation (HSE) [16]. HSE defines an extensive Cognitive Vision System (CVS) which transforms acquired image values into semantic descriptions of human behaviour and synthetic visual representations. A sketch of this system can be seen in Fig. 1.1. Motion segmentation is located in the Image Signal Level (ISL in the figure), where the sequence of image data is processed by segmenting potential targets, see appendix B for more information.

In this work, the focus is placed on one of the main HSE tasks: motion segmentation. Nevertheless, What is motion segmentation? This refers to the extraction process of moving objects from a video sequence. During these three decades, different techniques have been used for motion segmentation such as background subtraction, temporal differencing and optical flow. Even though many algorithms have been proposed in the literature [15, 69, 46, 45], the problem of identifying moving objects in a complex environment is still far from being completely solved. The information obtained from this step is the base for a wide range of applications such as smart surveillance systems, control applications, advanced user interfaces, motion based diagnosis, identification applications among others [15]. Nevertheless, motion segmentation is still an open and significant problem due to dynamic environmental conditions such as illumination changes, shadows, waving tree branches in the wind, etc. And difficulties with physical changes in the scene. However, in this thesis the problem is tackled without setting any kind of restrictions on the nature of the scene.

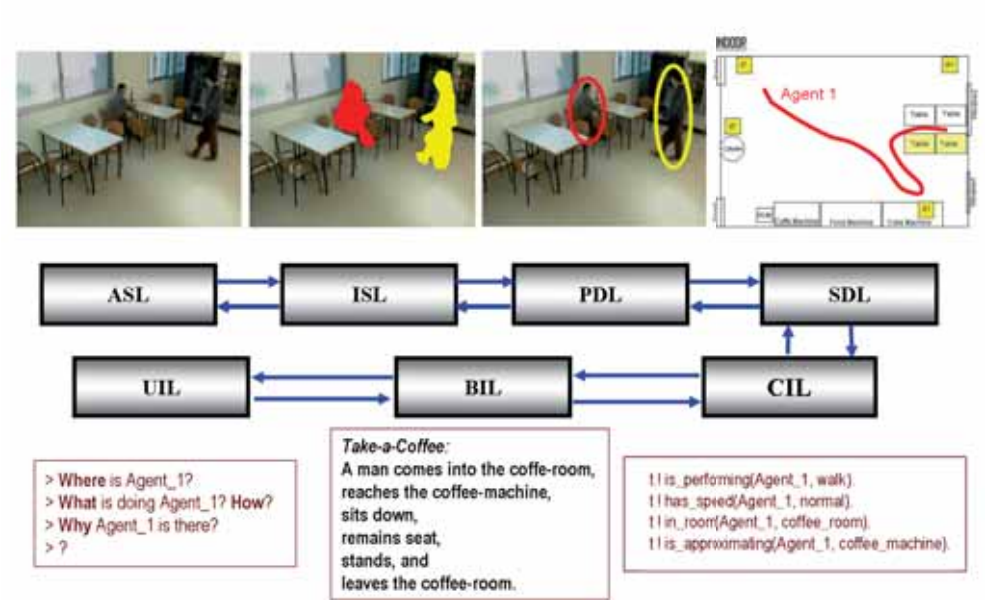


Figure 1.1: Overall architecture for Human Sequence Evaluation (HSE) [17].

1.1 Potential Applications

Human motion analysis has attracted great interests from computer vision researchers due to its promising application in many areas such as virtual reality, smart surveillance systems, advanced user interfaces, motion analysis, model-based coding, among others. Motion Segmentation is a significant issue in a human motion analysis system since the subsequent processes are greatly dependent on it.

Smart Surveillance Systems

Video surveillance is an important application domain. The video data is currently used only as a forensic tool, such as banks and supermarkets where the video data is usually recorded in tapes or stored in video archives. However, the video data can be used in real time.

A smart surveillance is needed in order to generate useful and helpful information, for instance to alert security officers when a burglary is in progress, or a suspicious behaviour such as wandering around and repeatedly looking into the cars in a parking lot, to avoiding false alarms such as animals wandering around, wind blowing, etc. Moreover, face and gait recognition is used with the purpose of access control.

Other smart surveillance applications besides security applications are measure traffic flow, monitor pedestrian congestion in public spaces, compile consumer demographics in shopping malls, etc. A smart surveillance system can bring a lot of benefits but it have to be balanced regarding privacy.

Potential applications	
"Smart" surveillance systems	Access control Parking lots Supermarkets, department stores Vending machines, ATMs Traffic Prevent terrorist attacks Statistical studies
Virtual reality	Interactive virtual worlds Games Virtual studios Character animation Teleconferencing (e.g., film, advertising, home-use)
Advanced user interfaces	Social interfaces Sign-language translation Gesture driven control Signalling in high-noise environments (e.g airports,factories)
Motion analysis	Content-based indexing of sports video footage Personalised training in golf, tennis, etc. Choreography of dance and ballet Clinical studies of orthopaedic patients
Model-based coding	Very low bit-rate video compression

Table 1.1
Potential applications

Other kind of application which has currently more significance due to the terrorist attacks, and the growing fear among population and governments, is to detect packets or suspected objects which are abandoned in places like airports, or undergrounds in order to avoid the terrorist attacks.

Virtual Reality

Another different kind of application domain is virtual reality. The main application is the interaction between the virtual and the real world. One of the most important objectives is to represent humans in the physical space in a virtual space. Tools like internet can be used as a medium to interact among virtual worlds. The interaction inside this virtual worlds can be improved between the participants with the used of cues such as gestures, head pose, and facial expressions.

Other application in virtual reality domain is related to the computer games. The realism of virtual humans and simulated actions in computer games are achieved thanks to the knowledge obtained of the acquisition of human body model, the re-

trieval of body pose, the human behaviour analysis, etc. Other applications are virtual studios, motion capture for character animations (synthetic actors), teleconferencing, etc.

Advanced User Interface

Vision is a useful complement of speech recognition and natural language understanding for a natural and intelligent dialogue between human and machine. This can create more useful and friendly interfaces, which allow interaction between machine and users in a more personable way. More detailed cues can be obtained by gestures, body poses, and facial expressions. That can be helpfully in speech recognition problems such as environmental noise and distance. Therefore, these systems can work independently without be affected by the surrounding environment. Vision can also improve other applications such as sign-language translation, gesture driven controls, signalling in high-noise environment such as factories or airports, or can be helpful for problems with phoneme disambiguation, or helping lip reading.

Motion Analysis

Motion analysis can be useful for the evaluation and training of athletic performance. It lies personalised training systems for various sports; these systems would observe the skills of the pupils and make suggestions for improvement. The gait analysis also aims at providing medical diagnosis and treatment support. The human gait can also be used as a new biometric feature for personal identification. Using the motion analysis can interpret video data sequences automatically using content-based indexing helping human efforts in sorting and retrieving images or video in a huge database. Video-based human-motion analysis is also useful for choreography of dance and ballet, and furthermore, for orthopaedic clinical studies.

Model-based Coding

Model-based image coding can use a low bit-rate video compression for more effective image storage and transmission. For instance, by encoding only the motion part of the scene and by avoiding sending the background part in videophone calling much money can be saved.

1.2 Segmentation Difficulties

Detection of regions that correspond to moving objects such as vehicles and people in natural scenes is a significant and difficult problem. Efficient segmentation simplifies the processing on subsequent steps on analysis [17]. Nevertheless, fast and reliable motion segmentation is an open and difficult problem due to dynamic changes in naturals scenes, such as illumination changes, weather conditions, camouflage, among others. Or due to the algorithm employed to perform the segmentation. Some of them cannot be solved at this level and need to use a posterior HSE level.

Depending on the Scene	Depending on the Algorithm
Global Illumination Changes (Gradual and Sudden)	Ghosts (Hole objects)
Local Illumination Changes (Shadows and Highlights)	Sleeping Objects
Weather Conditions	Incorporated Objects
Background in Motion (i.e. Waving trees, Flowing water)	Sleeping Person
Camouflage	Bootstrapping
	Foreground Aperture

Table 1.2
SEGMENTATION DIFFICULTIES

1.2.1 Due to Scene Conditions

Motion segmentation difficulties due to the scene are (i) *global and (ii) local illumination changes*, *(iii) weather conditions*, *(iv) background in motion*, and *(v) camouflage*. Some examples are shown in Fig 1.2.

- *Global illumination changes*: Lighting conditions in the scene can classified Background pixels erroneously as foreground. These changes can be *gradual* or *sudden* depending on the velocity of the illumination change. Normally, gradual illumination changes are when the illumination of the scene changes gradually with the time of the day. Then, the appearance of the background will be very different at different moments of the day. For instance, it can occurs due to a natural phenomenon such as at dawn, when the day gets lighter, or at dusk, when the day get darker. Sudden illumination occur when the illumination changes suddenly. This problem can appear in both indoor and outdoor scenes, e.g. for an indoor scene when a light is turned on/off, and for an outdoor scene when the sun is covered by clouds. In Fig. 1.2.(a) can be seen an example of gradual illumination change.

- *Local illumination changes –Shadows and Highlights–:* Foreground objects can project shadows, which is an area where direct light from a light source cannot reach due to obstruction by the foreground object. Shadows can be different from the background model in chrominance and intensity, and similar to the foreground objects in the scene. Shadows can be a big problem if the foreground objects are not correctly segmented from their shadows. Then, errors are propagated through the next levels such as classification, tracking, ... For instance, a person could be classified as an animal due to his shadow. Or different objects can be considered only one object because the shadows are joined them together. An example of shadow can be seen in the Fig. 1.2.(b). Highlight are the opposite to a shadow because they do the scene lighter. An highlight example can be seen in Fig. 1.2.(c).

- *Weather conditions:* Specific climatic conditions such as rain, snow, etc, can prevent from getting a clearly background. Therefore, it can be sometimes

Figure 1.2: Example of some motion segmentation difficulties, as mentioned in the text. (a) Gradual illumination changes. (b) Shadows. (c) Highlights. (d) Wheather conditions. (e) Background in motion. (f) Camouflage.

detected as foreground. Some of this weather phenomenon also present changes in the illumination, such as when the clouds cover the sun. In the Fig. 1.2.(d) can be seen how the background can change due to weather conditions.

- *Background in motion:* Sometimes the background presents movement such as the waving tree branches and bushes blowing in the wind, or the water of a river. Therefore, these cases are wrongly segmented as foreground because they exhibits motion and belongs to the background. Usually appears in outdoor scenes. An example of background in motion is the fire such as it can be seen in the Fig. 1.2.(e).
- *Camouflage:* Pixel features of the background model can be considered similar as the foreground pixel ones. Therefore, the foreground object will not be segmented. This problem happens both indoor and outdoor scenes. For instance a woman with a green coat cannot be correctly distinguished from a grass field. In the Fig. 1.2.(f) can be seen an example of camouflage.

1.2.2 Due to Algorithm limitations

Depending on the algorithm employed some of the next difficulties can be found: (i) *ghosts*, (ii) *sleeping objects*, (iii) *incorporated objects*, (iv) *sleeping person*, (v) *bootstrapping* and (vi) *foreground aperture*. Some of these problems cannot be solved at this level and need to be handled at posterior HSE level. The first four difficulties are very similar. After the four explanations, an example shows the relation among them.

- *Ghosts:* If an object which belongs initially to the background begins to move, or it is moved, then the object and the place which the object held in the background are both segmented as foreground pixels. Therefore, the place of the initial background object will be erroneously segmented as foreground. For instance, a car which is parked in a parking lot starts to move. Then, the moving car is correctly segmented, however the initial place where the car was is wrongly detected as foreground indefinitely.
- *Incorporated Objects:* It happens when an object which does not exhibits motion is not added into the background. Therefore, it is segmented as foreground indefinitely thereby impeding other foreground segmentations. For instance when a car parks in a parking lot.
- *Sleeping objects:* When a foreground object which is in motion stop his motion and it is immediately added to the foreground. The problem is that this object has to be longer detected as foreground. A method which has this problem seriously is frame difference, because any object which does not have motion throughout two frames (or more depending on the method used) would be not segmented as foreground. The approaches with this problem does not have the two problems above mentioned.
- *Sleeping person:* This difficulty is related to the last two difficulties, since if the object is an interest object like a person, then this object have not to be part

of the background. This difficulty cannot be solved at the segmentation level because at this level the interesting and uninteresting objects cannot be distinguished. Next, different examples are explained to understand the difference between these difficulties.

For instance, consider a car which stops in a traffic light. This car must keep on being detected within the scene as foreground, because it will continue his way when the traffic light turns on. In case that the car is not detected as foreground then sleeping objects difficulty appears. Instead of stop in a traffic light, the car parks in a parking lot. Then, it must not be detected indefinitely as foreground. Therefore, incorporated objects difficulty will appear if the car is detected indefinitely as foreground. The main difference between these difficulties is the time which objects are without motion. Nevertheless, if it is an agent instead of a car, the agent must not be detected as a part of the background, neither when he stops in a traffic light, nor when he stops because he is talking with another agent. The last case shows how it is impossible in this level distinguish between the interesting object (agent) and the object which is not interesting (the car).

- *Bootstrapping:* Some approaches need to be initialised using a training period without any foreground object. Nevertheless, a training period without any foreground object is not possible in some circumstances. For instance, in a crowed street the people walking during the training period can be incorporated into the background model, thereby building it wrongly. Normally, this difficulty only appears in approaches which uses background models.

- *Foreground aperture:* It happens when the interior pixels of a moving object cannot be segmented as foreground pixels because of the similarity between them. When it happens the border of the homogeneous object is normally segmented, however the interior pixels are not segmented as foreground pixels. This difficulty can appear in methods without background model or when the foreground homogeneous object in motion belongs to the background model.

1.3 Approaches and Contributions

This Thesis is mainly divided in two parts. The first one, firstly presents a study of motion segmentation problems. Based on this study, a novel algorithm for mobile-object segmentation from a static background scene is also presented. This approach is demonstrated robust and accurate under most of the common problems in motion segmentation. The second one tackles the problem of shadows in depth. Firstly a bottom-up approach based on a chromatic shadow detector is presented to deal with umbra shadows. Secondly, a top-down approach based on a tracking system has been developed in order to enhance the chromatic shadow detection. A sketch of the system is shown in Fig. 1.3

- In the first part of this Thesis, our first contribution is a novel theoretical case analysis of motion segmentation problems, where the performance of each cue

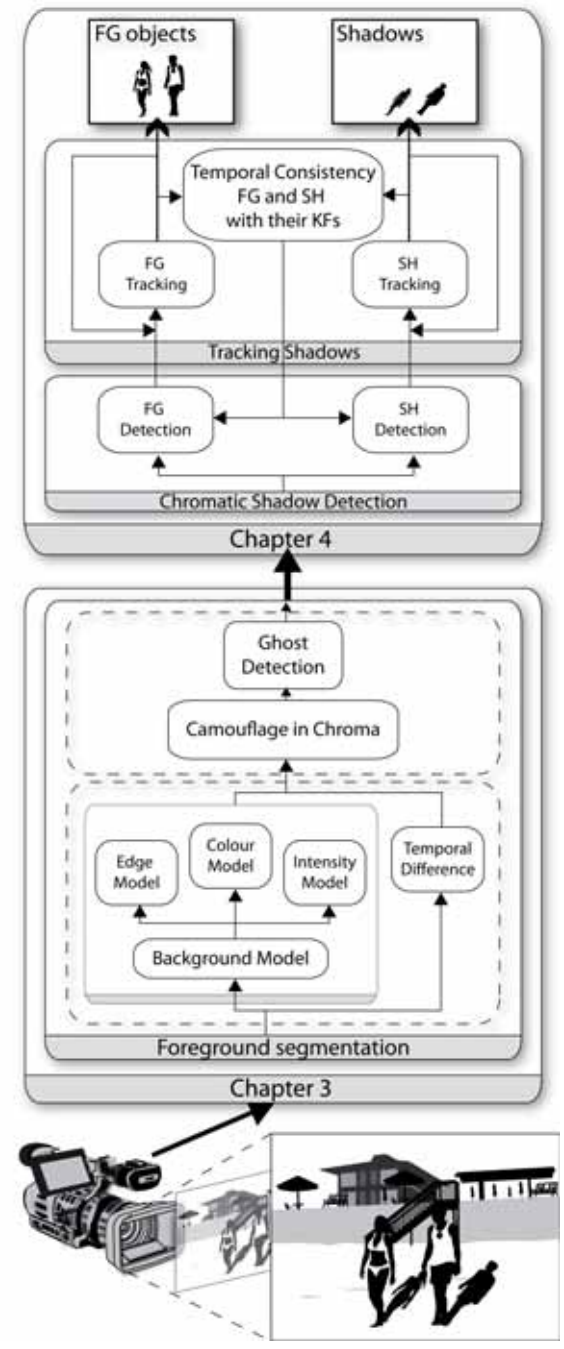


Figure 1.3: Motion segmentation architecture. FG represents the foreground, SH the shadows, and KF the Kalman filters.

used in the literature for segmentation (intensity, colour, and edges) is exhaustively evaluated, showing the advantages of every cue and when a cue can be or cannot be applied. To the best of our acknowledge, current state-of-the-art considers chromatic spaces only, and they do not address most of the problems identified in our case analysis.

- Our second contribution is a new architecture which handles the main problems observed in such a case analysis. The new hybrid approach fuses (i) the knowledge from these three cues and (ii) a temporal difference algorithm because each cue solves a particular problem identified in the case analysis. With this hybrid approach, we enhance the colour and edge models to solve both global/local illumination changes (shadows and highlights) and camouflage in intensity. Cue models employed in the hybrid approach have been improved over existing ones, furthermore their combination is a step forward the current state-of-the-art.

- The colour model is also changed into a new chromatic cone model instead to use the chromatic cylinder model employed in many motion segmentation approaches [25, 22, 33] . It uses chrominance angle distortion instead of chromatic distortion. For the same chromaticity line the chromatic distortion used in the above mentioned papers depends on the brightness distortion, while the chrominance angle distortion is invariant to the brightness. The invariant chromatic cone model is more robust towards chromatic shadows because these shadows (umbra shadows) modifies both the brightness and the chromaticity.

- A newly invariant gradient model is employed in order to identify the different gradients of the scene. As argued in [41, 53], the gradient model has to be invariant towards global and local illuminations changes, such as shadows. The gradient model presented in this Thesis uses a newly combination of gradient magnitudes and gradient directions which is invariant to illumination changes.

- Local information is exploited to cope with a very challenging problem such as the camouflage in chroma. Thus, in order to solve the problem the region enclosed for the colour and edge cues are combined with the illumination masks provided by the colour model. On the other hand, the intensity cue is used when pixels are beyond the dynamic range. Since, they are saturated or do not have enough chrominance, and colour and edge cues are not available. Additionally, temporal difference is also included to segment motion when these three cues are not available, such as that background not visible during the training period.

- Ghost are coped by the approach combining the segmentation obtained using all models with the segmentation obtained using edge cue and temporal difference algorithm. Furthermore, the system is able to cope with the bootstrapping problem by means of a motion filter which is iterated until convergence. The filter is used to remove moving pixels during a training set.

- As a result, our hybrid motion segmentation approach obtains very accurate and robust motion segmentation in both indoor and outdoor scenarios, as quantitatively and qualitatively demonstrated in the experimental results, by comparing

our approach with most best-known state-of-the-art approaches. The hybrid approach is based on a collaborative architecture, in which each model is devoted to specific tasks. These are performed by a particular algorithm, but they can be substituted by enhanced ones without modifying the architecture itself. Hence, this structured framework combines in a principal way the main advantage of each cue. In this way, by taking advantage of several cues, the system is allowed to benefit from all the cues capabilities, thereby simultaneously coping not only with ghosts, global and local illumination, and dark and light camouflage; but also, handling saturation, and lack of colour.

Nonetheless, motion segmentation has to deal with shadows to avoid distortions when detecting moving objects. Most segmentation approaches dealing with shadow detection are typically restricted to penumbra shadows such as the hybrid approach presented. Therefore, such techniques cannot cope well with umbra shadows. Consequently, umbra shadows are usually detected as part of moving objects.

- In this second part, we present two main novelties: (i) a bottom-up approach for detection and removal of chromatic moving shadows in surveillance scenarios. (ii) a top-down approach based on kalman filters to detect and track shadows.

- In the Bottom-up part the shadow detection approach apply a novel technique based on gradient and colour models for separating chromatic moving shadows from moving objects. Firstly, both a chromatic invariant colour cone model and an invariant gradient model are built to perform automatic segmentation while detecting potential shadows. Hereafter, the regions corresponding to potential shadows are grouped by considering "a bluish effect" and an edge partitioning. Lastly, (i) temporal similarities between local gradient structures and (ii) spatial similarities between chrominance angle and brightness distortions are analysed for all potential shadow regions, in order to finally identify umbra shadows. The resulting shadow detection can (1) detect and remove chromatic moving shadows (umbra shadows) and (2) penumbra shadows, while several other methods are restricted to the latter.

- However, in some cases the separation between a foreground object and a shadow region can fail. Occasionally, a part of the foreground object or the shadow is not accurately segmented due to segmentation problems, e.g. camouflage. Therefore, the shadow detection can miss-classify a shadow as being a part of a foreground object. In order to solve this problem a top-down approach has been developed. After detection of the objects and shadows both are tracked using Kalman filters, in order to enhance the chromatic shadow detection, when it fails to detect a shadow.

- In the Top-down part, firstly a data association between the blobs (FG and SH blobs) and the Kalman filters is performed. Secondly, an event analysis is carried out, in order to detect the different cases: object match, new object, lost object, object splitting and object merging. Taking this information into account, the Kalman filters are managed. Furthermore, occlusion handling is

managed based on a Probabilistic Based Model (PAM). Thus, based on this association temporal consistency is evaluated in the association between FGs and SHs and their respective Kalman Filters over time. Consequently, a number of cases are studied: FG and SH match, new shadow and lost shadow. Finally, the tracking results are feedback to the chromatic shadow detector to improve the object and shadow detection. Thus, chromatic shadows are correctly detected in cases with the mentioned segmentation problems.

- Thus, thanks to the data association between FG and SH we have achieved: (i) enhance the chromatic shadow detection by detecting shadows which were not possible to detect before. (ii) improve the segmentation for high level processes, such as detection and tracking, by avoiding shadows. (iii) a more robust tracking, since (1) the PAM and the KF tracker are more robust and correctly updated, and (2) erroneous created KFs are deleted.

- Qualitative and quantitative results of tests for both outdoor and indoor sequences from well-known databases validate the presented approach. Overall, our approach gives a more robust and accurate shadow detection and foreground segmentation compared to the state-of-the-art methods. Unlike other approaches, our method does not make any a-priori assumptions about camera location, surface geometries, surface textures, shapes and types of shadows, objects, and background. Experimental results show the performance and accuracy of our approach in different shadowed materials and illumination conditions.

Summarizing, in the first part of this thesis, firstly a case analysis of motion segmentation problems is presented by taking into account the problems associated with different cues, namely colour, edge and intensity. Secondly, a new hybrid approach which fuses colour, edge, intensity cues and temporal differencing to handle non-physical changes (such as global or local illumination changes), and physical changes (such as bootstrapping and ghosts) is developed [27, 25]. This architecture can cope with illumination changes, problems with the sensor dynamic range, and also with two of the three possible camouflages: camouflage in intensity and chroma. Furthermore, it can also solve the bootstrapping problem, can cope with ghosts, and can obtain a segmentation even when the background models are not available. In the second part, firstly a bottom-up approach for detection chromatic shadows is presented. Then, the contribution of this bottom-up process is threefold: (i) We combine an invariant colour cone model and an invariant gradient model to improve foreground segmentation and detection of potential shadows. (ii) We extend the shadow detection to cope with chromatic moving cast shadows by grouping potential shadow regions and considering "a bluish effect", edge partitioning, spatial similarities between local gradient structures, and temporal similarities between chrominance angle and brightness distortions. (iii) Unlike other approaches, our method does not make any assumptions about camera location, surface geometries, surface textures, shapes and types of shadows, objects, and background. Secondly, a top-down approach is presented in order to enhance chromatic shadow detection. A kalman filter is used in order to track the foreground objects and the shadows. Consequently, thanks to the data association between FG and SH we have achieved: (i) enhance the chromatic

shadow detection by detecting shadows which were not possible to detect before. (ii) improve the segmentation for high level processes, such as detection and tracking, by avoiding shadows. (iii) a more robust tracking, since (1) the PAM and the KF tracker are more robust and correctly updated, and (2) erroneous created KFs are deleted.

1.4 Document Outline

The remaining of the document is structured as follows. Next chapter presents a comprehensive state of the art related to motion segmentation. The different algorithms presented in the state of the art are analysed based on (i) the type of cues or structure employed; (ii) the method used to obtain the foreground region; and (iii) the procedure used to update the model.

In the Chapter 3, firstly a case analysis of anomalies derived from the different cues used for motion segmentation is presented. This leads to our approach to tackle segmentation. Section 3.2 explains the proposed hybrid approach, and describes how intensity, colour, edge and temporal difference are used to solve the aforementioned problems, such as camouflage in chroma, and ghosts by fusing the four cues. The experimental results for the motion segmentation approach are described in section 3.3, where our approach performance is widely analysed using indoor and outdoor sequences from several popular databases, and compared with several well-known motion segmentation approaches. Lastly, the final section concludes the main contributions of the chapter and discusses future work.

Next, Chapter 4 presents a bottom-up and top-down approaches for chromatic shadow detection. Firstly, the related methodology in the field of shadow detection is discussed in section 4.1, along with our contributions to this subject. In section 4.2, the theoretical concept of our approach is outlined. The bottom-up algorithm for foreground segmentation, along with the detection and removal of chromatic moving shadows are described in section 4.3. The top-down process used to enhance the shadow detection based on kalman filters to track the shadows is described in section 4.4. Finally, we present experimental results in section 4.5 and a brief discussion in section 4.6.

Finally, Chapter 5 presents a general discussion about the approaches and results obtained in this Thesis. For each topic related to the presented contributions we point out the remaining open issues and future directions of research.

Chapter 2

State of the Art

Detecting regions that correspond to moving objects such as vehicles and people in natural scenes is a significant and difficult problem which provides a focus of attention and simplifies the processing on subsequent analysis steps. Fast and reliable motion segmentation is an open and difficult problem due to dynamic changes in natural scenes such as global and local illumination changes, camouflages, repetitive moving objects —for instance waving flags, or moving leaves of a tree— or due to physical changes in the scene, among others.

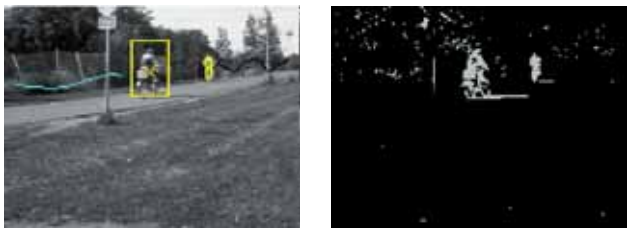
2.1 Motion Segmentation

Frequently used techniques for motion segmentation are background subtraction, frame differencing, a combination of both, or optical flow. Even though many algorithms have been proposed in the literature [15, 69, 46, 45], the problem of identifying moving objects in complex environment is still far from being completely solved.

2.1.1 Background Subtraction

Background subtraction is the most commonly used technique for motion segmentation in static scenes [42, 50, 32]. It attempts to detect moving regions in an image by differencing the current image and a reference background image in a pixel-by-pixel manner. The background model is created by averaging images over time in an initialization period. Therefore, pixels are classified as foreground where the difference is above a threshold whose calculation depends on the approach. Then, numerous approaches update over time the background model with new images to adapt it to dynamic scene changes. After this, some approaches employ a morphological post-processing operations such as erosion, dilation and closing, and also employ connected components over the foreground pixel map to reduce the effects of noise and to detected potential targets.

There are a large number of different algorithms within this basic scheme of background subtraction. Nonetheless, they differ in (i) the type of cues or structure employed to build the background model; (ii) the method used to obtain the foreground



(a) Sample frame (b) Segmentation

Figure 2.1: Sample frame based on [20] by Heikkila et al.

region; and (iii) the procedure used to update the model.

A simple version of the background subtraction scheme is employed by Heikkila and Silven [20], who classify a pixel value in the current image as foreground if it is over a predefined threshold compared with the background model:

$$|I_t - B_t| > \tau \tag{2.1}$$

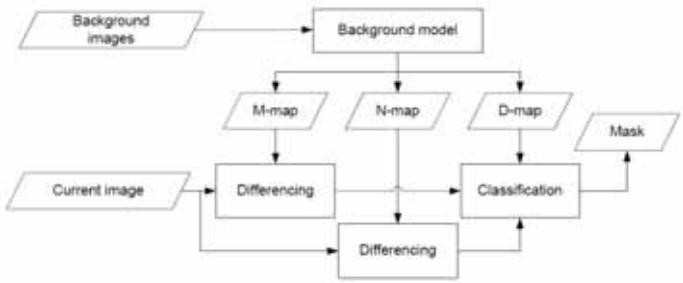
where B_t is the reference background at time t, I_t the current frame, and τ is a predefined threshold. Their approach updates the background model in order to guarantee reliable motion detection using a first order recursive filter as follows:

$$B_{t+1} = (1 - \alpha)B_t + \alpha I_t \tag{2.2}$$

where α denotes the adaptation rate that weights the current model versus the new observation. This approach performs well at obtaining the foreground moving pixels even when they stop moving. However, methods using similar schemes are extremely sensitive to changes of dynamic scenes such as gradual illumination changes, or physical changes such as ghosts. A sample frame is shown in Fig. 2.1.

In order to overcome these difficulties statistical approaches can be applied [69]. These approaches model either each pixel or group of pixels statistically. This allows building adaptive background models while providing robustness to the above-stated background conditions. Usually, model statistics are continuously updated in order to provide an adaptive approach. In order to classify if a pixel is foreground or background, authors compare current pixel values with the statistics of background model. These approaches are more efficient in front of noise, illumination changes, shadows, etc.

Haritaoglu et al. in W^4 [19] use a model of background subtraction built from order statistics of background values during a training period. The background scene is modelled by representing each pixel by three values: its minimum and maximum intensity values, and the maximum intensity difference between consecutive frames observed during this training period. Furthermore, W^4 uses a filter to exclude foreground objects during training period, such as moving people, based on median value of the pixels. Pixels are classified as foreground if the difference between the current value and the minimum and maximum values are greater than the values of the maximal interframe difference. Background model pixels are updated using the $pixel - based$ and $object - based$ updating conditions. The first condition updates



(a) Sketch Approach

(b) Sample frame

(c) Segmentation

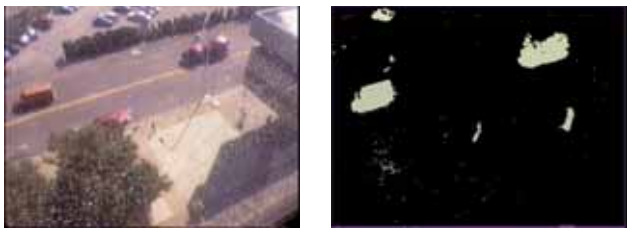
Figure 2.2: Sketch and sample frame based on [19] by Haritaoglu et al.

the background model periodically to adapt it to illumination changes in the scene, whereas the second one updates the background model to adapt it to physical changes, when new objects are deposited or removed in the background scene. Later, Neighbour pixels are grouped and blobs are classified using heuristics. Poses are identified by means of projection histograms. KFs and textural temporal templates are used to track detected targets. Therefore, the approach is able to detect and track people, isolated or in groups, in outdoor scenes, and considering several poses. However, this approach is rather sensitive to shadows and lighting changes, since the only cue used is the pixel intensity. A sketch of the algorithm and a sample frame is shown in Fig. 2.2.

In order to solve possible illumination changes and learn gradual changes in time, Wren et al. in Pfinder [72] proposed the modelling of the colour of each pixel with a single Gaussian, using YUV space. Each scene pixel is modelled using a Gaussian colour distribution. Thus, outliers are assumed to be foreground pixels, and are therefore segmented. Visible pixels are updated using a single adaptive filter. Segmented pixels are grouped into blobs and each blob is modelled using spatial and colour components. Blobs are associated with body parts using a log likelihood measure and tracked by means of of Kalman Filters (KF). However, a single Gaussian model cannot handle multiple backgrounds, such as waving trees, and the tracker just attempt to detect and track one person, in upright posture, in indoor scenes. A sample frame is shown in Fig. 2.3.

(a) Sample frame (b) Segmentation

Figure 2.3: Sample frame based on [72] by Wren et al.



(a) Sample frame (b) Segmentation

Figure 2.4: Sample frame based on [65] by Stauffer and Grimson.

Handling Background in motion

In general the above mentioned approaches obtain a good segmentation and have been used in real-time surveillance applications for long time. However they do not cope the problem of background in motion. Any approach that relies on motion to perform segmentation is liable to consider as foreground any moving background pixel, considering pixels such as the branch of a tree as a foreground. The next described approaches try to get an accurate segmentation putting special effort to solve this problem.

Stauffer and Grimson [65, 64] address the multiple backgrounds problem by using a Mixture of Gaussians to build a background colour model for every pixel. Pixels from the current frame are checked against the background model by comparing them with every Gaussian in the model until a matching Gaussian is found. If so, the mean and variance of the matched Gaussian is updated, otherwise a new Gaussian with the mean equal to the current pixel colour and some initial variance is introduced into the mixture. Moreover, the least probable Gaussian distribution is replaced if none of values match with it. Therefore, long-term still foregrounds are included. However the number of Gaussians employed has to be predefined. A sample frame is shown in Fig. 2.4. An improvement of the MoG can be found in Zivkovic et al. [77, 78], where the parameters of a MoG model are constantly updated, while also simultaneously selecting the appropriate number of components for each pixel.

Toyama et al. [66] in Wallflower use a three-component system to handle many canonical anomalies for background maintenance. Their work processes the images at various spatial scales: pixel, region, and frame levels. The pixel-level component per-

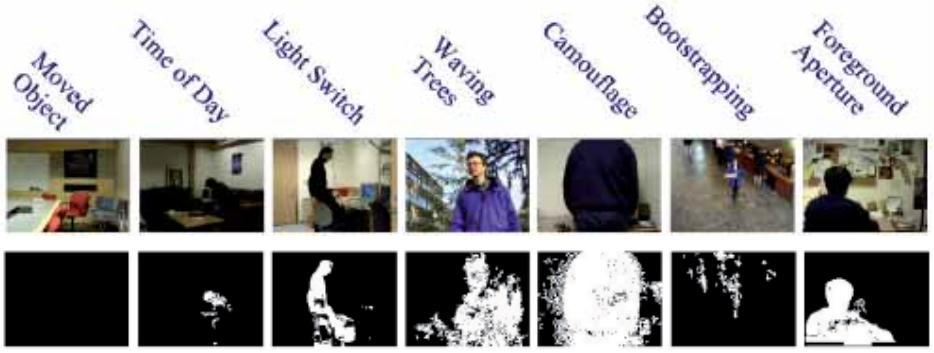


Figure 2.5: Sample frames based on [66] by Toyama et al.

(a) Sample frame (b) Segmentation

Figure 2.6: Sample frame based on [10] by Elgammal et al.

forms a Wiener filtering to make probabilistic predictions of the expected background; the region-level component fills in homogeneous regions of foreground objects; and the frame-level component detects sudden, global changes in the image and swaps in better approximations of the background. Reasonably good foreground detection is achieved in some cases, like in a scene with moving objects or with strong illumination changes (such as those caused by turning on/off the light switch). However, it fails when tackling small motion background or local illumination problems. A sample frames are shown in Fig. 2.5.

Elgammal et al. [10] use a non-parametric Kernel Density Estimation (KDE) to model the background. The model keeps a sample of intensity values for each pixel in the image and uses this sample to estimate the probability of any newly observed intensity value. The background model is updated continuously in order to adapt background changes. In addition to colour-based information, their system incorporates region-based scene information to match, not only with the corresponding pixel in the background model, but also to nearby pixel locations. This approach can handle the problem with small background motion such as tree branches. A sample frame is shown in Fig. 2.6. Mittal et al. [44] use adaptive KDE for managing background in motion. The optical flow is also used in such a work for detection of moving objects in conjunction with a normalized colour representation. In this way, the approach is able to manage complex background, but computational costs are severe.

Figure 2.7: Sample frames based on [6] by Chen et al. Segmentation in red colour.

(a) Sample frame (b) Segmentation

Figure 2.8: Sample frame based on [36] by Li et al.

Although small background motion is solved using the above mentioned motion segmentation approaches. More sophisticate approaches has been employed in order to solve more complicate dynamic backgrounds, thereby avoiding motion such as the fountain water flow, the sea waves or strong camera jitter.

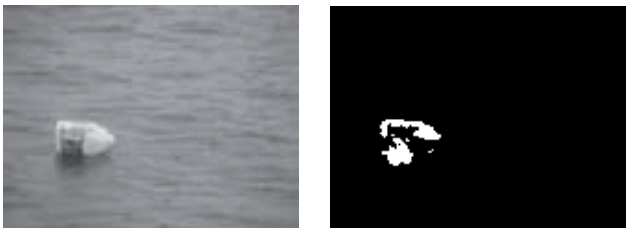
Chen et al. [6] combine pixel-based and block-based approaches to model complex background. For block-based background modelling uses contrast histograms using grey or colour images; however they have problems with camouflages and shadows. A sample frames are shown in Fig. 2.7.

Li et al. [36] and Sheikh et al. [60] use Bayesian networks to cope with dynamic backgrounds. Li et al. uses a Bayesian framework that incorporates spectral, spatial, and temporal features to characterize the background appearance. It uses colour, gradient, and temporal information based on a Bayes rules to detect foreground and background pixels. A sample frame is shown in Fig. 2.8. Sheik et al. first use non-parametric density estimation methods over a joint domain range representation to model the background as a single distribution; therefore multi-modal spatial uncertainties can be handled. Secondly, they use temporal information with the background difference. Finally, they proposes a MAP-MRF (maximum a posteriori - Markov Random Field) for object detection enforcing spatial context in the process. As a result, the algorithms can cope with dynamic backgrounds accurately. A sample frame is shown in Fig. 2.9.

Zhong et al. [76] in his approach firstly model the dynamic textures using an first-order linear model called Autoregressive Moving Average Model (ARMA). Latterly, a Kalman filter algorithm is used in estimating the intrinsic appearance of the dynamic texture. The foreground object regions are then obtained by thresholding

(a) Sample frame (b) Segmentation

Figure 2.9: Sample frame based on [60] by Sheikh et al.



(a) Sample frame (b) Segmentation

Figure 2.10: Sample frame based on [76] by Zhong et al.

the weighting function used in the robust Kalman filter. A sample frame is shown in Fig. 2.10. It works well for periodical changes in a scene but it is difficult to predict background changes with varying frequency in the natural scene.

The use of layers for image decomposition based on the neighbouring pixels is presented in [49] to avoid dynamic backgrounds. A different type of approach is used by Maddalena et al. [37] who use neural networks to overcome this problem. Mahadevan et al. [38] uses the information of the salience points and the neighbour pixels to cope the problem.

Handling Local and global illumination

The motion detection approaches described above obtain an accurate segmentation in indoor and outdoor scenarios in general and specially coping background in motion problem. Nevertheless, large number of them are susceptible to both local, such as shadows and highlights, and global illumination changes, like at dawn or dusk, or when the sun becomes covered by clouds. Most aforementioned methods fail under these circumstances. A sample frame is shown in Fig. 2.11.

There are numerous different approaches to solve global and local illuminations problems [51]. Nonetheless, they differ in the type of the cue and method employed.

Horprasert et al. [22] present a statistical background colour model which use colour chrominance and brightness distortion in RGB space. Using these distortions, this approach classifies the current pixel as original background, shaded background or shadow, highlighted background or moving foreground pixel. A sample frame is shown in Fig. 2.12. An improvement of this work is presented by Kim et al. [33] who build a cylinder to detect foreground objects in RGB space. They also quantized the

(a) Sample frame (b) Segmentation (c) Segmentation

Figure 2.11: Sample frame of background in motion and illumination change problems. a) Original frame, b) Segmentation based on [64] by Stauffer and Grimson. c) Segmentation based on [60] by Sheikh et al, the background in motion is solved but shadow problem is not coped.

(a) Sample frame (b) Segmentation

Figure 2.12: Sample frame based on [22] by Horprasert et al. Shadows in red colour.

background values for each pixel into codebooks, which represent a compressed form of background model for a long image sequences. Therefore, it can also coped part of the background in motion problem. Furthermore, the use of a layered background model can adapt it to physical changes in the scene. Nevertheless, anomalies in the dynamic range prevent to obtain an accurate segmentation. A sample frame is shown in Fig. 2.13.

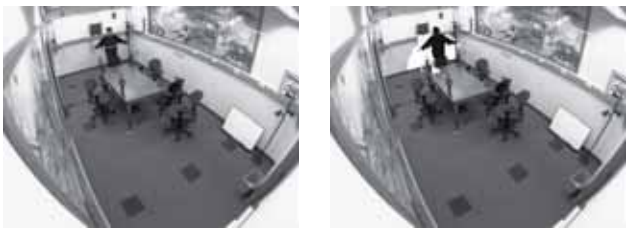
In order to avoid shadows, other spaces are used such as Cucchiara et al. [8]. They use the HSV space colour model to avoid local illumination problems. In [7], a more complex model is used which is able to detect shadows and ghosts using HSV space. The approach classifies the pixels as moving visual object, uncovered background, background, ghost, or shadow. A sample frame is shown in Fig. 2.14.

So colour is a suitable cue to handle problems with local and global illumination changes. Nevertheless, there are a lot of problems when colour is used, such as the change of illuminant. Two main approaches are employed in order to deal with this problem: colour invariant normalisations and colour constancy methods.

The first one, uses the invariant descriptors of the image to normalise the pixel values from the image to achieve the invariance. Therefore, images are transformed to such invariant colour spaces; however the structure of the colours in the image is sometimes lost. Different colour normalisation methods are chrominance coordinates, normalised chromatic coordinates, non-iterative comprehensive colour normalisation, L1L2L3 normalisation and m1m2m3 normalisation. Nevertheless, when the image is

(a) Sample frame (b) Segmentation

Figure 2.13: Sample frame based on [33] by Kim et al.



(a) Sample frame (b) Segmentation

Figure 2.14: Sample frame based on [7] by Cucchiara et al. Fg. segmentation in black and shadows in white.

transformed into this invariant colour space some intensity information is normally removed, which is needed to perform accurate motion segmentation. Furthermore, the normalisation approaches cannot normally cope with all the cases in only one normalisation. Nonetheless, the methods which are explained above sometimes are performed in motion segmentation. Vanrell et al. [67] is a good example which present a motion segmentation approach which uses a comprehensive colour normalisation based on background information to cope with illumination changes.

The second one, colour constancy methods, tries to recover the scene illuminant, in order to remove it from the images. Different colour constancy methods are based on the white estimation, the recovery of illuminant, gamut methods, Bayesian method and methods of neural nets. Nevertheless, the best colour constancy approaches need calibrated images and are high-time consuming.

Based on colour constancy methods, some approaches use the intrinsic images to remove shadows and to cope with the illuminant problem. Intrinsic image decomposition separates one image into two images: one which records variation in reflectance (reflectance image) and another which represents the variation in the illumination across the image (illumination image).

Given that shadows can cause changes in both intensity and colour illumination, Finlayson et al. [14] try to compute an invariant image which depends only on reflectance. Hence, the approach searches for a function of image chromaticity, which is invariant to the changes in illuminant colour and intensity. First, the approach find a 1D grey-scale image representation which is illuminant invariant at each image pixel and free of shadows. However, the colour information from the image is

(a) Sample frame

(b) Recovered shadow-free image

Figure 2.15: Sample frame based on [14] by Finlayson et al.

(a) Sample frame

(b) Reflectance Image

(c) Illumination Image

Figure 2.16: Sample frame based on [71] by Weiss et al.

also lost in this 1D representation. Then, they extend the 1D representation to an equivalent 2D chrominance representation which is also locally illuminant invariant and, therefore, shadow free. Finally, the approach recovers a full-colour 3D image representation which is the same as the original image but with shadows removed. Nonetheless, part of the colour information is lost when removing the effect of the scene illumination at each pixel in the image, thereby increasing camouflage problem. A sample frame is shown in Fig. 2.15 Other approaches such as Nadimi et al. [47] use a multistage approach, where the bluish effect from the illuminant scene is used plus a spatio-temporal ratio test and a dichromatic reflection model in order to remove shadows.

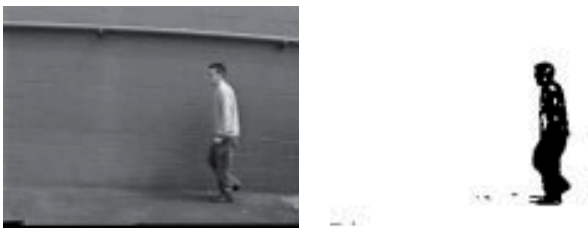
Weiss [71] tries also to extract the intrinsic images. Nevertheless, his approach uses edge cues instead of colour cues to obtain the reflectance image. This process requires several frames from a sequence to determine the reflectance edges of the scene. A reflectance edge is an edge which persists throughout 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 the position of the shadows. A sample frame is shown in Fig. 2.16

There are other approaches which use different techniques such as the normalised cross correlation to eliminate local illuminations (e.g. shadows). However, these techniques are not usually employed because of their problems with camouflage.

The edge cues are also used in some approaches for motion segmentation. Jabri et

(a) Sample frame (b) Segmentation

Figure 2.17: Sample frame based on [29] by Jabri et al.



(a) Sample frame (b) Segmentation

Figure 2.18: Sample frame based on [43] by Mckenna et al.

al. [29] use a statistical background modelling and subtraction approach which combine colour (RGB space) and edge information. The background model is computed in two distinct parts: the colour model and the edge model. On the one hand, a colour 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 colour channel, thereby yielding horizontal and vertical difference images. The background model is continuously updated. Background subtraction is performed by subtracting the colour and edge channels separately using confidence maps, and then combining the results to obtain the foreground pixels. A sample frame is shown in Fig. 2.17

McKenna et al. [43] also use colour and edge information to model the background. The background model combines pixel RGB and chrominance values with local image gradients. The motion segmentation consists of three separate background models which are combined to obtain the foreground pixels. The first model is built using the mean and variance for every channel (in RGB) to perform a typical background subtraction, thereby adapting these parameters for each new frame. A foreground mask is obtained comparing the current pixel with the model (mean and variance parameters). The second background model is computed using the mean and variance from the chrominance values for every pixel to handle the shadows. However, the use of chrominance information increases the problems with camouflage, for instance a dark green coat is not distinguishable in front of grass. To handle this kind of problem, the approach has a third background model which uses gradient information. Gradients are estimated using the sobel masks in the horizontal and vertical

(a) Sample frame (b) Segmentation

Figure 2.19: Sample frame based on [30] by Javed et al.

(a) Sample frame (b) Segmentation

Figure 2.20: Sample frame based on [21] by Heikkil*ä* et al.

directions. Therefore, the background model is represented using the gradient means and the magnitude variance for every pixel. The foreground is detected comparing the current pixel with the edge model parameters. Therefore, a pixel is considered foreground if either chrominance or gradient information supports that classification. However, hard-edge shadows are still segmented as foreground and the foreground region segmented contain holes. To handle problem with holes, the approach use the foreground mask obtained with the first model to fill them. A sample frame is shown in Fig. 2.18

Javed et al. [30] present a method that uses multiple cues, based on colour 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. At the pixel level, two statistical models of gradients and colour 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 colour 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 colour based subtraction results if more than 50 percent of the results are considered foreground, thereby using only gradient subtraction results to handle detect global illumination changes. Nevertheless, ghosts are not eliminated if the background contains a high number of edges, while shadows are not eliminated either. A sample frame is shown in Fig. 2.19

Figure 2.21: Sample frame based on [30] by Javed et al. Shadows in white colour.

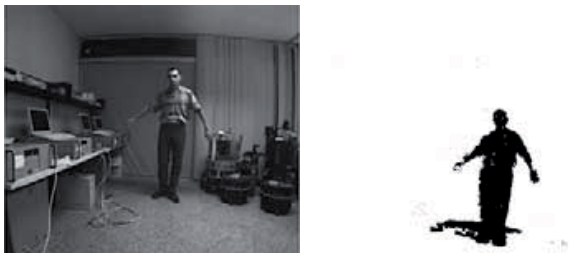
Other approaches use textures to avoid global and local illumination problems. Heikkil*ä* et al. [21] use textures to model the background. The texture features are extracted by using a modified Local Binary Pattern operator (LBP) trying to cope with some of the motion segmentation problems such as background in motion and global illumination changes. A sample frame is shown in Fig. 2.20. Leone et al. [35] use texture descriptors based on the coefficients of a Gabor functions decomposition and photometric properties in order to solve shadows. A sample frame is shown in Fig. 2.21 However, these methods are noise-dependent, the noise in the scene can do that the textures fails. An improvement of this approach is carried by Yao et al. [74], where the textures obtained using the LBP are combined with a colour model based on the RGB space. Another approach which uses textures is proposed by Amato et al. [1], which uses chromacity information plus textural information in order to avoid local and global illumination changes. However, texture based approaches usually suffer of serious failures with camouflages and local illuminations (e.g shadows). Furthermore, the selection of the size of patches is an intrinsic problem of the textural approaches because it is dependent of the scene and the objects in the scene.

2.1.2 Temporal Differencing

Approaches based on temporal difference attempt to extract moving regions by making use of a pixel-by-pixel difference between consecutive frames in a video sequence [69]. This kind of method is very adaptive to dynamic scene changes. Nevertheless, it generally fails to extract the entire relevant pixels of moving objects, thereby causing foreground aperture problem. This method also cannot cope with sleeping objects problem. Approaches based on this method normally incorporate additional techniques in order to detect these stopped objects. A typical temporal differencing scheme consists of comparing the current frame with the last frame. Therefore, the pixels are considered foreground if the difference is over a threshold: Basic Schema Temporal difference

$$|I_t - I_{t-1}| > \tau \tag{2.3}$$

The segmentation results depend only on the threshold method used for binarization. To improve typical temporal difference scheme, some approaches use a three-frame differencing. Furthermore, to overcome typical temporal differencing defects, some approaches use hybrid algorithms which combine a three-frame differencing with an adaptive background subtraction model.



(a) Sample frame (b) Segmentation

Figure 2.22: Sample frame based on [62] by Spagnolo et al.

Shen [61] is an example of hybrid algorithm which uses RGB, HSI colour spaces, fuzzy information and temporal difference techniques to achieve motion segmentation. The segmentation is executed in two steps. In the first step, a fuzzy classification is carried out by considering the mobility of pixels which is generated by combining the results from separately thresholded difference images of each RGB channel. In the second step, falsely detected pixels from the first step are eliminated by using the previous segmentation result and the motion information obtained from successive frames. Finally, the HSI colour space is used to eliminate shadows. Spagnolo et al. [62] also use temporal information to get an accurate segmentation. The approach combines the use of radiometric similarity between regions to compare pixels, both in the temporal image analysis and in the background subtraction. A sample frame is shown in Fig. 2.22

2.1.3 Optical Flow

Lastly, motion segmentation based on optical flow uses characteristics of flow vectors of moving objects over time to detect change regions in an image sequence. These methods can segment moving objects in video sequences even from a moving camera. However, most of these methods are computationally highly expensive and very sensitive to noise. Moreover, most of them cannot be executed in real-time without specialized hardware [69]. The approach from Mittal et al. [44] above mentioned in the background subtraction approaches use a hybrid approach which combine optical flow in conjunction with a normalized colour representation in order to get the motion segmentation. A sample frame is shown in Fig. 2.23

Another approach which also uses MAP-MRF is presented by Bugeau et al. [5]. Firstly, the camera motion is computed and the images rectified. Secondly, the approach restrict momentarily the analysis to a subgrid of "moving" pixels (i.e. not belonging to camera motion) defining a descriptor to characterize them. The descriptor is formed by three different groups of features: the first group is composed of the coordinates of the point, the second group contains its motion, and the last one contains discriminant photometric features. Thirdly, the pixels selected are merged into clusters consistent for both colour and motion using a Mean Shift algorithm with automatic multidimensional bandwidth selection. Finally, from the clusters, the complete pixel-wise segmentation of moving objects is found using a MAP-MRF framework.

(a) Sample frame (b) Segmentation

Figure 2.23: Sample frame based on [44] by Mittal et al.

Figure 2.24: Sample frames based on [5] by Bugeau et al. Segmentation in red colour.

The use of spatial, dynamic and photometric features allows the extraction of moving foreground objects even in presence of illumination changes and fast variations in the background. However, as the approach only work on a subgrid of pixels, and because it do not model the background, it is not able to get an accurate segmentation of the moving object. A sample frames are shown in Fig. 2.24.

2.1.4 Discussion

Once the advantages and drawbacks of existing approaches have been detailed, we can identify the main contributions of our motion segmentation approach w.r.t. the state of the art:

A novel theoretical case analysis of motion segmentation problems is presented, where the performance of each cue used in the literature for segmentation (intensity, colour, and edges) is exhaustively evaluated, showing the advantages of every cue and when a cue can be or cannot be applied. To the best of our acknowledge, current state-of-the-art considers chromatic spaces only, and they do not address most of the problems identified in our case analysis.

Our hybrid algorithm uses intensity, colour, edges cues and temporal difference, because each cue solves a particular problem identified in the case analysis. Cue models have been improved over existing ones, furthermore their combination is a step forward the current state-of-the-art.

The Chromatic invariant cone model achieves better segmentation results than the commonly used cylinder model [22, 33]. The invariant gradient model combines

magnitudes and orientations for edge segmentation while avoiding false edges due to intense global illumination changes [43, 29].

Using chromaticity only, the assessment of whether a foreground region is a shadow, a change of global illumination or a dark camouflage is not possible [22, 33]. Approaches using HSV [8, 61] also exhibit this problem.

Other techniques are not be able to work with shadows, highlights and global illumination changes [65, 19, 72]. We cope with those problems without a significant increase of computational cost. Other approaches fuse colour and edges cues without addressing shadow removal [29, 30].

Using the combination of cues, we are able to solve the ghost problem on-the-fly, instead of requiring a predefined time period or not cope the problem [22, 33, 10, 66, 43, 29].

Finally, the resulting shadow detection can (1) detect and remove chromatic moving shadows (umbra shadows) and (2) penumbra shadows, while several other methods are restricted to the latter.

Chapter 3

Enhancing Motion-based Segmentation

Detection of regions that correspond to moving objects such as vehicles and people in natural scenes is a significant and difficult problem. But efficient segmentation simplifies the processing on subsequent steps of analysis [17]. The information obtained from this step is the basis for a wide range of applications such as smart surveillance systems, control applications, advanced user interfaces, and motion basis diagnosis, among others [15]. Nevertheless, fast and reliable motion segmentation is an open and difficult problem due to dynamic changes in natural scenes such as global and local illumination changes (i.e. shadows and highlights), camouflages, repetitive moving objects (waving flags, moving tree leaves) or due to physical changes such as bootstrapping, and ghosts, among others [66]. Frequently used techniques for motion segmentation are background subtraction, frame differencing, a combination of both, or optical flow [15, 69, 46, 45, 52]. Even though many algorithms have been proposed in the literature, the problem of identifying moving objects in complex environment is still far from being completely solved.

In this chapter, a novel approach which overcomes most known techniques used for motion segmentation is proposed. The main advantages of the our approach are: (i) The novel theoretical case analysis, most employed cues in the literature for motion segmentation are exhaustively analysed in order to find their advantages and drawbacks. (ii) A new architecture, based on such an analysis, so that a new hybrid approach which fuses colour, edge, intensity cues and temporal difference has been developed [27, 25]. (iii) The proposed method can handle non-physical changes (such as global or local illumination changes and camouflages), physical changes (such as bootstrapping and ghosts), and sensor dynamic range problems. (iv) Models are also improved: a chromatic invariant cone model enhances colour segmentation; an invariant gradient model combining magnitude and orientation improves edge segmentation avoiding false edges (due to intense global illumination changes). Furthermore, (v) our approach is able to detect dark camouflages, which is distinguished from shadows and changes of global illumination. Our technique can also (vi) detect ghost problems on-the-fly without increasing the computational cost. Real-time processing can

be achieved because the method can be parallelizable, due to the pixel-wise nature of the approach.

The contribution in this chapter is organized as follows. The next chapter presents a case analysis of anomalies derived from the different cues used for motion segmentation. This leads to our approach to tackle segmentation. Section 3.2 explains the proposed algorithm, and describes how intensity, colour, edge and temporal difference are used to solve the aforementioned problems, such as camouflage in chroma, and ghosts by fusing the four cues. The experimental results are described in section 3.3, where our approach performance is widely analysed using indoor and outdoor sequences from several popular databases, and compared with several well-known motion segmentation approaches. Lastly, final section concludes this contribution and discusses future work.

3.1 A Case Analysis of motion segmentation problems

Colour Information obtained from a recording camera is based on the sensor response s^c —for Lambertain or perfect matte surfaces— and depends on three components: the illuminant spectral power distribution $L(\lambda)$, the object reflectance distribution $R(\lambda)$, and the sensor sensitivity $S^c(\lambda)$, following the equation:

$s^c = \int_\lambda L(\lambda) R(\lambda) S^c(\lambda) d\lambda$,

where λ denotes the wavelength, and $c \in \{R, G, B\}$ the colour channel, see Fig. 3.1. Therefore, changes in the illumination —in both brightness and chrominance components— modify the sensor response, see Fig. 3.2. The object reflectance may considerably depend on the both the incident-light angle and the viewing angle. It also may present strong specular components with no information about the object colour. Finally, it depends on the sensor sensitivity, see Fig. 3.3.

In addition, the sensor dynamic range must be taken into account. This is defined as the ratio between the maximum possible signal versus the noise signal in dark. Thus, very low or very high brightness distort the observed response. Consequently, these effects should be considered as a source of potential errors during both background modelling and image segmentation.

Very dark pixels are beyond the sensor dynamic range, since they do not have enough brightness for reliably compute their chrominance. A similar problem appears with very light pixels, which have at least one channel component saturated.

A series of experiments with a Macbeth board were designed to explore these phenomena, see Fig. 3.4. Experiments show as a wrong background model may be built depending on the illumination conditions during the training step of the background model (red line in Fig. 3.4). A Macbeth board was first illuminated with a constant light source. Then, the diaphragm was modified in a series of time steps, thereby changing the received luminance. The background was modelled during 50 frames. Then, 650 more frames were acquired while changing the aperture.

Fig. 3.5 shows a case analysis of the potential segmentation problems using the combination of three background models: colour, edges and intensity, and the pixel value within the sensor dynamic range.

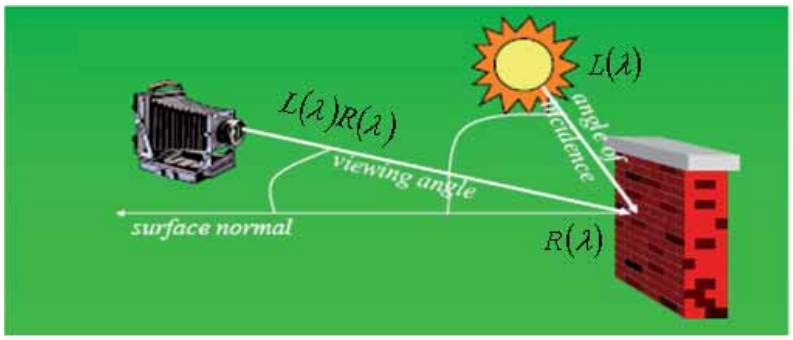


Figure 3.1: Sensor response. The sensor response depends on the illuminant wavelength, and on the object reflectance, apart from the sensor sensitivity. (Figure modified from CS410 notes, Draper, 2006).

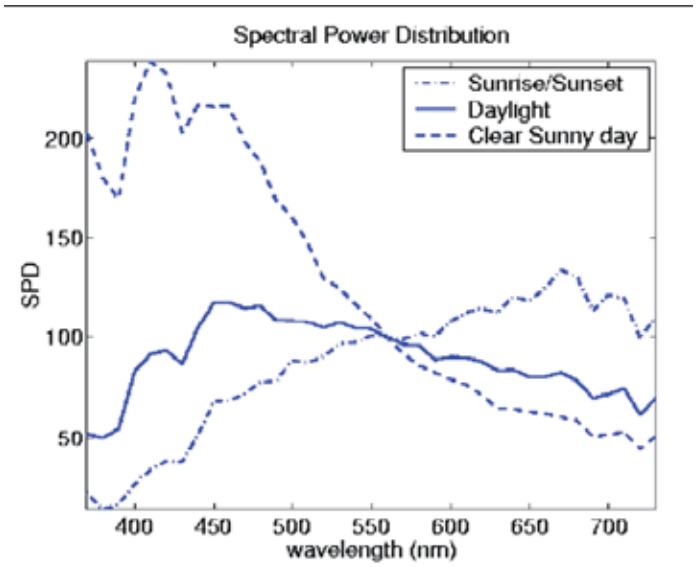


Figure 3.2: Illuminant Spectral Power Distribution. The illuminant SPD may vary, thereby affecting the observed colour. (Figure modified from CS320 notes, Jepson, 2005).

Edges from very dark pixels with not enough brightness can be hidden since they are beyond sensor dynamic range. And a similar problem appears with very light pixels. Consequently, cases beyond sensor dynamic range should be addressed using an intensity model, because both colour and edge models are not suitable, thereby classifying the pixels as foreground (case FgI) or background (case BgI) depending on their intensity.

There could be pixels whose Bg. colour model can be computed, although the

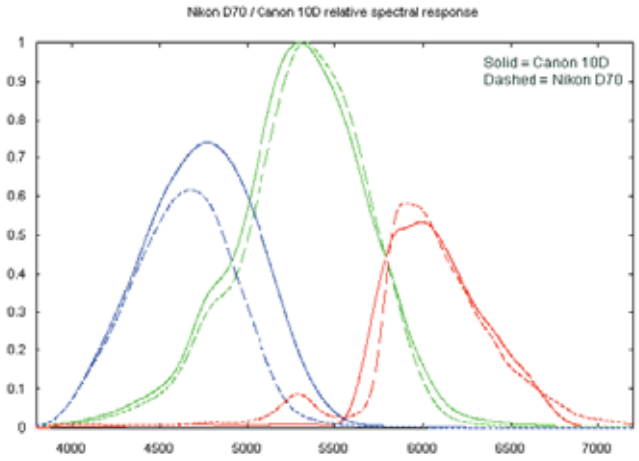


Figure 3.3: Sensor sensitivity. Different sensors present a different response to the same stimulus. (Figure from http://astrosurf.com/build/70v10d/eval_htm).

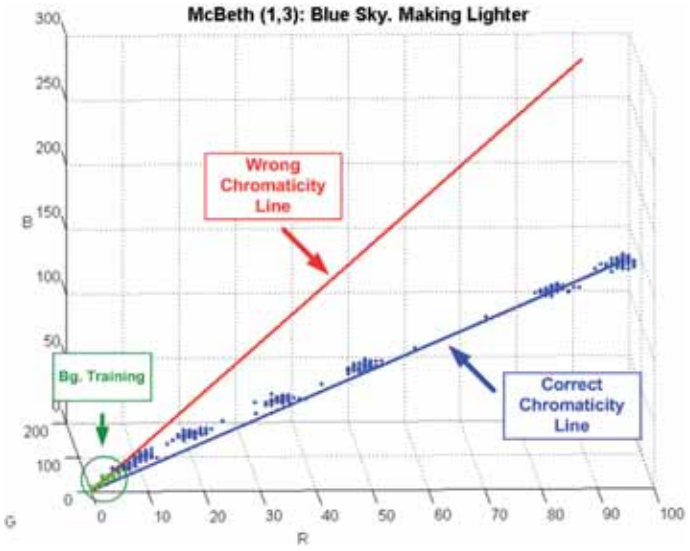
current image pixels are beyond the sensor dynamic range. Here, neither chrominance nor edge cues can be used. In such a case, the brightness component of the colour model can be used as a suitable cue, thereby classifying them as dark/light foreground (case DF/LF) or background (case BB).

Changes in illumination, despite of being local or global, sudden or gradual (such as shadows or highlights) are all supposed to entail just variations in the observed brightness, but not in the chrominance. Thus, a pixel can be considered as foreground using colour and edge models in the following situations: (i) a pixel is considered foreground using the colour model when it differs in chrominance with the model (case FgC); (ii) using the edge model when it shows a gradient change respect to the model (case FgE). Otherwise the pixel is classified as background (cases BgE, shadow (S) or highlight (H)).

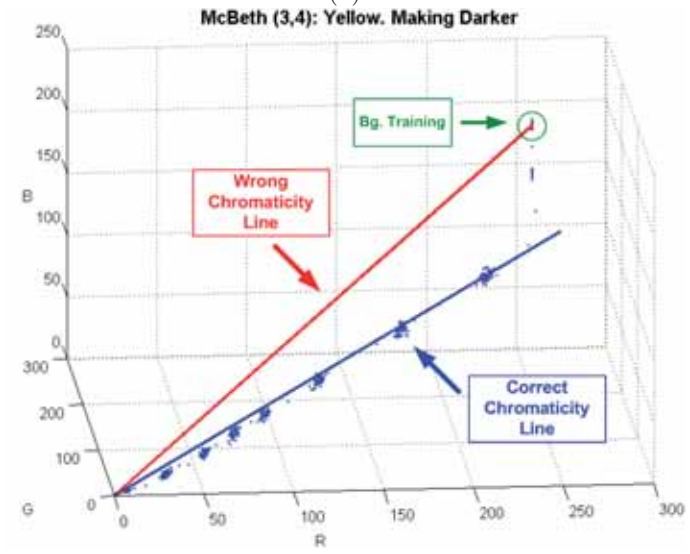
Foreground pixels whose lower and higher brightness cannot be distinguished from shadows and highlights, are considered dark/light camouflage (DC/LC, respectively).

Hence, fusing the three models may overcome some of the segmentation problems such as changes in illumination conditions, camouflage in intensity and camouflage in chroma, as long as the illuminant has a plain spectral power distribution.

However, there are other *anomalies* that cannot be disambiguated with the colour, edges and intensity cues, which are not taken into account in this thesis. Firstly, foreground pixels with the same chrominance, brightness, and gradient as the background model can not be segmented, so such pixels are considered camouflaged (CaC, CaB, and CaE respectively). Secondly, intense shadows and highlights (IS/IH) can be classified as DF or LF, and shadows and highlights (S/H) over zones beyond the sensor dynamic range can be considered as foreground (FgI). Thirdly, edges of sharp shadows and highlights (SS, SH) can be segmented (FgE). Finally, local and global changes in the illuminant chrominance (CI), as well as gleaming surfaces (GS) may cause false-positive segmentations. So there is still a lot of ground to cover.



(a)



(b)

Figure 3.4: Experiments on a Macbeth board to test the sensor dynamic range. Background pixels are drawn in green. The red line denotes the modelled chrominance line, whereas the blue one corresponds to the correct one. (a) This corresponds to a blue checker which is not observed with enough light during the modelling process. (b) In this case, the chrominance of a yellow checker is modelled while some of the channels are saturated. Consequently, there are noticeable deviations between the inferred and correct chrominance in both cases.

Case Analysis: Bg.Model vs. Image Comparison (pixel wise)

Model	Range ▶ Cues ▼	Model BSDR		Image BSDR			Image ISDR						
BGM	Chrom.	x	x	x	x	x	Similar						Diff.
	Brightness	x	x	Lower	Similar	Higher	Lower		Similar		Higher		x
BIM	Int	Sim.	Diff.	-	-	-	-		-		-		-
BEM	Edges	x	x	x	x	x	Sim.	Diff.	Sim.	Diff.	Sim.	Diff.	-
Classif ication	Case	BgI	FgI	DF	BB	LF	S	FgE	BgE	FgE	H	FgE	FgC
	Anomalies	CaI	S H CI GS	IS CI GS	CaB	IH CI GS	DC CaE	SS	CaC CaE		LC CaE	SH	CI GS

Figure 3.5: Labelling: Beyond (BSDR) or Inside (ISDR) Sensor Dynamic Range; Shadow (S), Highlight (H), Background using Chrominance (BgC), Brightness (BB), Edges (BgE) or Intensity cues (BgI); Foreground using Chrominance (FgC), Edges (FgE) or Intensity (FgI), Dark Foreground (DF) and Light Foreground (LF); Camouflage using Chrominance (CaC), Edges (CaE), Brightness (CaB) or Intensity (CaI), Dark Camouflage (DC), Light Camouflage (LC), Sharp Shadows (SS), Sharp Highlight (SH), Intense Shadows (IS), Intense Highlight (IH), Change of Illuminant (CI), Gleaming Surface (GS). Cues: **'x'** it cannot be used; **'-'** it is no relevant. See text for details.

3.2 Multicue Image Segmentation

The segmentation task is next presented following a statistical background-subtraction approach based on the case analysis presented before. Our approach addresses the analysed cases by combining three background models and a temporal difference algorithm.

Firstly, background models are built and automatic threshold selection for them are computed. Next, image segmentation using these models is presented and finally, an approach to combine the models to obtain an accurate segmentation solving camouflage in chroma and ghosts is showed.

3.2.1 Background Modelling

The approach combines three background models and a temporal difference algorithm. A sketch of the Background-Modelling Module is shown in Fig 3.6. A Background Colour Model (BCM) will consist of a chromatic invariant cone representation which separates both chrominance and brightness component; a Background Edge Model (BEM) will make use of magnitudes and invariant gradient orientations; a Background Intensity Model (BIM) will compute the mean and standard deviation for each pixel intensity; and a Temporal Differencing (TD) algorithm will evaluate the changes between three consecutive frames.

The background is modelled on a pixel-wise basis [65, 19, 43], which provides the necessary representation accuracy. Training is carried out by using a window of T frames. A motion filter $\left|I^c_{a,t} - \widetilde{I}^c_a\right| < \max\left(\kappa_m \sigma^c_a, \epsilon\right)$ is used to remove moving pixels during a training set of T frames, where $I^c_{a,t}$ and $\widetilde{I}^c_a$ are the current image value and median value of pixel $'a'$ for each channel $c \in \{R, G, B\}$ respectively, σ^c_a is the correspondent standard deviation, κ_m sets the confidence region, and ϵ is a small positive quantity. This process is iterated until convergence. Then, just those pixels with a representative number of valid values in the T frames are taken into account for background modelling.

Pixel values of colour, edge and intensity obtained from motion filter are used to build the background models. On the one hand, pixels whose RGB values are beyond the dynamic range of the sensor are used to build BCM and BEM. On the other hand, pixels values beyond the sensor dynamic range are used to build BIM. Those pixels considered in motion are not valid to build any background model and will be evaluated using a temporal difference algorithm.

Background Colour Model (BCM)

The BCM is computed according to the chromatic-invariant cone representation shown in Fig. 3.7: first, the RGB mean $\boldsymbol{\mu}_a = \left(\mu^R_a, \mu^G_a, \mu^B_a\right)$ and standard deviation $\boldsymbol{\sigma}_a = \left(\sigma^R_a, \sigma^G_a, \sigma^B_a\right)$ of every image pixel a during the time period $t = [1 : T_1]$ are computed.

Once each RGB component is normalised by their respective standard deviation σ^c_a, two distortion measures are established during the training period: the brightness distortion, $\alpha_{a,t}$, and the chrominance angle distortion, $\beta_{a,t}$. The brightness distortion

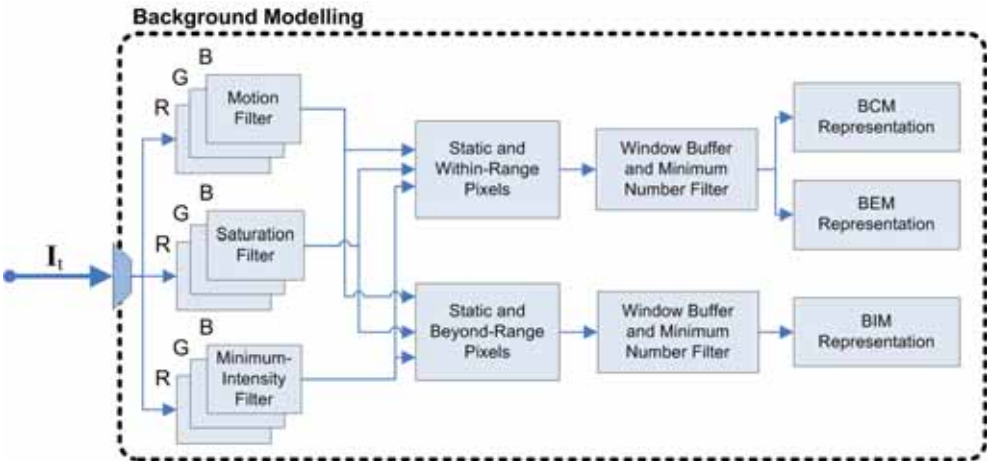


Figure 3.6: Background modelling approach. See text for details.

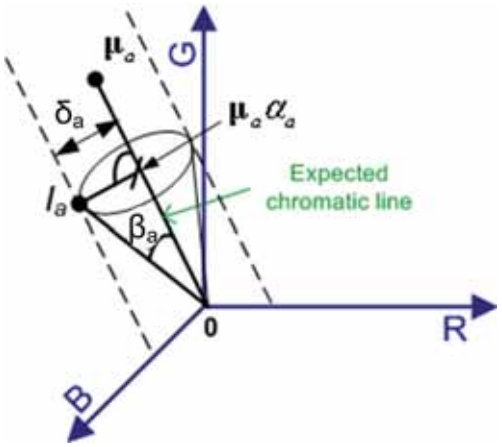


Figure 3.7: Colour-model representation. $\boldsymbol{\mu}_a$ represents the expected RGB colour value for a pixel a, while $\mathbf{I}_a$ is the current pixel value. The line $\overline{\mathbf{0}\boldsymbol{\mu}_a}$ shows the expected chromatic line —all colours along this line have the same chrominance, but different brightness. α_a and β_a give the current brightness and chrominance angle distortion, respectively.

can be computed by minimising the distance between the current pixel value $\mathbf{I}_{a,t}$ and the chromatic line $\overline{\mathbf{0}\boldsymbol{\mu}}_a$. The angle between $\overline{\mathbf{0}\boldsymbol{\mu}}_a$ and $\overline{\mathbf{0}\boldsymbol{I}}_a$ is, in fact, the chromatic angle distortion. Thus, the brightness and the chromatic angle distortions are given by:

$$\alpha_{a,t} = \frac{\frac{I^R_{a,t}\mu^R_a}{(\sigma^R_a)^2} + \frac{I^G_{a,t}\mu^G_a}{(\sigma^G_a)^2} + \frac{I^B_{a,t}\mu^B_a}{(\sigma^B_a)^2}}{\left(\frac{\mu^R_a}{\sigma^R_a}\right)^2 + \left(\frac{\mu^G_a}{\sigma^G_a}\right)^2 + \left(\frac{\mu^B_a}{\sigma^B_a}\right)^2}, \tag{3.1}$$

$$\beta_{a,t} = \arcsin \frac{\sqrt{\sum_{c=R,G,B}\left(\frac{I^c_{a,t} - \alpha_{a,t}\mu^c_a}{\sigma^c_a}\right)^2}}{\sqrt{\sum_{c=R,G,B}\left(\frac{I^c_{a,t}}{\sigma^c_a}\right)^2}}. \tag{3.2}$$

Finally, the Root Mean Square over time of both distortions for each pixel is computed: $\bar{\alpha}_a$ and $\bar{\beta}_a$, respectively:

$$\bar{\alpha}_a = RMS\left(\alpha_{a,t} - 1\right) = \sqrt{\frac{1}{T_1}\sum_{t=0}^{T_1}\left(\alpha_{a,t} - 1\right)^2}, \tag{3.3}$$

$$\bar{\beta}_a = RMS\left(\beta_{a,t}\right) = \sqrt{\frac{1}{T_1}\sum_{t=0}^{T_1}\left(\beta_{a,t}\right)^2}, \tag{3.4}$$

where 1 is subtracted to $\alpha_{a,t}$, so that the brightness distortion is now distributed around zero: positive values mean brighter pixels, whereas negative ones mean darker pixels, with regard to the learnt values. These values are used as normalising factors so that a single threshold can be set for the whole image. This 4-tuple $BCM =< \boldsymbol{\mu}_a, \boldsymbol{\sigma}_a, \bar{\alpha}_a, \bar{\beta}_a >$ constitutes the pixel colour background model.

Background Edge Model (BEM)

The BEM is built as follows: first the Sobel edge operator is applied to each colour channel in horizontal and vertical directions. This yields both horizontal $G^c_{x,a,t} = S_x * I^c_{a,t}$ and vertical $G^c_{y,a,t} = S_y * I^c_{a,t}$ gradient image for each frame during the training period $t = [1:T]$, where $c \in \{R, G, B\}$ denotes the colour channel.

Then, each background pixel gradient is modelled using the gradient mean $\mu_{Gx,a} = (\mu^R_{Gx,a}, \mu^G_{Gx,a}, \mu^B_{Gx,a})$ and $\mu_{Gy,a} = (\mu^R_{Gy,a}, \mu^G_{Gy,a}, \mu^B_{Gy,a})$, and gradient standard deviation $\sigma_{Gx,a} = (\sigma^R_{Gx,a}, \sigma^G_{Gx,a}, \sigma^B_{Gx,a})$ and $\sigma_{Gy,a} = (\sigma^R_{Gy,a}, \sigma^G_{Gy,a}, \sigma^B_{Gy,a})$ computed from all the training frames for each channel.

Then, the magnitudes of the gradient mean μ_G and standard deviation σ_G are computed in order to build the background edge model. The orientation of the gradient (μ_θ and σ_θ) is also computed to avoid false edges created by illumination changes.

$$\mu^c_{G,a} = \sqrt{(\mu^c_{Gx,a})^2 + (\mu^c_{Gy,a})^2}; \quad \sigma^c_{G,a} = \sqrt{(\sigma^c_{Gx,a})^2 + (\sigma^c_{Gy,a})^2}, \tag{3.5}$$

$$\mu^c_{\theta,a} = \arctan\left(\frac{\mu^c_{Gy,a}}{\mu^c_{Gx,a}}\right); \qquad \sigma^c_{\theta,a} = \arctan\left(\frac{\sigma^c_{Gy,a}}{\sigma^c_{Gx,a}}\right), \tag{3.6}$$

where $c \in \{R, G, B\}$ denotes the colour channel. Thus, $BEM =< \mu^c_{G,a}, \sigma^c_{G,a}, \mu^c_{\theta,a}, \sigma^c_{\theta,a} >$.

Background Intensity Model (BIM)

Finally, the BIM consist on a 2-tuple given by the mean pixel intensity, μ^I_a and its standard deviation σ^I_a. It is computed for those non-in-motion pixels which have a representative number of values beyond sensor dynamic range. So, $BIM =< \mu^I_a, \sigma^I_a >$.

3.2.2 Automatic Threshold Selection

The thresholds employed for the segmentation task are automatically computed for each model, as shown next.

Background Colour Model (BCM)

The BCM is completed by an automatic threshold computation for a given detection rate when a new frame is acquired. First, the normalised distortions are calculated for each pixel:

$$\breve{\alpha}_{a,t} = \frac{\alpha_{a,t}}{\bar{\alpha}_a}; \; \breve{\beta}_{a,t} = \frac{\beta_{a,t}}{\bar{\beta}_a}. \tag{3.7}$$

This process is repeated during a temporal window of T_2 frames in order to avoid errors due to an insufficient number of samples. Subsequently, the histograms of both accumulated measures $\breve{\alpha}_{a,t}$ and $\breve{\beta}_{a,t}$ are computed taking into account all pixel distortions during the temporal window. Detection rates are used to set lower and higher brightness distortion thresholds, $\tau_{\alpha 1}, \tau_{\alpha 2}$, and a chrominance threshold, τ_β, which can be over one to give a confidence region, since the motion filter has eliminated the outliers and restrict the confidence region for background pixels.

Two thresholds $\tau_D = \kappa_D \tau_{\alpha 1}$ and $\tau_L = \kappa_L \tau_{\alpha 2}$ are set for both dark and light foreground cases, where the current pixel is beyond the sensor dynamic range. Usually $\kappa_D = \kappa_L = \kappa$ is a factor that specifies the confidence region. Summarizing, the BCM thresholds are $BCM_\tau =< \tau_{\alpha 1}, \tau_{\alpha 2}, \tau_\beta, \tau_D, \tau_L >$.

Background Edge Model (BEM)

The BEM uses three thresholds for edge pixel segmentation. A minimum magnitude gradient threshold (τ_e) is learnt to know when an edge can be compared using its oriented gradient. An oriented gradient threshold (τ_θ) and maximum magnitude gradient threshold (τ_G) are learnt for pixel segmentation according to BEM. The

thresholds are computed as $\tau^c_{e,a} = max(\kappa_e \sigma^c_{G,a}, \epsilon)$, $\tau^c_{\theta,a} = max(\kappa_\theta \sigma^c_{\theta,a}, \overline{\sigma}^c_{\theta,a})$, and $\tau^c_{G,a} = max(\kappa_G \sigma^c_{G,a}, \overline{\sigma}^c_{G,a})$, where κ_e,κ_θ, and κ_G are the factors that set the confidence region, and $\kappa_e << \kappa_G$; and $\overline{\sigma}^c_{\theta,a}$ and $\overline{\sigma}^c_{G,a}$ are the average standard deviation computed over the entire image to set a minimum positive quantity. Summarizing, $BEM_\tau =< \tau^c_{e,a}, \tau^c_{\theta,a}, \tau^c_{G,a} >$.

Background Intensity Model (BIM)

The threshold used for pixel segmentation according to BIM is computed as $\tau^I_a = \max\left(\kappa^I \sigma^I_a, \epsilon\right)$, where κ^I is the factor that sets the confidence region, and ϵ is a small positive quantity. So, $BIM_\tau =< \tau^I_a >$.

Temporal Differencing (TD)

Finally, the threshold for temporal differencing segmentation is automatically computed for a given detection rate. A standard deviation is calculated during the three first frames to evaluate. The histogram of accumulated measures is computed, taking into account all pixel standard deviation during the three frames. Detection rate is used to set threshold τ_r, avoiding outliers. The threshold is finally computed as $\tau_T = \max\left(\kappa_T \tau_r, \epsilon\right)$, where κ_T is the factor that sets the confidence region, and ϵ is a small positive quantity. Thus, $TD_\tau =< \tau_T >$.

3.2.3 Image Segmentation

The segmentation task is done in two steps. The first step is used to obtain the foreground regions for every model, and the second step is used to fuse the results achieved in the first step to cope the camouflage in chroma.

Input images can now be segmented by classifying the pixels according to computed background model and the current sensor response. A sketch of the Image-Segmentation Module is shown in Fig 3.8.

Thus, in the first step four general cases are considered, and a different model is applied in each one:

- BCM and BEM are applied to those pixels whose current values are inside the sensor dynamic range, and for which BCM and BEM could be built;
- the brightness component of BCM is applied to segment those pixels whose current values are beyond this range and have BCM;
- BIM is applied to those pixels which do not have enough values within the dynamic sensor range during the modelling process.
- and, TD is applied to those pixels whose background was not visible during the training period and there is no background model available.

As a result, a first step segmentation map $\mathbf{M}_{a,t}$ is computed at each time. Thus, pixels under the first condition are classified using BCM as background (BgC), highlight (H), shadow (S), or foreground (FgC); and using BEM as background (BgE),

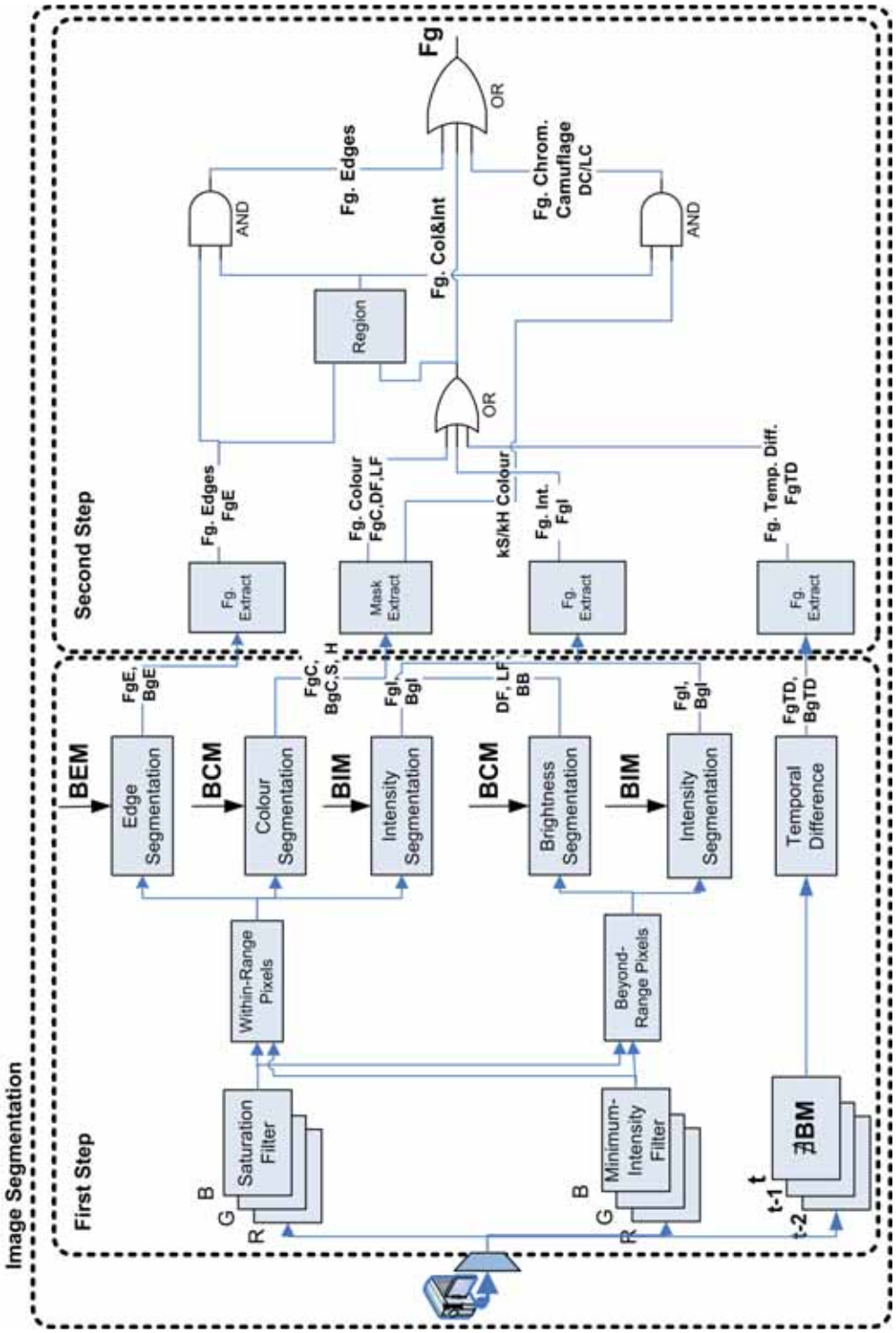


Figure 3.8: Image segmentation approach. As a result of applying background models to the current frame, pixels are classified in the first step according to the sensor dynamic range using BCM as foreground (FgC), background (BgC), shadow (S), and highlight (H); according to the BEM as foreground (FgE) and background (BgE); using the BCM on pixels beyond the sensor dynamic range, as dark foreground (DF), light foreground (LF), and background (BB); according to the BIM as foreground (FgI) and background (BgI); and according to the TD as foreground (FgTD) and background (BgTD). In a second step, pixels inside the region enclosed by the foregrounds from the first step are combined with a thresholded S and H mask in order to segment the foregrounds dark (DC) and light (LC) camouflage.

foreground (FgE). Those pixels under the second condition are classified as background (BB), or dark foreground (DF) and light foreground (LF); those under the third one as background (BgI) or foreground (FgI); and those under the last one as background (BgTD) or foreground (FgTD). The whole process is summarized according to the following equation:

$$\mathbf{M}_{a,t} = \tag{3.8}$$

$$= \begin{cases} \text{BgC} : \exists\text{BCM} \wedge \tau_m < I^c_{a,t} < \tau_n \wedge \tau_{\alpha 1} < \breve{\alpha}_{a,t} < \tau_{\alpha 2} \wedge \breve{\beta}_{a,t} < \tau_\beta \\ \text{S} : \exists\text{BCM} \wedge \tau_m < I^c_{a,t} < \tau_n \wedge \breve{\alpha}_{a,t} < \tau_{\alpha 1} \wedge \breve{\beta}_{a,t} < \tau_\beta \\ \text{H} : \exists\text{BCM} \wedge \tau_m < I^c_{a,t} < \tau_n \wedge \breve{\alpha}_{a,t} > \tau_{\alpha 2} \wedge \breve{\beta}_{a,t} < \tau_\beta \\ \text{FgC} : \exists\text{BCM} \wedge \tau_m < I^c_{a,t} < \tau_n \wedge \breve{\beta}_{a,t} > \tau_\beta \\ \text{BgE} : \exists\text{BEM} \wedge \tau_m < I^c_{a,t} < \tau_n \wedge \neg(F_\theta \vee F_G) \\ \text{FgE} : \exists\text{BEM} \wedge \tau_m < I^c_{a,t} < \tau_n \wedge F_\theta \vee F_G \\ \text{BB} : \exists\text{BCM} \wedge I^c_{a,t} < \tau_m \vee \tau_n < I^c_{a,t} \wedge \tau_D < \breve{\alpha}_{a,t} < \tau_L \\ \text{DF} : \exists\text{BCM} \wedge I^c_{a,t} < \tau_m \wedge \breve{\alpha}_{a,t} < \tau_D \\ \text{LF} : \exists\text{BCM} \wedge I^c_{a,t} > \tau_n \wedge \breve{\alpha}_{a,t} > \tau_L \\ \text{BgI} : \exists\text{BIM} \wedge I^c_{a,t} < \tau_m \vee \tau_n < I^c_{a,t} \wedge \left|I^I_{a,t} - \mu^I\right| < \tau^I_a \\ \text{FgI} : \exists\text{BIM} \wedge I^c_{a,t} < \tau_m \vee \tau_n < I^c_{a,t} \wedge \left|I^I_{a,t} - \mu^I\right| > \tau^I_a \\ \text{BgTD} : \nexists\text{BM} \wedge - \wedge \sigma_{a,t} < \tau_T \\ \text{FgTD} : \nexists\text{BM} \wedge - \wedge \sigma_{a,t} > \tau_T \end{cases}$$

where the sensor dynamic range is determined by τ_m, τ_n; $\breve{\beta}_{a,t}$ and $\breve{\alpha}_{a,t}$ are the normalised distortions for the current test image; and $\sigma_{a,t}$ is the standard deviation for the current and last two images.

Edge segmentation is achieved based on the following premises:

- Illumination changes modify the gradient magnitude but not the gradient orientation.
- Gradient orientation is not feasible where there are no edges.
- An edge can appear in a place where there were no edges before.

Assuming the first two premises, the oriented gradients will be compared instead of the gradient magnitudes for those pixels which have a minimum magnitude, in order to avoid the false edges due to illumination changes:

$$F_\theta = \left((\tau^c_{e,a} < V^c_{G,a,t}) \wedge (\tau^c_{e,a} < \mu^c_{G,a})\right) \wedge (\tau^c_{\theta,a} < |V^c_{\theta,a,t} - \mu^c_{\theta,a}|), \tag{3.9}$$

For those pixels satisfying the third premise, their gradient magnitudes are compared instead of their orientation magnitudes:

$$F_G = \left(\neg\left((\tau^c_{e,a} < V^c_{G,a,t}) \wedge (\tau^c_{e,a} < \mu^c_{G,a})\right)\right) \wedge (\tau^c_{G,a} < |V^c_{G,a,t} - \mu^c_{G,a}|), \tag{3.10}$$

where the $V^c_{\theta,a,t}$ and $V^c_{G,a,t}$ are the gradient orientation and magnitude for every pixel in the current image, respectively.

3.2.4 Camouflage in Chroma (case DC/LC)

Despite edge segmentation is less sensitive to global illumination changes than colour and intensity cue, problems like noise, foreground aperture and camouflage prevents accurate segmentation of foreground objects. Therefore, handle dark and light camouflage problems by using only edges is not feasible. Then, the brightness component of the colour model should be used to solve the foreground aperture difficulty by filling the foreground object.

Thus, in the second step, the region enclosed by the foreground pixels segmented in the first step are combined with the thresholded shadows and highlights in order to solve the foreground camouflage in chroma and avoid global and local illumination problems, thereby segmenting foreground pixels as dark (DC) and light camouflage (LC):

$$\begin{aligned} \text{DC} &= \mathit{Region}(FgC \vee DF \vee LF \vee FgI \vee FgE) \wedge k_{DI} S, \\ \text{LC} &= \mathit{Region}(FgC \vee DF \vee LF \vee FgI \vee FgE) \wedge k_{LI} H, \end{aligned} \tag{3.11}$$

where $k_{DI} = k_{LI}$ is a factor that specifies the confidence region.

In this second step, shadows (S) and highlights (H) are also modified due to DC/LC. Furthermore, to avoid noise generated from the edge cues, the foreground edges obtained for the BEM are filtering using the region created to cope with DC/LC problem.

An example of image segmentation where camouflage in chroma is solved can be seen in Fig. 3.9, where the agent near the crosswalk has the jeans dark camouflaged with the road. The whole process is summarised in Algorithm 1.

3.2.5 Ghost Detection

Segmented input regions are evaluated to assess whether they contain a ghost or a foreground region based on two premises:

- A ghost corresponds to an object which was represented in the background model. Therefore, the detected region must belong to the background model.
- A ghost cannot be in motion. Therefore, the detected region does not exhibit any motion.

Firstly, the boundary and the area from the detected region are compared with the foreground edges and the region enclosed by these edges. Thus, the foreground segmentation is compared with the foreground edges obtained from the edge cue to know the probability that a detected region belongs to the background model or to the current image. Then, the boundary and the area from the detected region are also compared with the foreground obtained from the temporal difference algorithm, to know the probability that the detected region is in motion. Finally, a region is considered a ghost based on the probabilities obtained in the first step. A region is detected as a ghost if:

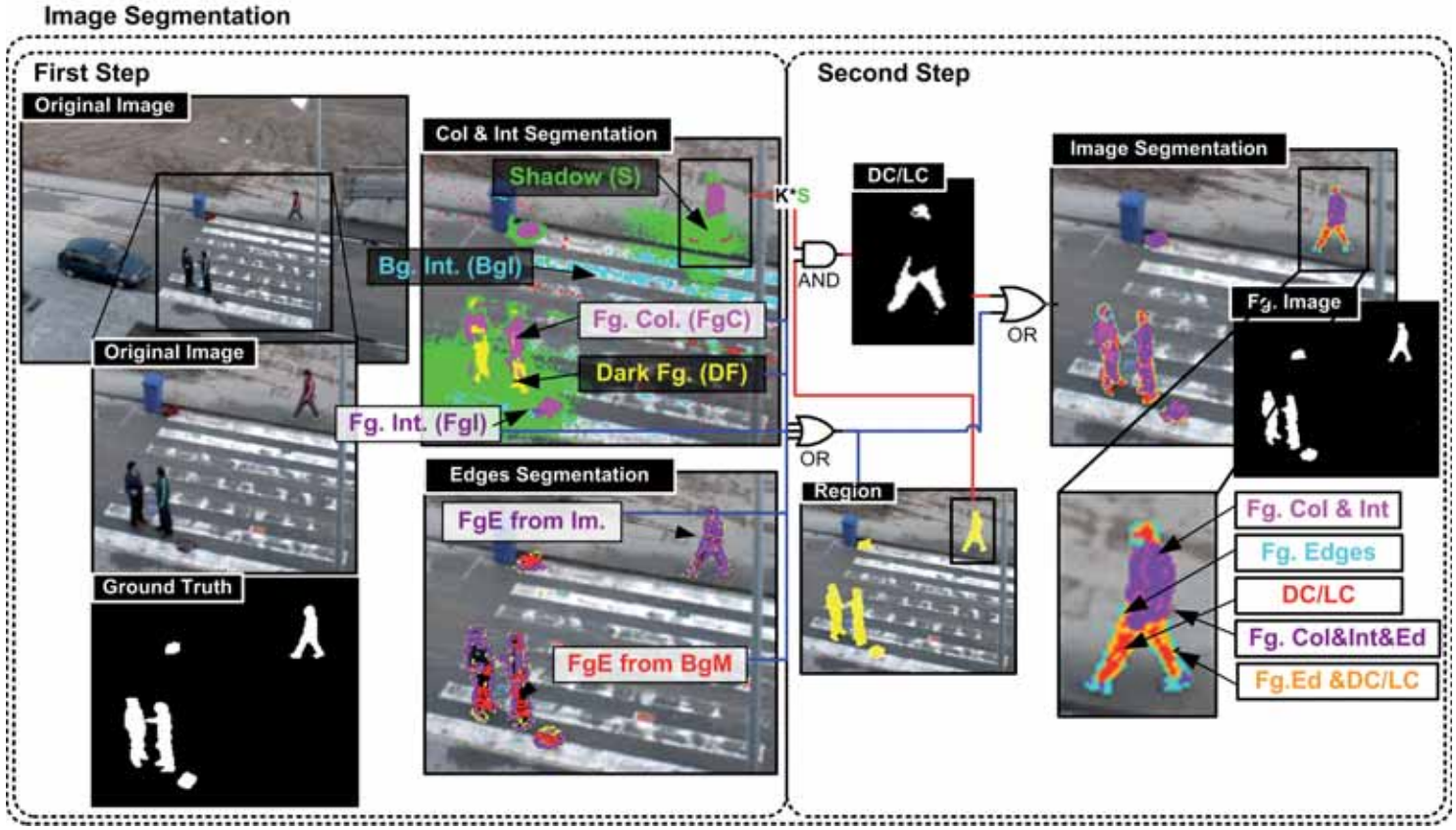


Figure 3.9: Approach and example of Image segmentation. Colour and intensity segmentation, and edge segmentation figures show fg. pixels and S/H masks results from the first step. Region figure shows the region enclosed by the fg. pixels from the first step. DC/LC figure shows how dark and light camouflage pixels are correctly segmented using a thresholded S/H combined with the foreground region. Finally, the image segmentation figure shows the final result of foreground segmentation.

Algorithm 1 Image Segmentation.

- **if** BCM and BEM exists for the current pixel ('a'), **then**:
  - **if** the current pixel ($I^c_{a,t}$) is within the sensor dynamic range ($\tau_m < I^c_{a,t} < \tau_n$), **then**:
    - **if** it has a different chrominance ($\breve{\beta}_{a,t} > \tau_\beta$) or different gradient ($F_\theta \vee F_G$), **then** foreground (FgC, FgE),
    - **else if** it has lower brightness ($\breve{\alpha}_{a,t} < \tau_{\alpha 1}$) and is outside the enclosed foreground region ($\notin Region(Fg)$), **then** shadow (S),
    - **else if** it has lower brightness and is inside the enclosed foreground region ($\in Region(Fg)$), **then** dark camouflage (DC),
    - **else if** it has higher brightness ($\breve{\alpha}_{a,t} > \tau_{\alpha 2}$) and is outside the enclosed foreground region ($\notin Region(Fg)$), **then** highlight (H),
    - **else if** it has higher brightness and is inside the enclosed foreground region ($\in Region(Fg)$), **then** light camouflage (LC),
    - **otherwise**, original background (BgC, BgE).
  - **else**
    - **if** it has lower brightness ($\breve{\alpha}_{a,t} < \tau_D$), **then** dark foreground (DF),
    - **else if** it has higher brightness ($\breve{\alpha}_{a,t} > \tau_L$), **then** light foreground (LF),
    - **otherwise**, original background (BB).
- **else if** BIM exists, **then**:
  - **if** it has lower or higher intensity ($\left|I^I_{a,t} - \mu^I\right| > \tau^I_a$), **then** foreground (FgI),
  - **otherwise**, original background (BgI).
- **otherwise**, no background was visible during the training period and temporal-differencing algorithm is applied
  - **if** it has different intensity over three frames ($\sigma_{a,t} > \tau_T$), **then** foreground (FgTD),
  - **otherwise**, original background (BgTD).

$$GD_b = \prod_{s\in\{B,A\}} \left(\frac{\sum\limits_{a\in R_b} (R^s_b \& FgEi^s)}{|R^s_b|} + k_d < \frac{\sum\limits_{a\in R_b} (R^s_b \& FgEb^s)}{|R^s_b|} \right) \\ \& \prod_{s\in\{B,A\}} \left(\frac{\sum\limits_{a\in R_b} (R^s_b \& FgTD^s)}{|R^s_b|} < \Gamma_m \right) \tag{3.12}$$

where the first equation line denotes the first premise and the second equation line the second one; a is the pixel position; R_b is the evaluated region and b is the number of region; $s \in \{B, A\}$ represents whether the boundary (B) or the area (A) is evaluated; $FgEi$ represents the foreground edges of the current image, and $FgEb$ of the background model; $|R_b^s|$ denotes the number of pixels of region b; K_m denotes a confidence region between the probability of belonging to the current image or to the background model; $FgTD$ represents the foregrounds obtained from the temporal difference; and τ_m is a threshold to detect if a region is in motion.

A sketch of the ghost detection approach can be seen in Fig. 3.10, where the images show how the ghost detection works in a real sequence, images are from the Hermes_Outdoor_Cam1 sequence.

3.3 Experimental Results and Comparison Evaluation

Our approach has been tested with several indoor and outdoor sequences under uncontrolled environments, where multiple segmentation problems like the ones mentioned in the introduction appear. These sequences are taken from both well-known public databases, and own ones. Successful segmentation results have been achieved for all of these sequences.

In order to evaluate the performance of the proposed approach in a quantitative way, ground-truth segmentation masks have been generated by manual segmentation. The sequences segmented are *Hermes_Outdoor_Cam1* from the HERMES database[1] (1612 frames @15 fps, 1392 x 1040 PX), *CVC_Zebra1* sequence from CVC database[2] (1343 frames @20 fps, 720 x 576 PX), *CVC_Machine* sequence from CVC database (797 frames @29 fps, 640 x 480 PX), *OneLeaveShopReenter1cor* from the CAVIAR database[3] (389 frames @ 25 fps, 384 x 288 PX) used in PETS 2004, and Hall_Monitor from the NEMESIS database[4] (300 frames, 352x240 PX). Furthermore, approaches from other authors [19, 65, 22, 33, 78, 70, 24] have been used for performance comparison.

Two well-known and most employed quantitative expressions have been utilized to evaluate the segmentation performance, Detection Rate (DR) (also called True Positive Rate) and False Alarm Rate (FAR) [51, 32]:

$$DR = \frac{TP}{TP + FN};\ FAR = \frac{FP}{TP + FP}, \tag{3.13}$$

where DR is the ratio between the number of correctly detected pixels to the total number of pixels in the ground truth data, and FAR is the ratio between the number of misclassified pixels to the total number of detected pixels. TP, FP and FN correspond to the true positive, false positive, and false negative pixels comparing the segmentation results with the ground truth data.

[1] http://www.hermes-project.eu
[2] http://iselab.cvc.uab.es
[3] http://homepages.inf.ed.ac.uk/rbf/CAVIAR
[4] http://www.ics.forth.gr/cvrl/demos/NEMESIS/hall_monitor.mpg

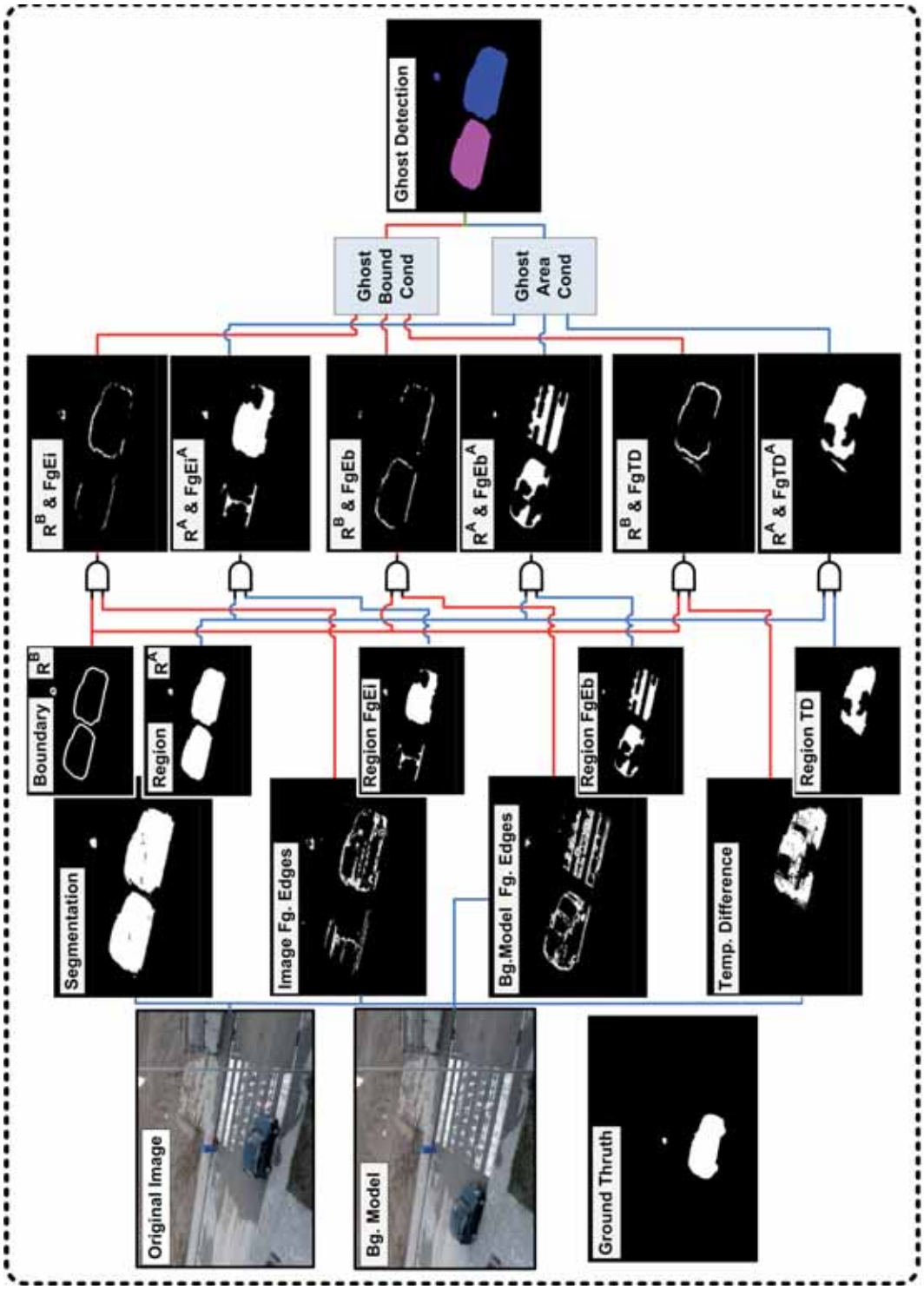


Figure 3.10: Approach and example of Ghost Detection. Boundary and area from the detected region are compared with the foreground edges and the region enclosed by these edges. In this way, the probability that a detected region belongs to the background model or to the current image is obtained. Then, the boundary and the area from the region detected are also compared with the foregrounds obtained from the temporal difference algorithm. In this way, the probability that detected region is in motion is obtained. Then, a region is considered a ghost based on the probabilities obtained in the first step. Ghost Detection image shows the correct detection of the ghost in magenta colour.

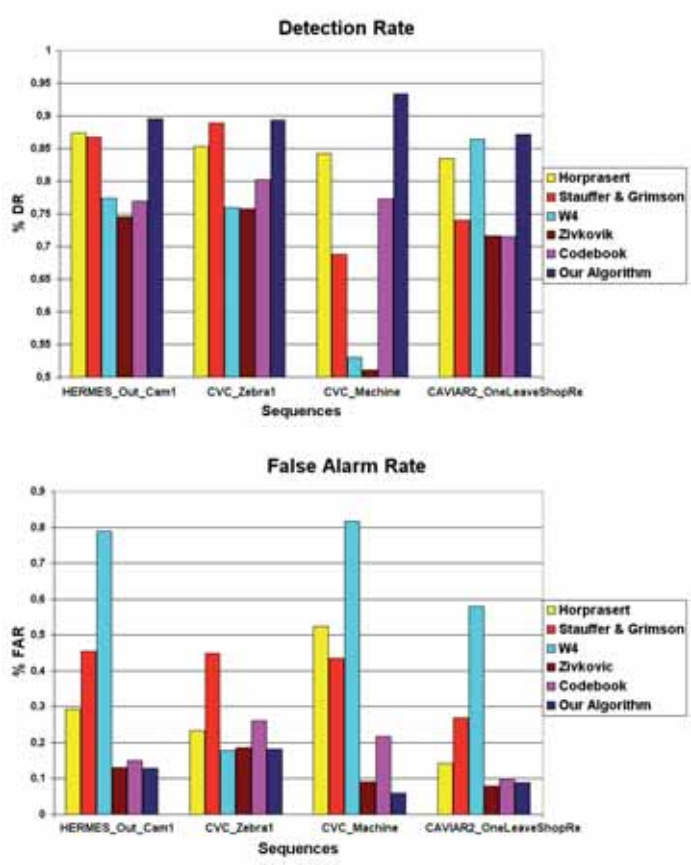


Figure 3.11: Detection Rate and False Alarm Rate results. First sequence is from HERMES Database, second and third sequences are from CVC Database, and fourth sequence is from CAVIAR Database. Our approach has been compared with different approaches [19, 65, 22, 78, 33] using a ground-truth manually segmented. Our algorithm obtains the best detection rate, maintaining the lowest false alarm rate in all the sequences evaluated.

Fig. 3.11 shows the results of the segmentation process using the DR and the FAR. Results show that our algorithm obtains the best DR with the lowest FAR in all the evaluated sequences. The Figs. 3.12, 3.13, 3.14, 3.15, 3.16, 3.17 and 3.18 show why our approach obtains the best results.

Fig. 3.12 shows the results to compare our approach with other approaches [19, 65, 22, 33, 78] on the *Hermes_Outdoor_Cam1* sequence. The first graph of Fig. 3.12 shows the number of false negative pixels segmented using the different approaches, and the second one shows the number of false positive pixels. Frames from 790 up to 1040 correspond to a gradual illumination change. Also, two cars appear into the scene and several persons are crossing the road through a crosswalk. Therefore, multiple motion segmentation difficulties appears in this sequence: (i) global illumination changes —

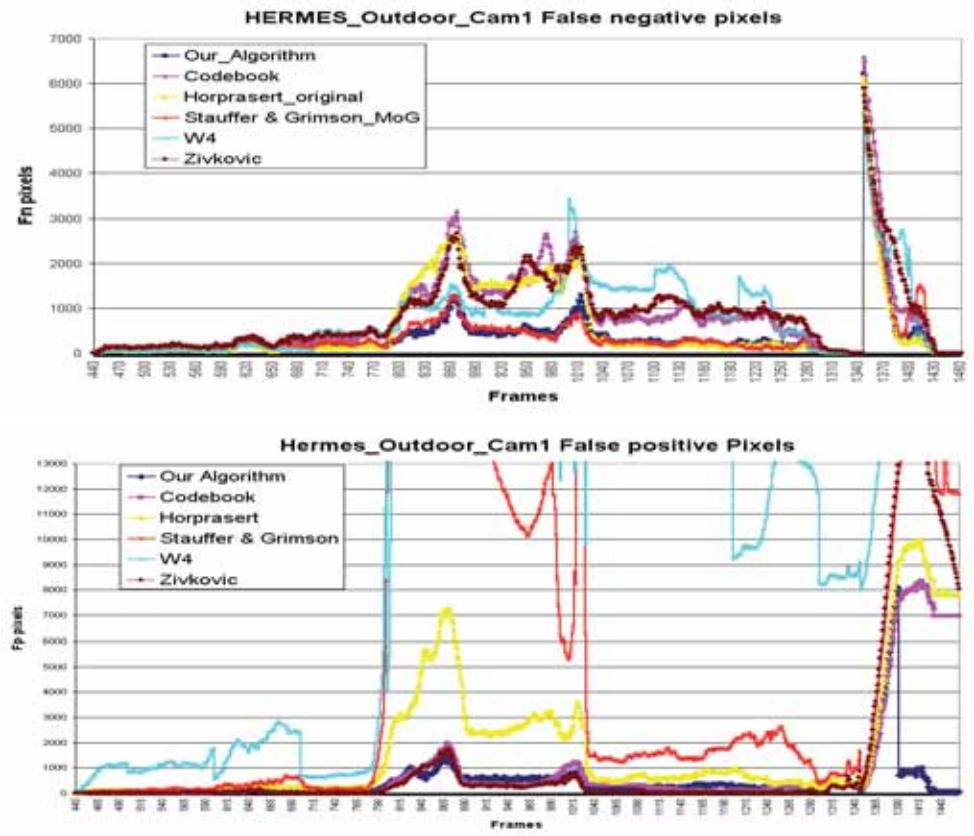


Figure 3.12: False negatives (first graph) and false positives (second graph) computed from comparing W^4 [19] (colour cyan), Stauffer and Grimson approach [65] (colour red), Horprasert et al. approach [22] (colour yellow), Codebook [33] (colour magenta), Zivkovic et al. approach [78] using a shadow detector [51] (colour brown) and our approach (colour blue), based on the ground truth from the HERMES database. Our approach obtains the best results. See text for details.

the scene get darker for an instant—, (ii) local illumination changes —shadows from agents and vehicles—, (iii) camouflage —trousers of agent three when he appears in the scene—, (iv) dark and light camouflage problems —dark camouflage of the trousers of the agent three when he is crossing the crosswalk and light camouflage of the white car with the grey road—, and (v) ghost problem —a car parked begins to move—.

In the aforementioned sequence, W^4 (cyan line) segments the illumination change as foreground, and also any shadows of cars and agents. The Stauffer and Grimson algorithm (red line) cannot always cope with the illumination change and also classifies the shadows as foreground. The Horprasert et al. approach (yellow line) cannot tackle the light camouflage (white car with grey road). Codebook (magenta line) is not able

to differentiate between the illumination change and camouflage in chroma, and the Zivkovic et al. approach (brown line) segments the illumination changes and all the shadows like the Stauffer and Grimson approach. However, the version analysed in this thesis included a shadow detection [51] which has also problems to distinguish between illumination changes and camouflage in chroma. As it can be seen our approach is robust to these problems and obtains the best segmentation among the approaches compared. A significant frame (number 864), shown in Fig. 3.13, has been selected for qualitative comparison.

Frames from 1340 up to 1460 (last 100 frames showed) in Fig. 3.12 correspond to a car parked which begins to move, therefore the problem of ghost appears. Our approach is the only one among the evaluated approaches that can cope with this problem as soon as it occurs as can be seen in the false positive graph. A significant frame (number 1411) can be seen in Fig. 3.14.

Figs. 3.13, 3.14, 3.15, 3.16, and 3.17 shows significant frames comparing our approach with the other approaches [19, 65, 22, 33, 78]. First row shows the original image, second row is the ground truth, from third up to eighth rows are the segmentation results: third row from W^4 approach [19], fourth row from Stauffer and Grimson approach [65], fifth row from Horprasert et al. approach [22], sixth row from Codebook approach [33], seventh row from Zivkovic et al. approach [78] using a shadow detector [51], and eighth row from own approach. In this figure can be seen why our approach performs better than other approaches. Our approach obtains more number of TP along with less number of FP and FN pixels, showing that our algorithm can tackle global and local illumination problems, problems beyond the dynamic range, chroma and intensity camouflage problem, bootstrapping problems, and ghost problem.

In the *CVC_Zebra1* sequence, four people are involved during the scene. Further, several vehicles cross the scene in a front plane, and people walk beside various street lamps and trees. W^4 segments the shadows as foreground and have problems with the updating process. Stauffer and Grimson approach has problems with shadows and gradual illumination changes. Horprasert et al. cannot solve the light camouflage problem (white shirt with the grey road) and cannot cope with saturation problem of the sky with gradual illumination changes. Codebook cannot also cope with the light camouflage problem and the saturation problem simultaneously. Zivkovic et al. approach has also problems with the illumination changes and camouflages, furthermore the updating system has the sleeping person problem [66] (fg. pixels are segmented as bg. because the updating system incorporate the fg. motionless objects to the bg. model). Our algorithm is robust to these problems. Fig. 3.15 presents one frame where the light camouflage problem described above is observed (white shirt with the grey road).

In the *CVC_Machine* sequence an agent enters the scene and interacts with a vending machine, see Fig. 3.16. This scene presents strong illumination changes, and big saturation with the wall. Dark and light camouflages are also present in the scene (agent in front of wall). Our algorithm can satisfactorily manage strong illumination changes, saturation problems, and dark and light camouflage avoiding sleeping person anomaly. Zivkovic et al. can manage the strong illumination change using the updating system, but it increases considerably the sleeping person anomaly.

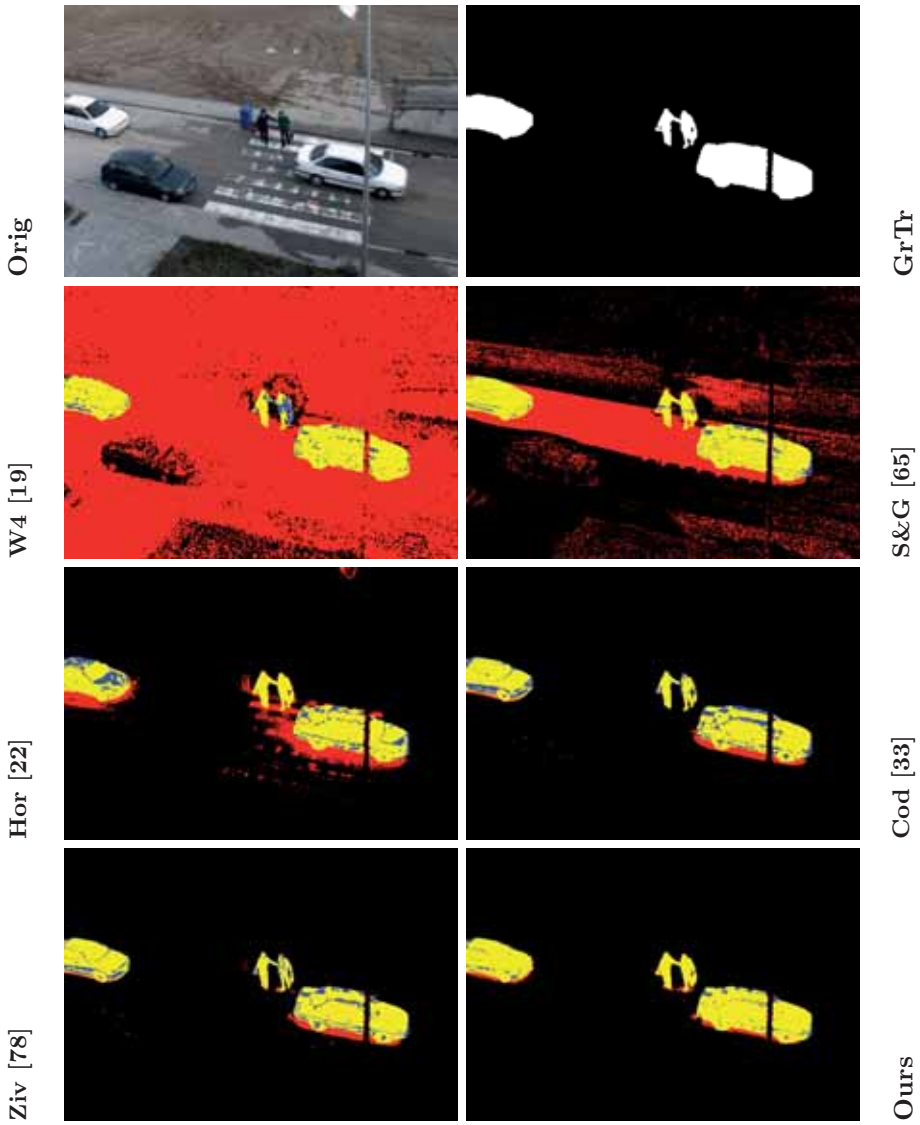


Figure 3.13: Foreground segmentation comparative using HERMES database. First image and second image are an Original image and the ground truth from HERMES Outdoor sequence, light camouflage problem. From third image up to eight image are the segmentation results using the W^4 approach [19], the Stauffer and Grimson approach [65], Horprasert et al. approach [22], the Codebook approach [33], the Zivkovic et al. approach [78] using a shadow detector [51], and our approach, respectively. Segmentation results are coloured in yellow for TP pixels, blue for FN pixels, and red for FP pixels. Our algorithm obtains more number of TP along with less number of FP and FN. Light camouflage problem, white car with grey road, and soft illumination change.

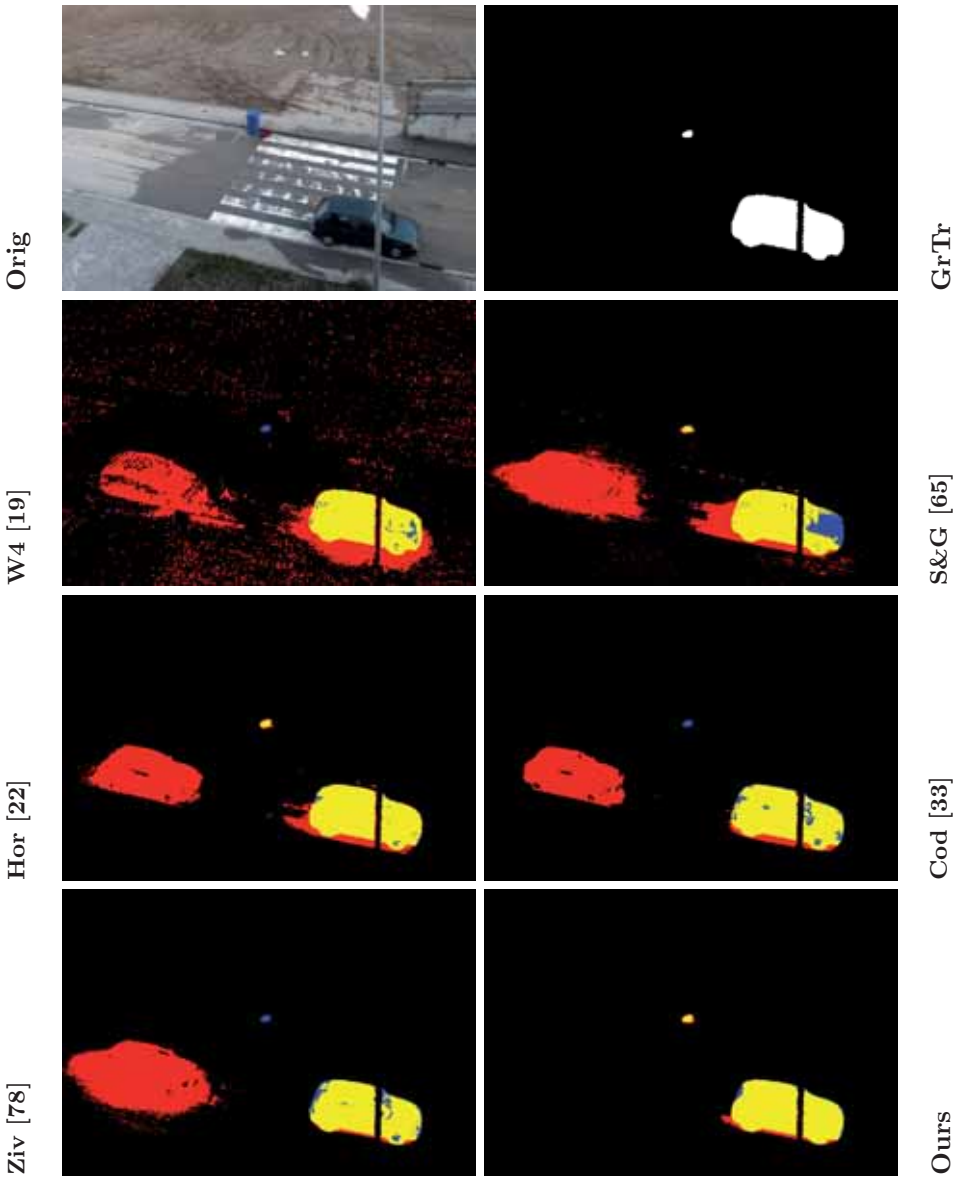


Figure 3.14: Foreground segmentation comparative using HERMES database. First image and second image are an Original image and the ground truth from HERMES Outdoor sequence, ghost problem. From third image up to eight image are the segmentation results using the W^4 approach [19], the Stauffer and Grimson approach [65], Horprasert et al. approach [22], the Codebook approach [33], the Zivkovic et al. approach [78] using a shadow detector [51], and our approach, respectively. Segmentation results are coloured in yellow for TP pixels, blue for FN pixels, and red for FP pixels. Our algorithm obtains more number of TP along with less number of FP and FN. Ghost problem.

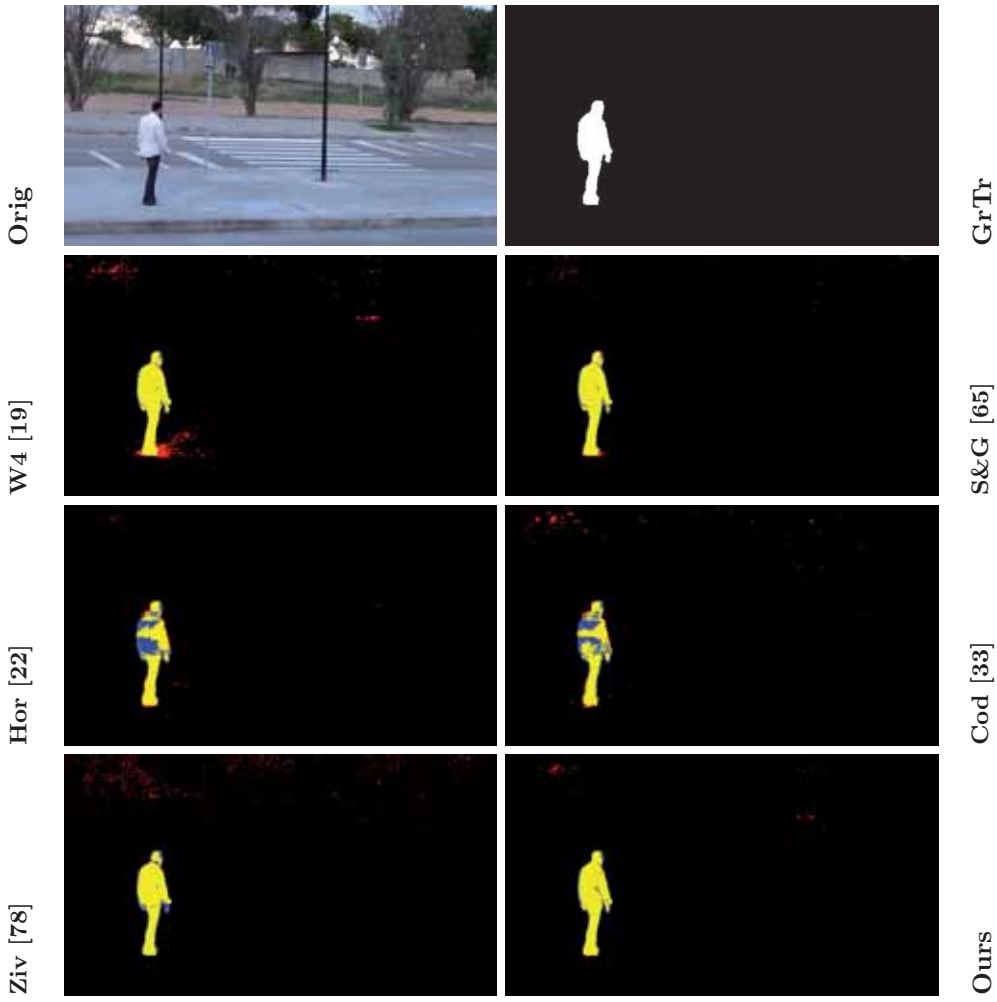


Figure 3.15: Foreground segmentation comparative using CVC database. First image and second image are an Original image and the ground truth from CVC Zebra1 sequence. From third image up to eight image are the segmentation results using the W^4 approach [19], the Stauffer and Grimson approach [65], Horprasert et al. approach [22], the Codebook approach [33], the Zivkovic et al. approach [78] using a shadow detector [51], and our approach, respectively. Segmentation results are coloured in yellow for TP pixels, blue for FN pixels, and red for FP pixels. Our algorithm obtains more number of TP along with less number of FP and FN. Light Camouflage due to white shirt with grey road, and sky saturation.

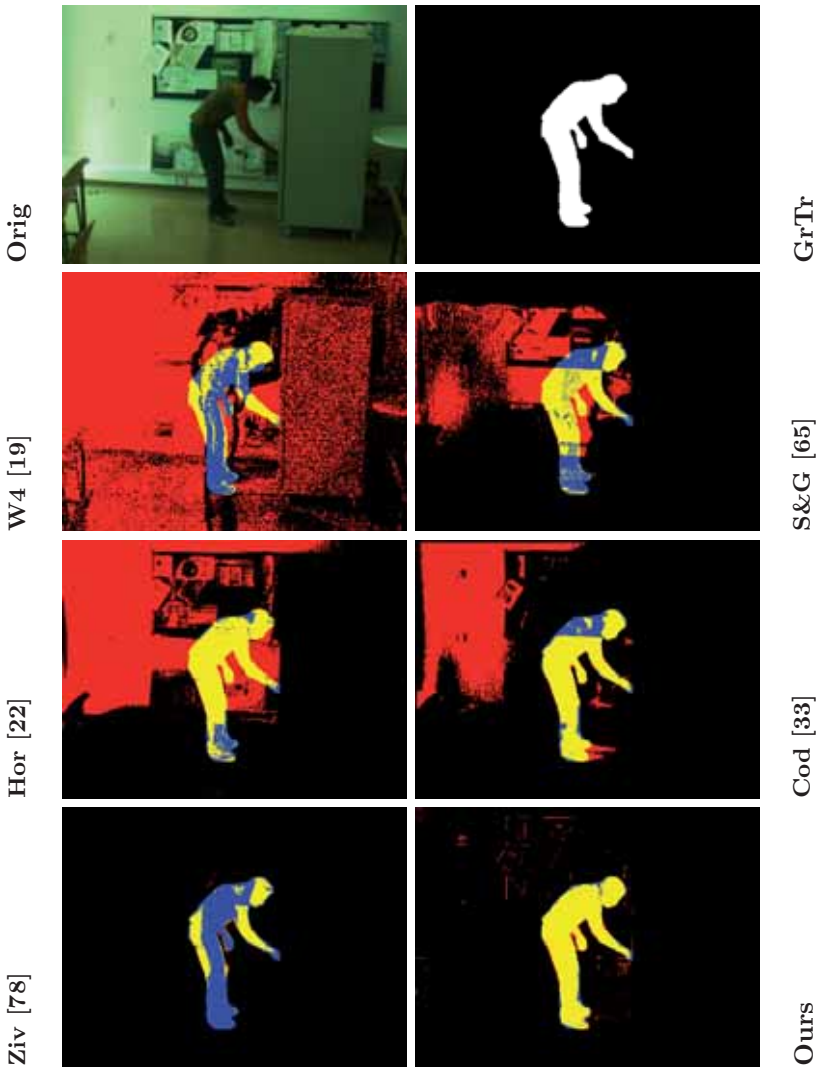


Figure 3.16: Foreground segmentation comparative using CVC database. First image and second image are an Original image and the ground truth from CVC Machine sequence. From third image up to eight image are the segmentation results using the W^4 approach [19], the Stauffer and Grimson approach [65], Horprasert et al. approach [22], the Codebook approach [33], the Zivkovic et al. approach [78] using a shadow detector [51], and our approach, respectively. Segmentation results are coloured in yellow for TP pixels, blue for FN pixels, and red for FP pixels. Our algorithm obtains more number of TP along with less number of FP and FN. Strong illumination and saturation problem.

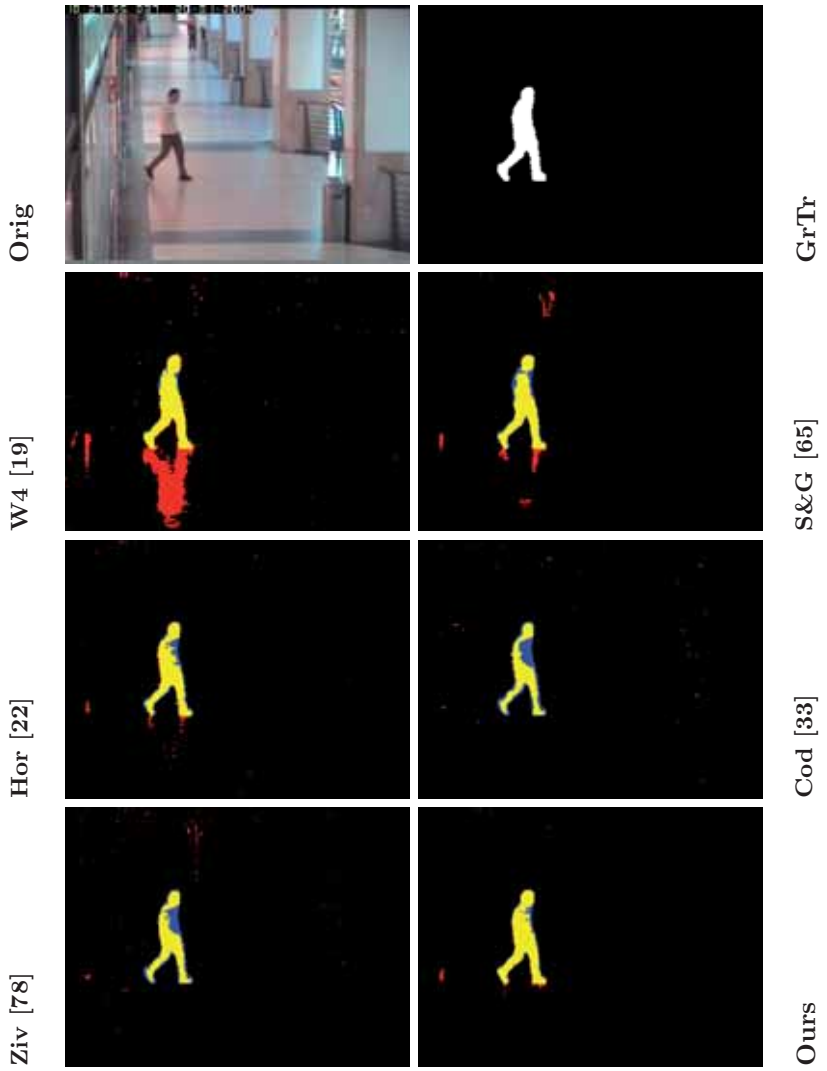


Figure 3.17: Foreground segmentation comparative using CAVIAR database. First image and second image are an Original image and the ground truth from OneLeaveShopReenter1cor sequence. From third image up to eight image are the segmentation results using the W^4 approach [19], the Stauffer and Grimson approach [65], Horprasert et al. approach [22], the Codebook approach [33], the Zivkovic et al. approach [78] using a shadow detector [51], and our approach, respectively. Segmentation results are coloured in yellow for TP pixels, blue for FN pixels, and red for FP pixels. Our algorithm obtains more number of TP along with less number of FP and FN. Strong clutter and different illuminants in the same scene.

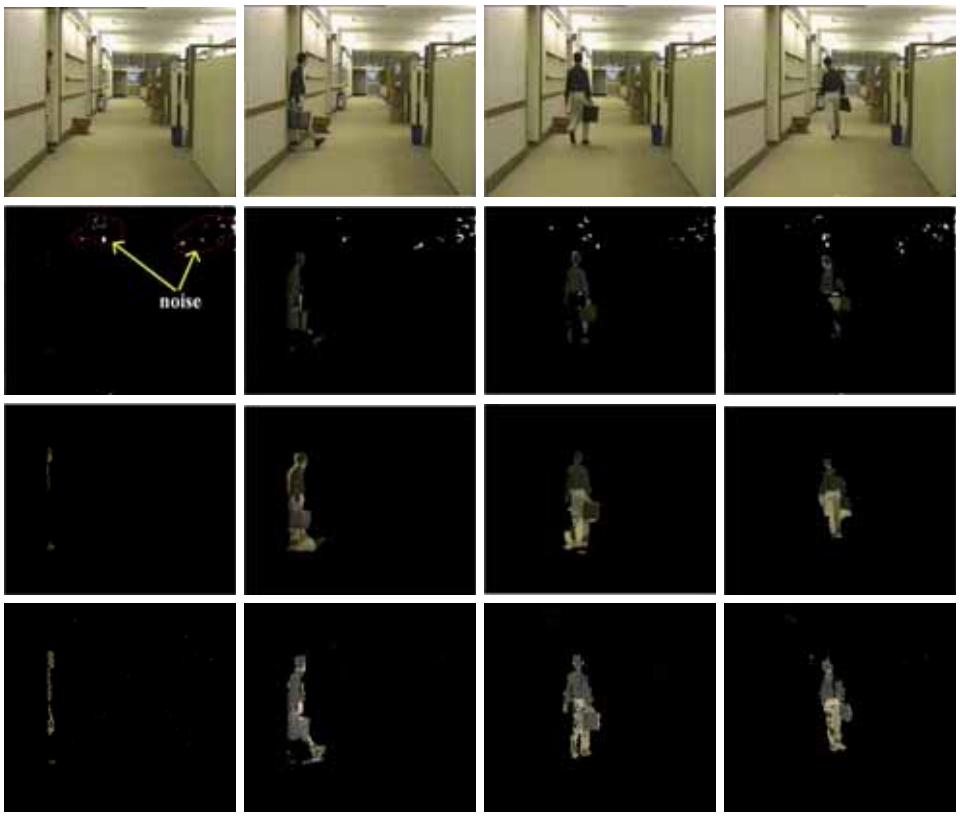


Figure 3.18: First row shows the original frames from the Hall_Monitor sequence of the NEMESIS dataset, second row shows the detection results of Wang et al. [70], and third row shows the detection results from Huang et al. [24]. These images have been obtained directly from [24]. Fourth row shows the segmentation results of our proposed approach without applying any morphological operation.

In the sequence *OneLeaveShopReenter1cor*, two agents are correctly segmented, see Fig. 3.17. The colour distribution of the background is very similar to the agents thus including strong clutter. Furthermore, several oriented lighting sources with different illuminant are present, dramatically affecting the agent appearance depending on its position and orientation (bluish effect at right of the corridor, and reddish one at left). A significant frame of this sequence can be seen in the Fig. 3.17, where dark camouflage and shadows are correctly solved using our approach.

Fig. 3.18 shows frames from Hall_Monitor sequence (first row) comparing Wang et al. approach [70] (second row), and Huang et al. approach [24] (third row) with our approach (fourth row). The sequence shows challenging aspects due to noise, shadows, and camouflage. Wang et al. can not manage correctly noise and camouflage. Huang et al. is able to manage problems with noise, but shadows are not correctly removed, and their approach segments regions corresponding to background as foreground, such

as the region around the legs. Instead, our approach can cope with these issues thus enhancing segmentation.

By using a combination of cues, each of them is used in a very restrictive way without compromising the detection rate. Nevertheless, false positive rate is cut down. Other databases were tested but the available ground truth is not valid for pixel segmentation comparison, since ground truth describes the position of the bounding box where the foreground object was detected, instead of its manually segmented contour.

In order to show the accuracy of our approach it has been tested in a high number of databases, whose most of them are well-know databases from internet. The Figs. 3.24,3.25,3.19,3.35,3.36,3.26,3.33,3.27,3.28,3.29, 3.30, 3.31, 3.32, 3.34, 3.21 show selected frames with the results of our approach in well-knwon datasets. The figures depicted that our approach is able to work in all kind of scenes under uncontrolled environments. Independently on the scene type (indoor, outdoor), the camera resolution (high, low) or localization, the surface geometry or textures, the quality of the images (blurred images), the size, the shape, the type or the appearance of the objects or the background.

Datasets employed: PETS 2001[5], ATON[6], VS-PETS[7], CAVIAR[8], NEMESIS[9], HERMES[10], ATON[11], PETS 2006[12], CVC[13], ETHZ[14], UMIACS[15], MODLAB [16], VSSN06[17], VISOR[18], LSVN[19].

Fig. 3.19 shows some significant detection results from the above analysed HERMES_Outdoor_Cam1 sequence. Several agents and cars are correctly detected despite dark and light camouflage, ghost problems, and a soft illumination change.

A crosswalk sequence is analysed in *Zebra1* (CVC database, 1344 frames @ 25fps, 720x576 pixels). Four people are involved during the scene. Further, several vehicles cross the scene in a front plane, and people walk besides various streetlamps and trees, resulting in multiple, and partial camouflage of the agents.

Significant processed frames are shown in Fig. 3.20 depicting the detection results using our final approach. In this figure, it can be seen as the four agents are correctly detected, despite local illumination and camouflage problems. Thus, part of the agents are sometimes not detected due to the camouflage problem. Furthermore, some pixels belonging to the trees are detected because the background in motion problem is not being tackled in this work.

[5]ftp://ftp.pets.rdg.ac.uk/pub/PETS2001
[6]http://cvrr.ucsd.edu/aton/
[7]ftp://ftp.pets.rdg.ac.uk/pub/VS-PETS/
[8]http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/
[9]http://www.ics.forth.gr/cvrl/demos/NEMESIS/hall_monitor.mpg
[10]http://www.hermes-project.eu
[11]http://cvrr.ucsd.edu/aton/shadow/
[12]http://pets2006.net/
[13]http://iselab.cvc.uab.es/
[14]http://www.vision.ee.ethz.ch/datasets/
[15]http://www.umiacs.umd.edu/users/
[16]http://www.na.icar.cnr.it/ maddalena.l/MODLab/MODseq.html
[17]http://imagelab.ing.unimore.it/vssn06/
[18]http://www.openvisor.org/video_categories.asp
[19]http://vision.gel.ulaval.ca/ CastShadows/

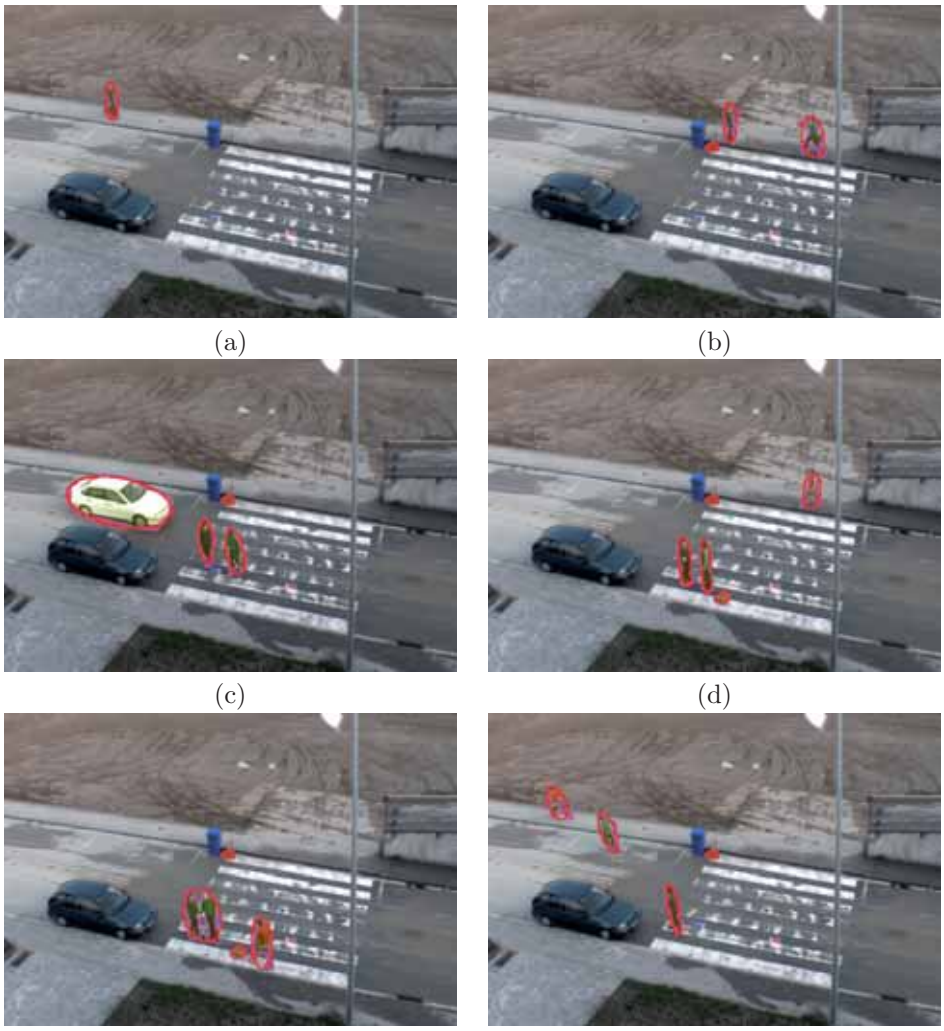
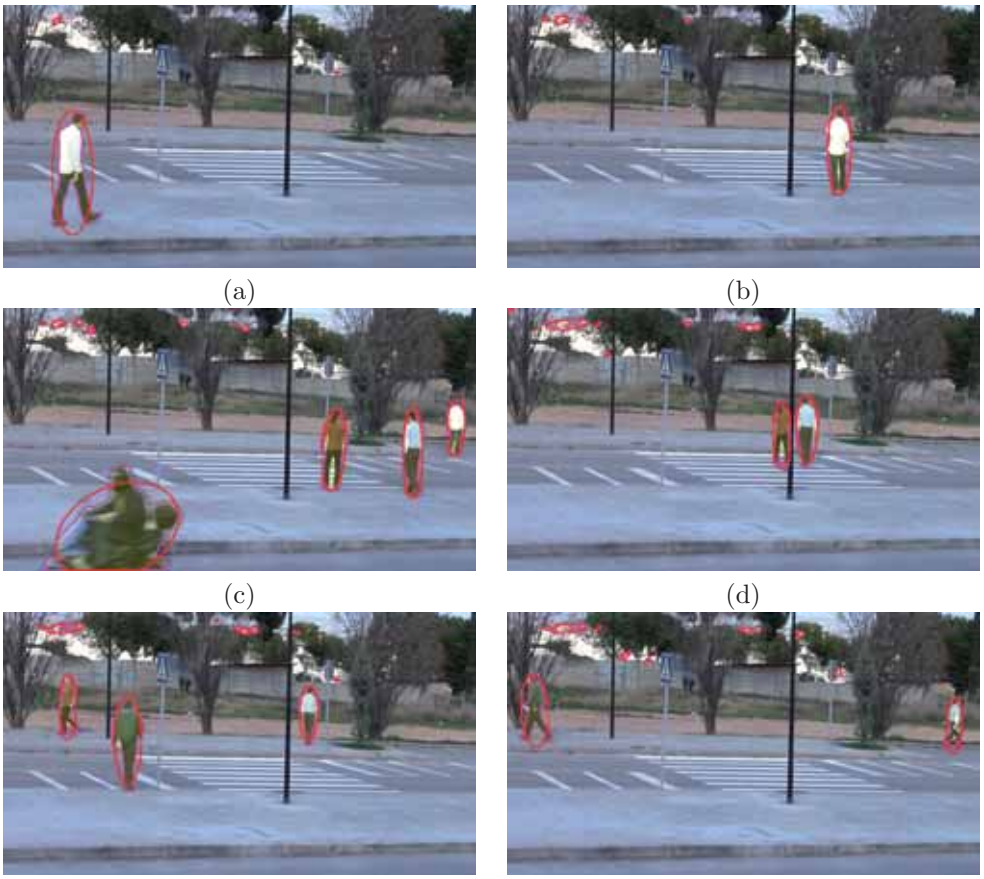


Figure 3.19: Foreground detection results from HERMES_Outdoor_Cam1 sequence using our final approach. Several agents and cars are involved in the scene and they are correctly detected despite dark and light camouflage, ghost problems, and a soft illumination change.

In the Fig. 3.21 some significant frames are shown. The agent in the sequence is accurately detected despite the strong illumination change, the saturation in the left part of the scene, and the reflected shadow over the floor.

In the sequence *OneLeaveShopReenter1cor* (CAVIAR dataset2, 389 frames @ 25 fps, 384 x 288 pixels), two agents are segmented simultaneously, in spite of motion segmentation problems such as camouflage, local illumination problems, among others. The background colour distribution is so similar to agent one that it constitutes



(a) (b) (c) (d) (e) (f)

Figure 3.20: Foreground detection results from Zebra1 sequence using our final approach. The four agents are correctly detected. However, the camouflage problem sometimes appear. Furthermore, some tree pixels are erroneously detected because background in motion (waving tree) is not tackled in our approach.

Figure 3.21: Foreground detection results from CVC_Machine sequence using our final approach. The agent is accurately detected despite the strong illumination change, the saturation in the left part of the scene, and the reflected shadow over the floor.

a strong source of clutter. Furthermore, several oriented lighting sources with different illuminant are present, dramatically affecting the agent appearance depending on its position and orientation (notice the bluish effect on the floor on the right of the corridor, and the reddish one on the floor on the left of the corridor).

Significant processed frames are depicted in Fig. 3.22 showing the detection results using our approach fusing colour, intensity and edge cues. In this figure, it can be seen how the two agents are correctly segmented, thereby handling problems with shadows, dark and light camouflages, etc. Nevertheless, some parts of the agent one are not accurately segmented due to camouflage problem in intensity and chroma, see Fig. 3.22.(a). Furthermore, some shadows are segmented due to the problem with the floor reflectance and the different illuminants in the scene, see Fig. 3.22.(c).

The sequence *DATASET1_TESTING_CAMERA1* (PETS 2001 database, 2688 frames @ 29.97 fps, 768 x 576 pixels) presents a high variety of agents entering and leaving the scene: three isolated people, two groups of people, one isolated person, three cars and a person who exits from a parked car.

Some significant processed frames are depicted in Fig. 3.23, showing the detection results using our final approach in the sequence DATASET1_TESTING_CAMERA1 from the PETS2001 database. In this figure, it can be seen how the agents and cars are correctly detected despite all motion segmentation problems which can be found in this sequence. Nevertheless, some part of the car shadows are erroneously detected because intense local illumination, such as intense shadow, causes edges segmentation. Moreover, some part of the roof and a part of the window is sometimes detected because of BCM problems, when handling the global illumination changes. Furthermore, the window and the parked green car are indefinitely detected because this approach does not have updating process, and it cannot cope with incorporated objects problem. An updating process is required in order to detect the motionless green car and incorporate it to the background.

Fig. 3.24 depicts significant processed frames, showing the detection results using our final approach in the VS_PETS sequence. In this figure, it can be seen how football players and the ball are correctly detected despite the players are always in the scene. Advertisements from the upper part of the frames are also detected because they are changing along the sequence. Furthermore, our algorithm is able to detect all the mobile objects without taking into account the size.

Significant processed frames are depicted in Fig. 3.25, showing the detection results using our final approach in the Intelligentroom indoor sequence from the ATON database. In this figure, it can be seen how the agent is correctly detected despite the low quality of the image, which presents a strong noise, blurred image and shadows.

Fig. 3.26 depicts significant processed frames, showing the detection results using our final approach in CienciesCNM3 sequence from CVC database. In this figure, it can be seen how the agents are accurately in spite of the saturated sky, and the agents are partially occluded by the background.

Fig. 3.27 depicts significant processed frames, showing the detection results using our final approach in the Central_pedX1 sequence from the ETHZ database. In this figure, it can be seen how multiple agents, cars, bikes and motorbikes are detected despite the very low quality of the image, which presents a strong noise, and a blurred image. However, the camouflage problem sometimes appear due to the low

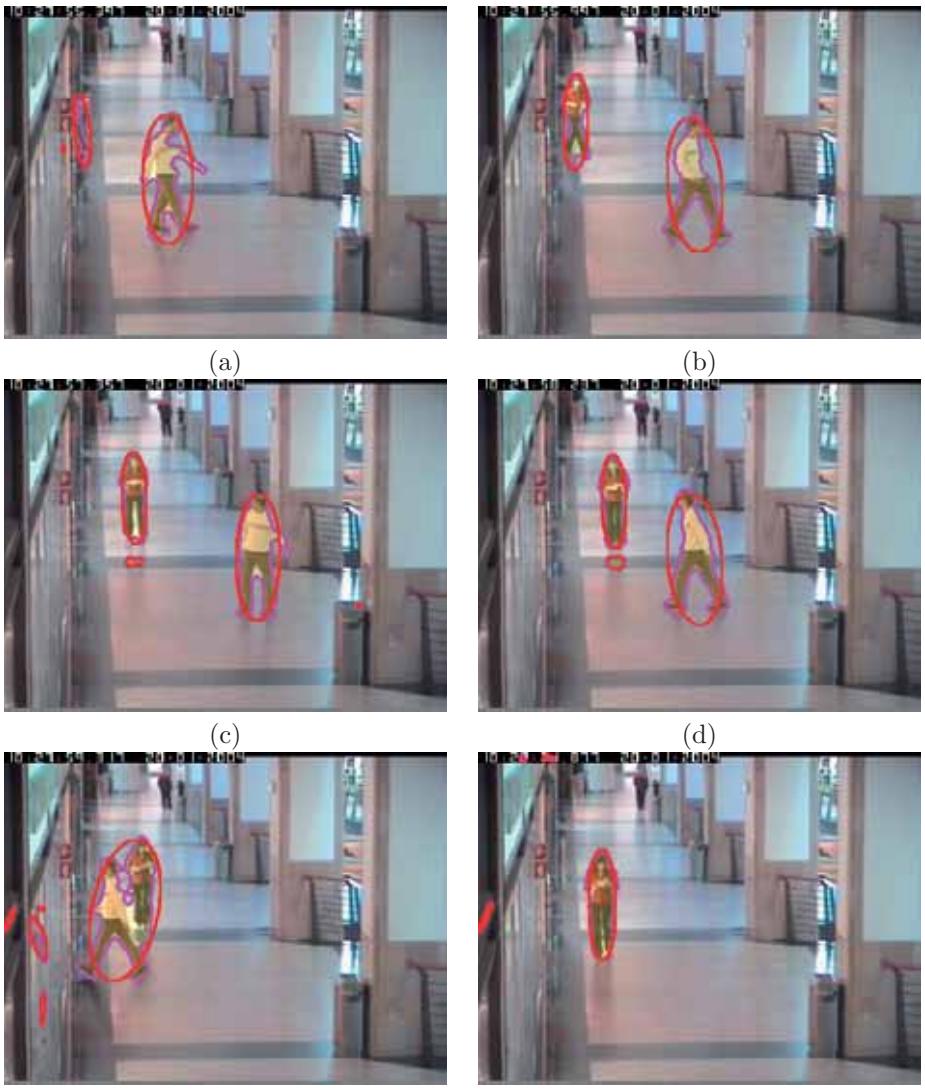


Figure 3.22: Foreground detection results from CAVIAR database using our final approach fusing colour, intensity and edge cues. The two agents are correctly detected despite the sequence exhibits a different illumination due to several different lighting sources. However, some shadows are segmented due to the problem with the floor reflectance and change of the illuminant. Furthermore, some parts of the agent one are not accurately segmented due to camouflage problem in intensity and chroma. Image notation: each red ellipse represents each detected object and magenta lines denote their contour.

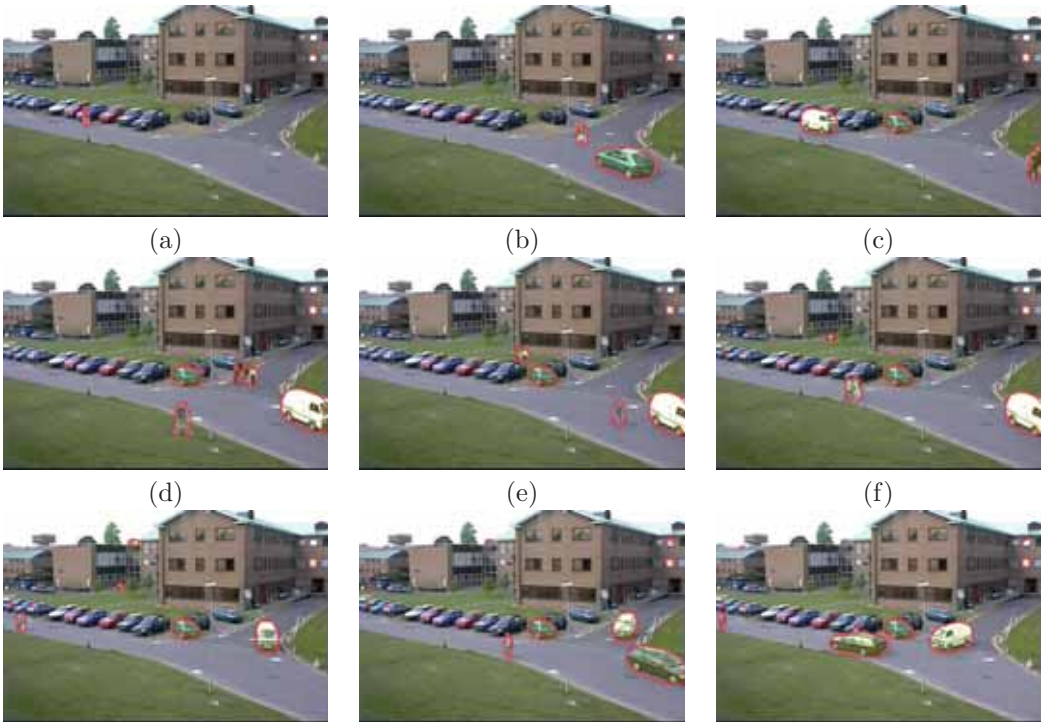


Figure 3.23: Foreground detection results from PETS 2001 database using our final approach. All the agents and vehicles are correctly detected. Nevertheless, some vehicle shadows are detected due to intense local illumination, and some part of the roof and a part of window is sometimes detected due to BCM problems with the global illumination changes. Furthermore, the window and the parked green car are indefinitely detected because this approach does not have updating process.

quality of the image. Furthermore, some tree pixels are erroneously detected because background in motion (waving tree) is not tackled in our approach.

Significant processed frames are shown in Fig. 3.28 depicting the detection results using our final approach in the Laboratory sequence from the ATON database. In this figure, it can be seen how the two agents are correctly detected despite the big problems with the camouflage and the shadows presented. The agents appears in the scene walking and running at different speeds. The detection is achieved thereby showing that our approach is invariable to the frame rate. However, the camouflage problem sometimes appear in some part of the agents. An updating process is required in order to incorporate the filling cabinet opened along the sequence into the background.

Fig. 3.29 depicts significant processed frames, showing the detection results using our final approach in the Rats BlackWhiteboxr sequence. In this figure, it can be seen how the rat is detected despite the problems with the camouflage and shadows

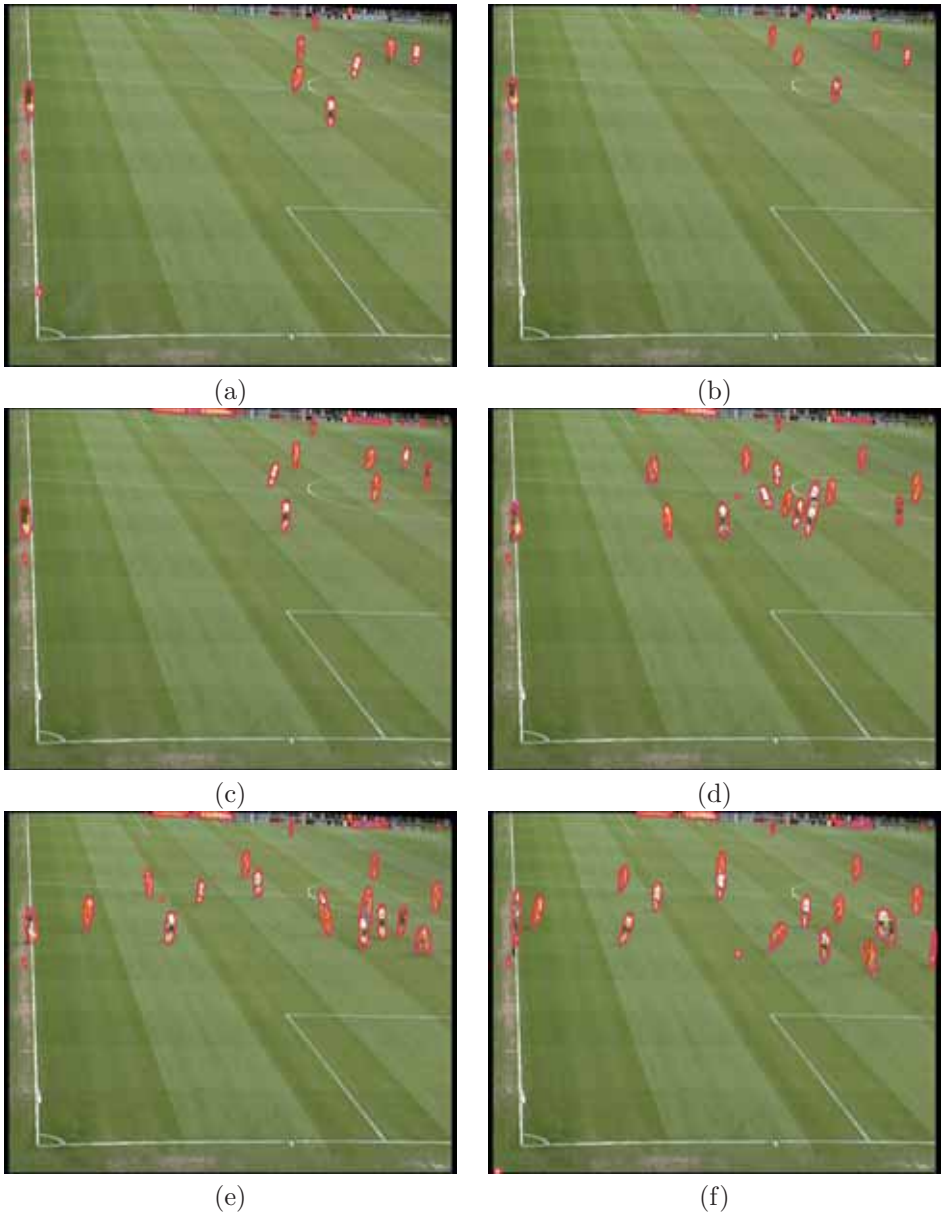


Figure 3.24: Foreground detection results from VS_PETS sequence using our final approach. The players and the fotball are correctly detected despite the players are always in the scene, hence our approach can detect all the foreground objects without taking into account the size. Advertisements from the upper part of the frames are also detected because they are changing along the sequence.

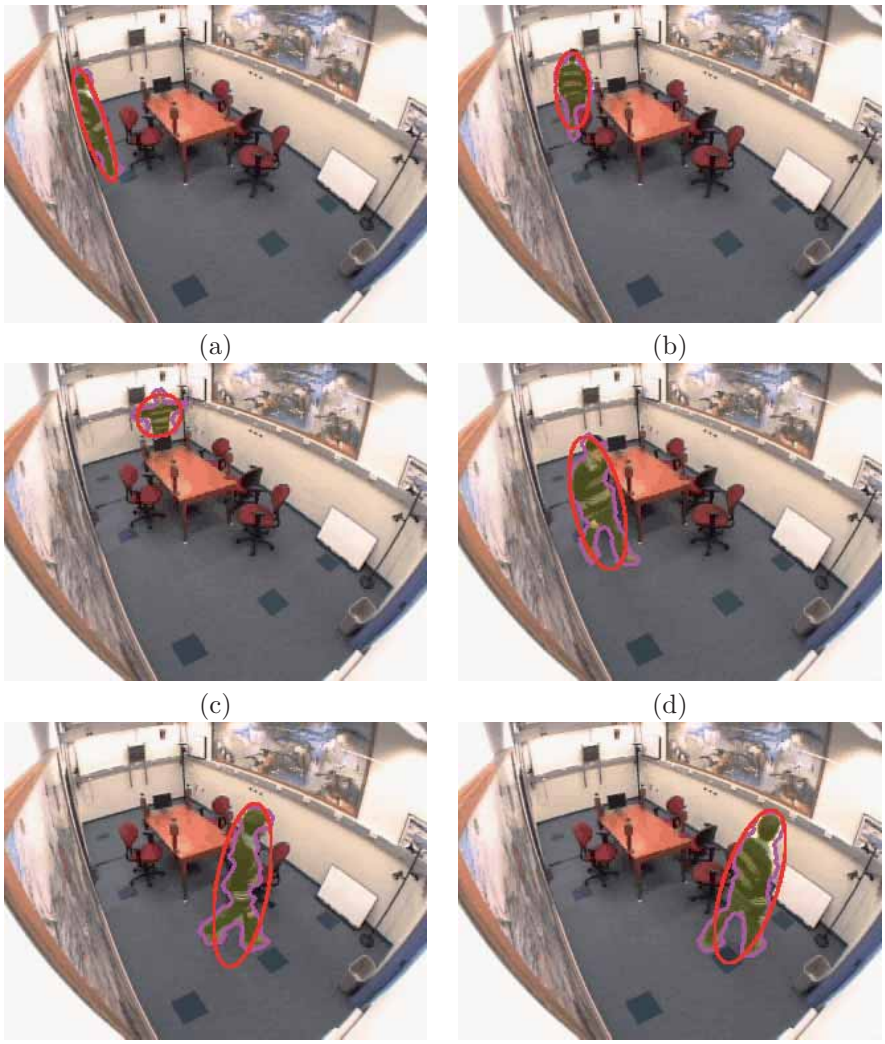


Figure 3.25: Foreground detection results from ATON_Intelligentroom sequence using our final approach. The agent is accuratelly detected despite the blurred and noisy image, and the shadows in the sequence.

presented. Notice that in spite of colour images the sequence does not have chrominance and all the detection achieved is using the dark and light camouflage process. However, the camouflage problem sometimes appear in some part of the rats due to the lack of change in intensity and chroma.

Fig. 3.30 depicts some significant frames, showing the detection results using our final approach in the Campus sequence from ATON database. In this figure, it can be seen how the agents and the cars are detected despite the problems with the

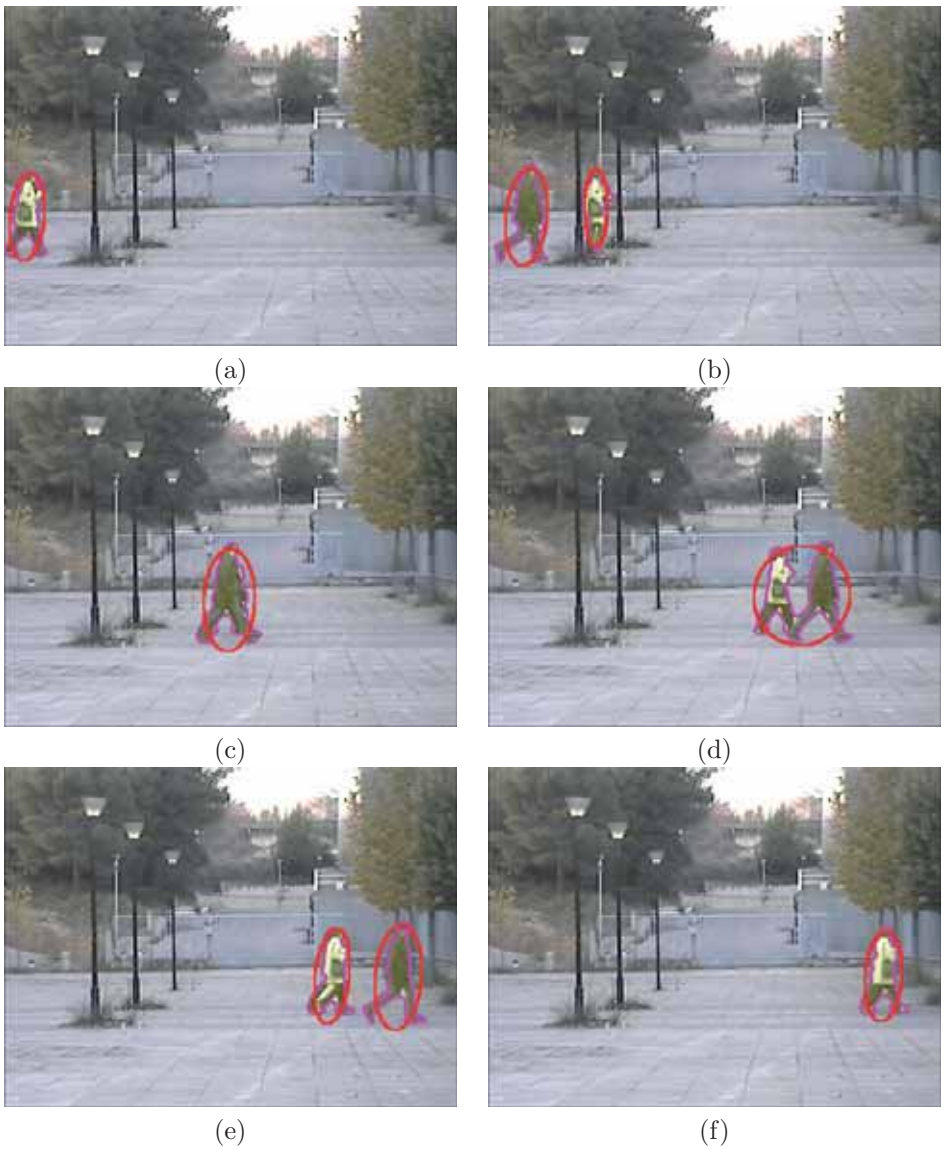
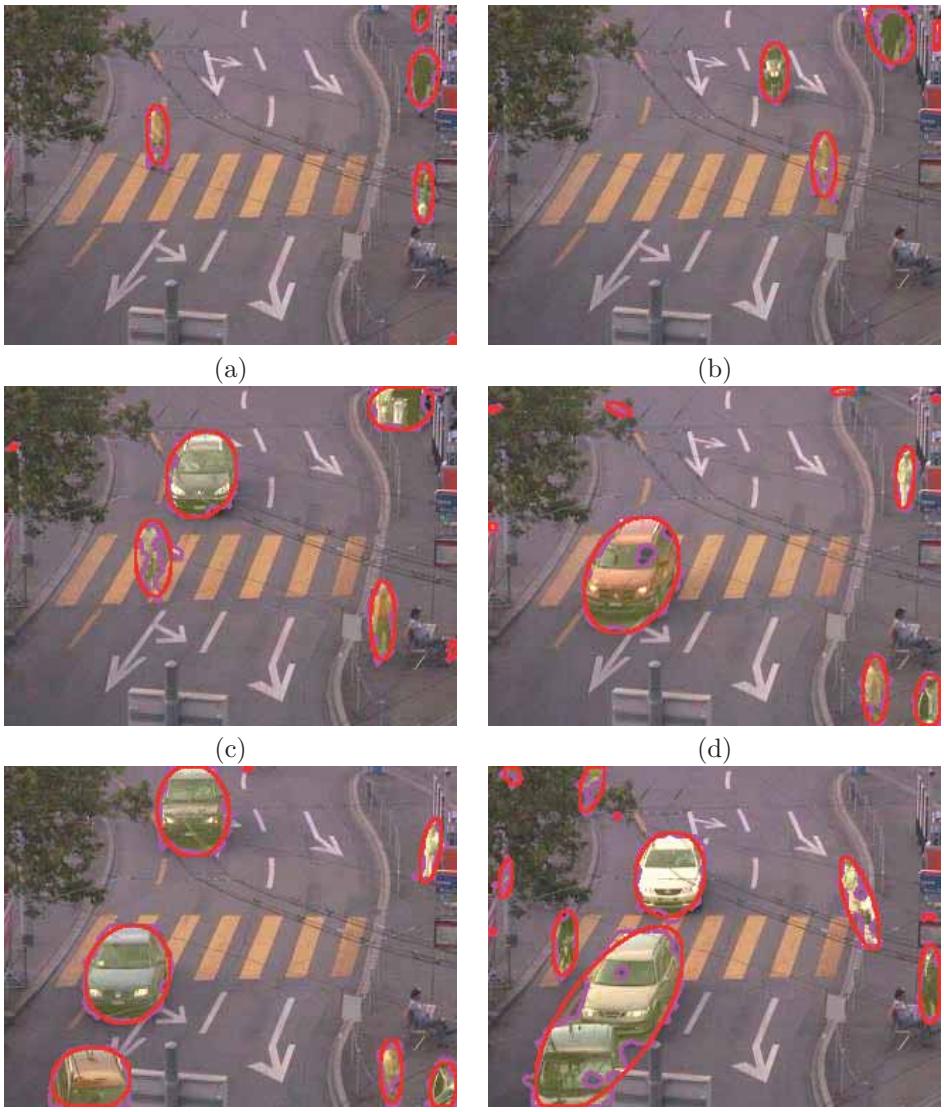


Figure 3.26: Foreground detection results from CVC_CienciesCNM3 sequence using our final approach. Several agents are involved in the scene and they are accurately detected in spite of the saturated sky, and the agents are partially occluded by the background.



(a) (b) (c) (d) (e) (f)

Figure 3.27: Foreground detection results from ETHZ_Central_pedX1 sequence using our final approach. Multiple agents, cars, bikes and motorbikes are detected despite the very low quality of the image, which presents a strong noise, and a blurred image. However, the camouflage problem sometimes appear due to the image low quality. Furthermore, some tree pixels are erroneously detected because background in motion (waving tree) is not tackled in our approach.

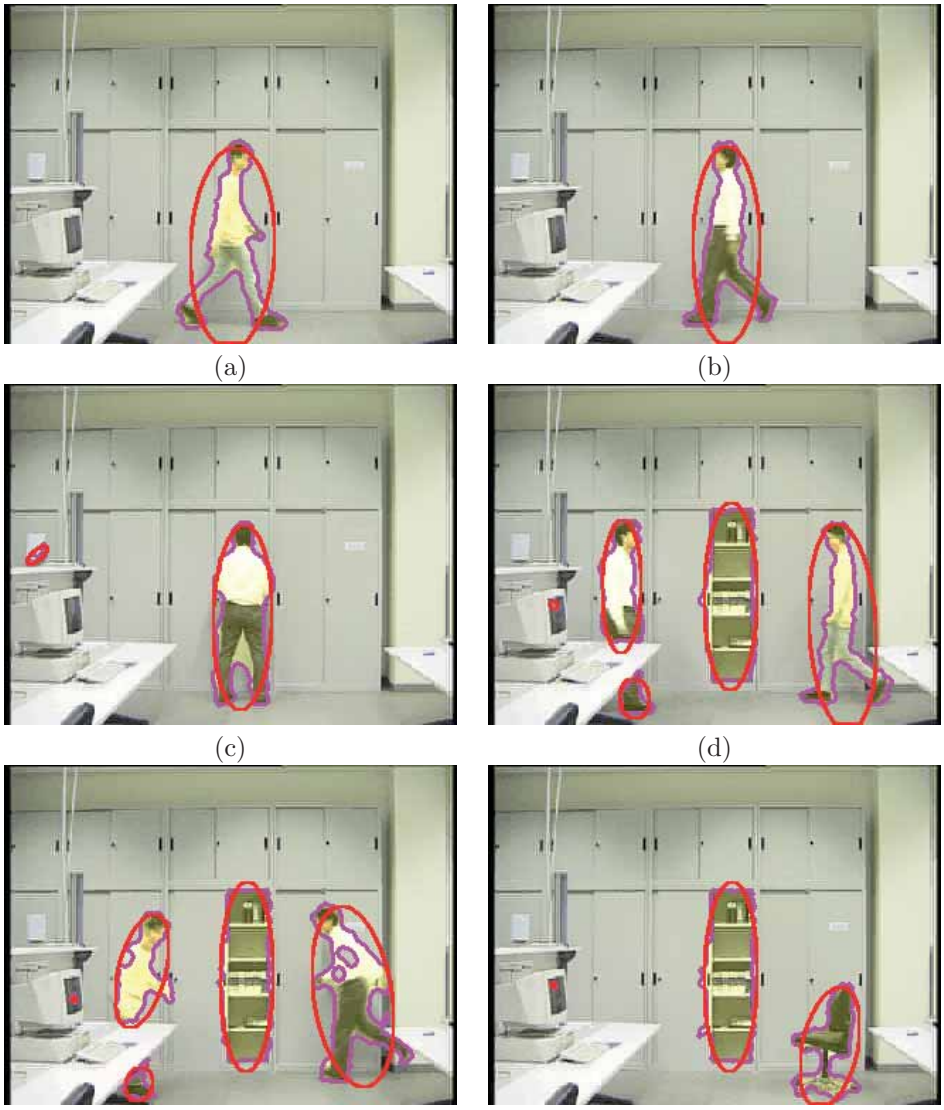
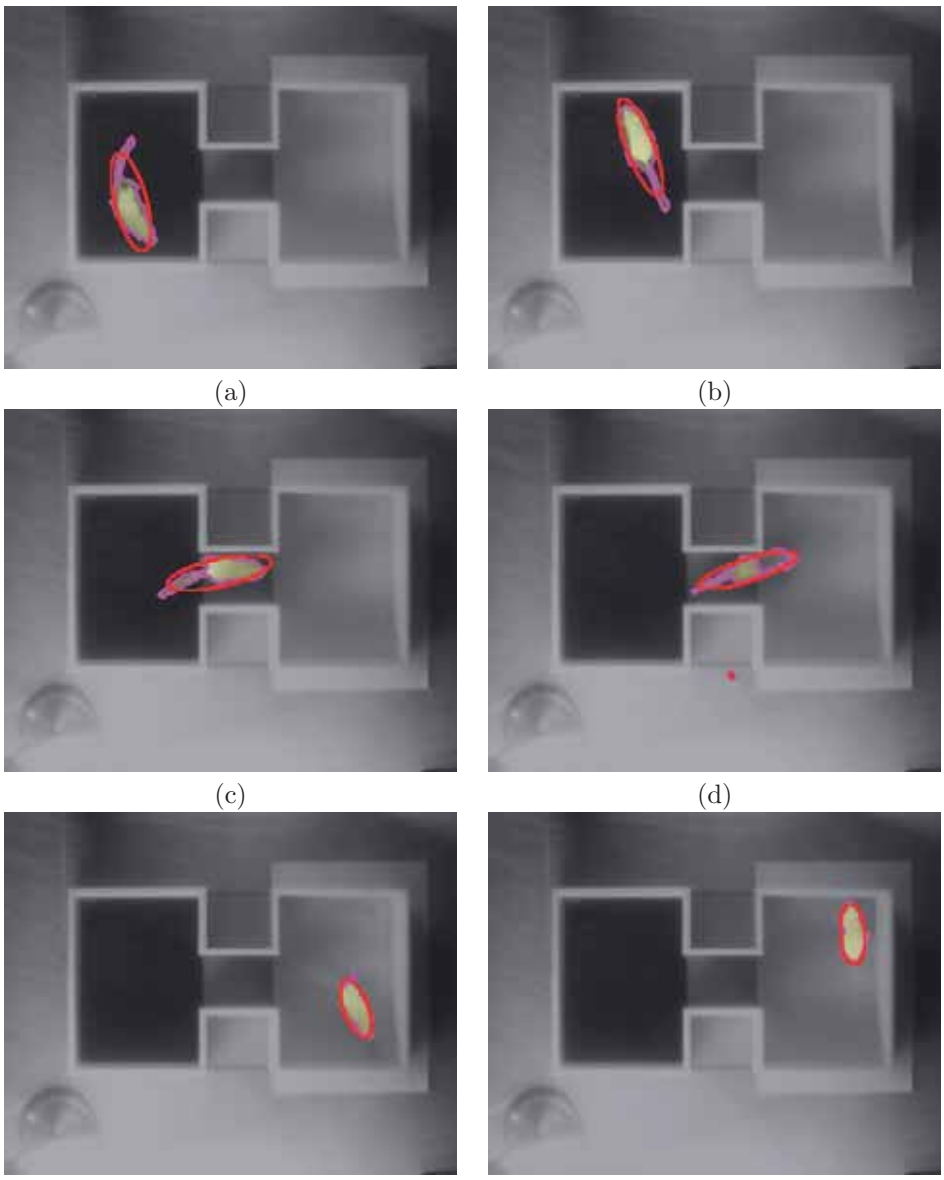


Figure 3.28: Foreground detection results from ATON_Laboratory sequence using our final approach. the two agents are correctly detected despite the big problems with the camouflage and the shadows presented. The agents appears in the scene walking and running at different speeds. The detection is achieved thereby showing that our approach is invariable to the frame rate. However, the camouflage problem sometimes appear in some part of the agents.



(a) (b) (c) (d) (e) (f)

Figure 3.29: Foreground detection results from Rats_BlackWhiteboxr sequence using our final approach. The rat is detected despite the problems with the camouflage and shadows presented. Notice that in spite of colour images the sequence does not have chromiance and all the detection achieved is using the dark and light camouflage process. However, the camouflage problem sometimes appear in some part of the rats due to the lack of change in intensity and chroma.

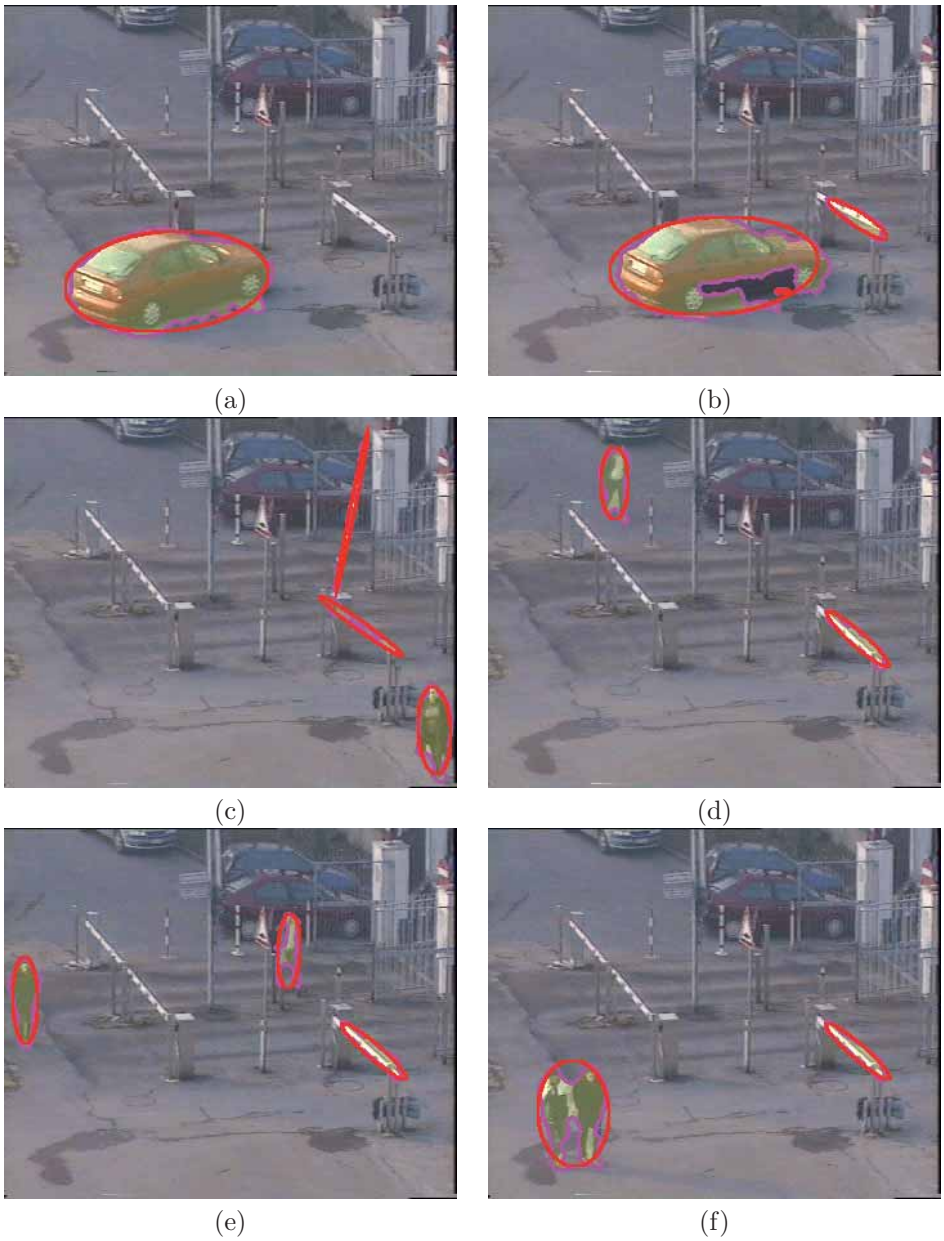


Figure 3.30: Foreground detection results from ATON_Campus sequence using our final approach. the agents and the cars are detected despite the problems with the low quality of the sequence, the blurred image and the strong shadows presented. However, the camouflage problem sometimes appear in some part of the car because the camouflage mask can not be correctly build.

low quality of the sequence, the blurred and noisy image and the strong shadows presented. However, the camouflage problem sometimes appear in some part of the car because the camouflage mask can not be correctly build.

Significant processed frames are depicted in Fig. 3.31, showing the detection results using our final approach in the Ismael sequence from UMIACS. In this figure, it can be seen how the agent is correctly detected despite the low quality of the image, and the big agent cluttered and occluded with background objects. However, sometimes the dark/light camouflage mask can not be correctly build due to camouflage in intensity and chroma at the same time. In this figure can be observed that our approach can obtain an accurate segmentation without taking into account the size of the foreground object.

Fig. 3.32 depicts some significant frames, showing the detection results using our final approach in the Msa sequence from MODLAB database. In this figure, it can be seen how the agent and the bag are correctly detected despite the problems with the reflected shadows in the floor and in the column. The bag is indefinitely detected because our approach does not have updating process, since this problem is not tackle in this approach.

Significant processed frames are depicted in Fig. 3.33, showing the detection results using our final approach in the indoor sequence S3_T7_A_Cam4 from the PETS2006 database. The scene presents multiple problems, saturation over the floor due to the different sources of illumination. Furthermore, multiple agents are in the scene with different colour appearance, therefore camouflage in intensity, in chroma and both are observed in the scene. The agents also exhibits strong shadows which are reflected over the floor. In this figure, it can be seen how all the multiple agents presented in the scene are correctly detected, thereby handling problems with strong shadows, saturations, dark and light camouflages, etc. Nevertheless, sometimes some parts of the agents one are not accurately segmented due to camouflage problem, see Fig. 3.33.(c). Furthermore, some shadows are segmented due to the problem with the floor reflectance and change in chroma, see Fig. 3.33.(i).

Significant processed frames are shown in Fig. 3.34 depicting the detection results using our final approach in the Camera1_070605 sequence from VISOR database, which is employed as main corpus in the VIDIVIDEO European project and in the VSNN06 conference. In this figure, it can be seen how the multiple agents are correctly detected despite the big camouflage between one of the agents with the floor and the columns, and the problems with the saturations and the soft shadows. The approach show an accurate detection independent on the number of the agents in the scene. In (e) one agent is reflected in the rear window and both are detected.

Fig. 3.35 depicts significant processed frames, showing the detection results using our final approach in a synthetic sequence from HERMES database. This sequence is a synthetic copy of the HERMES_Outdoor_Cam1 showed in the Fig. 3.19. In the Fig. 3.35, it can be seen how the agents and cars are correctly detected thereby showing that our approach is also able to work with synthetic images. The frames showed in the Fig. 3.35 are the same frames that are presented in the Fig. 3.19, both figures show similar results.

Significant processed frames are depicted in Fig. 3.36, showing the detection results using our final approach in a augmented reality sequence from HERMES

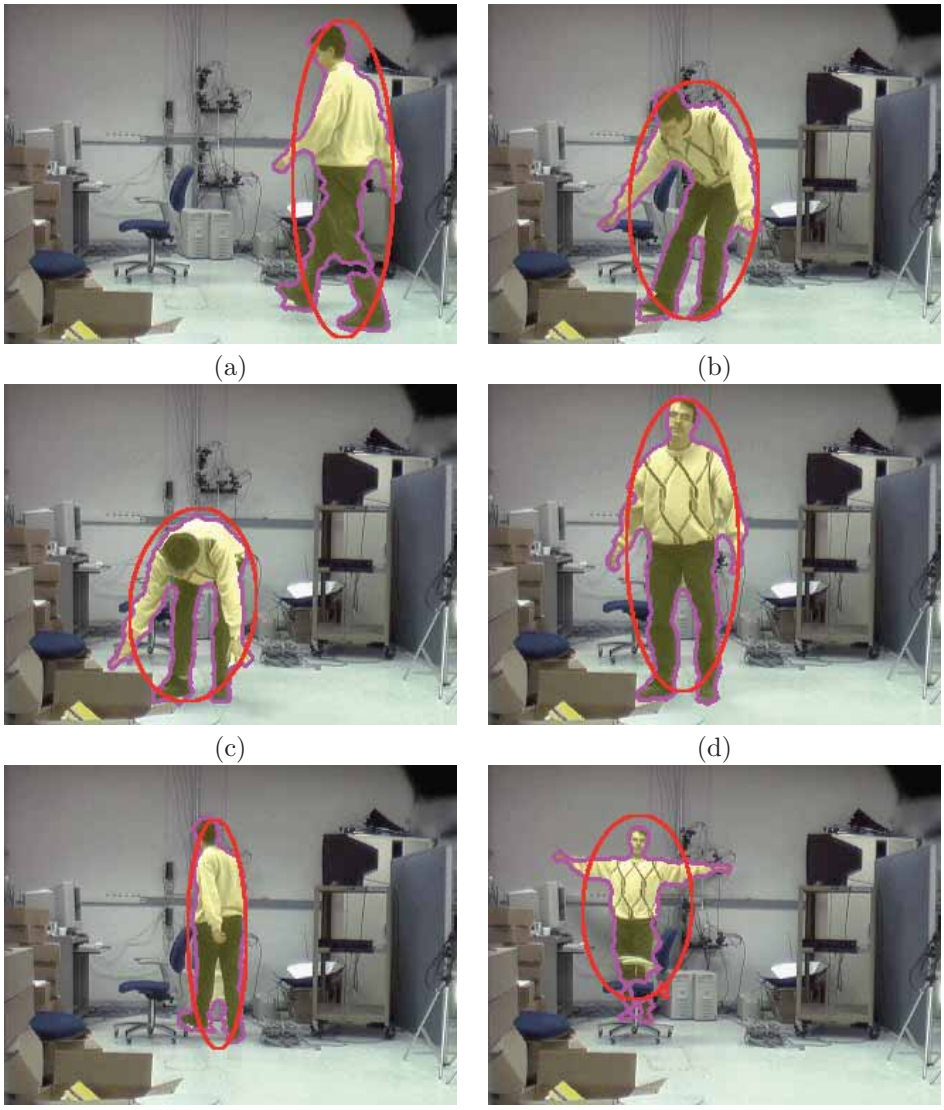
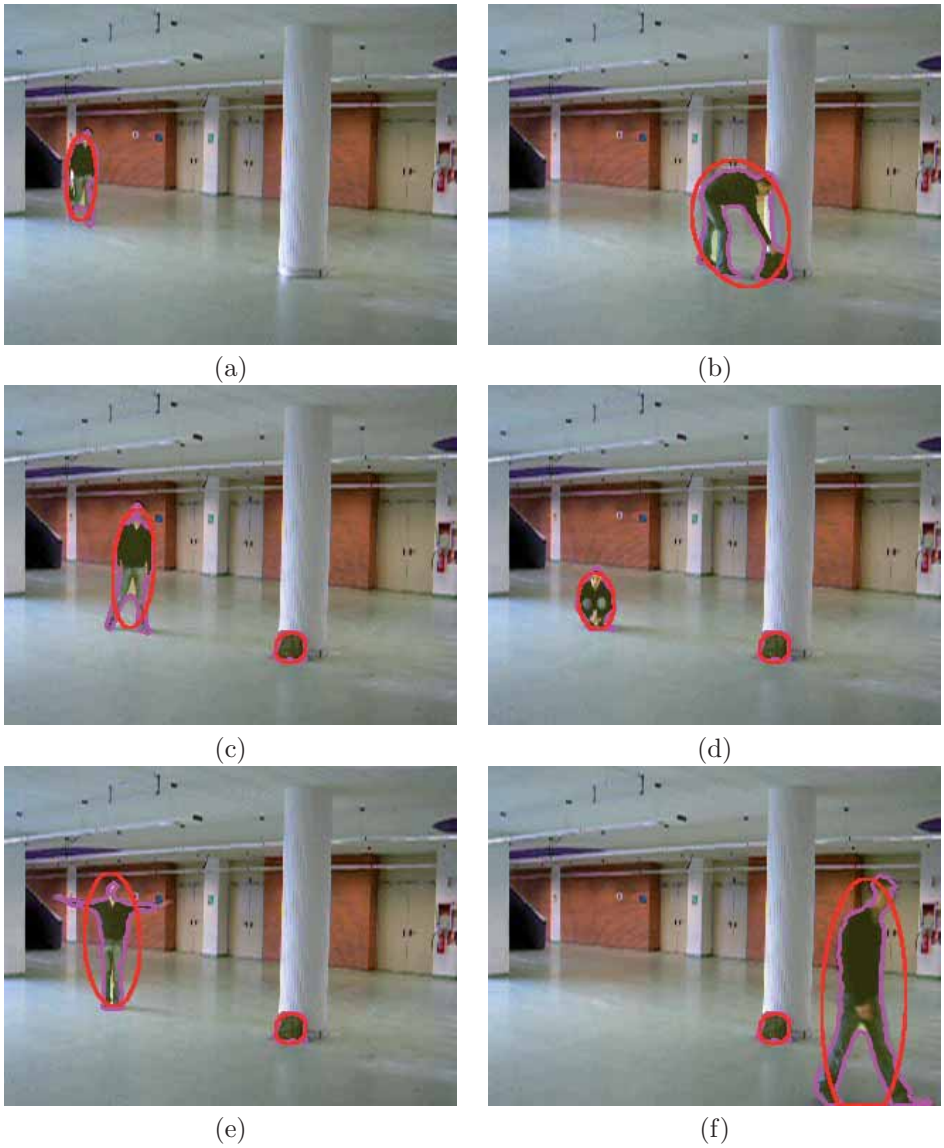


Figure 3.31: Foreground detection results from UMIACS_Ismail sequence using our final approach. The agent is correctly detected despite the low quality of the image, and the big agent cluttered and occluded with background objects. However, sometimes the dark/light camouflage mask can not be correctly build due to camouflage in intensity and chroma at the same time. Our approach obtains a accurate segmentation without taking into account the size of the foreground object.



(a) (b) (c) (d) (e) (f)

Figure 3.32: Foreground detection results from MODLAB_Msa sequence using our final approach. the agent and the bag are correctly detected despite the problems with the reflected shadows in the floor and in the column. The bag is indefinitely detected because it is not included in the background, since this problem is not tackle in this approach.

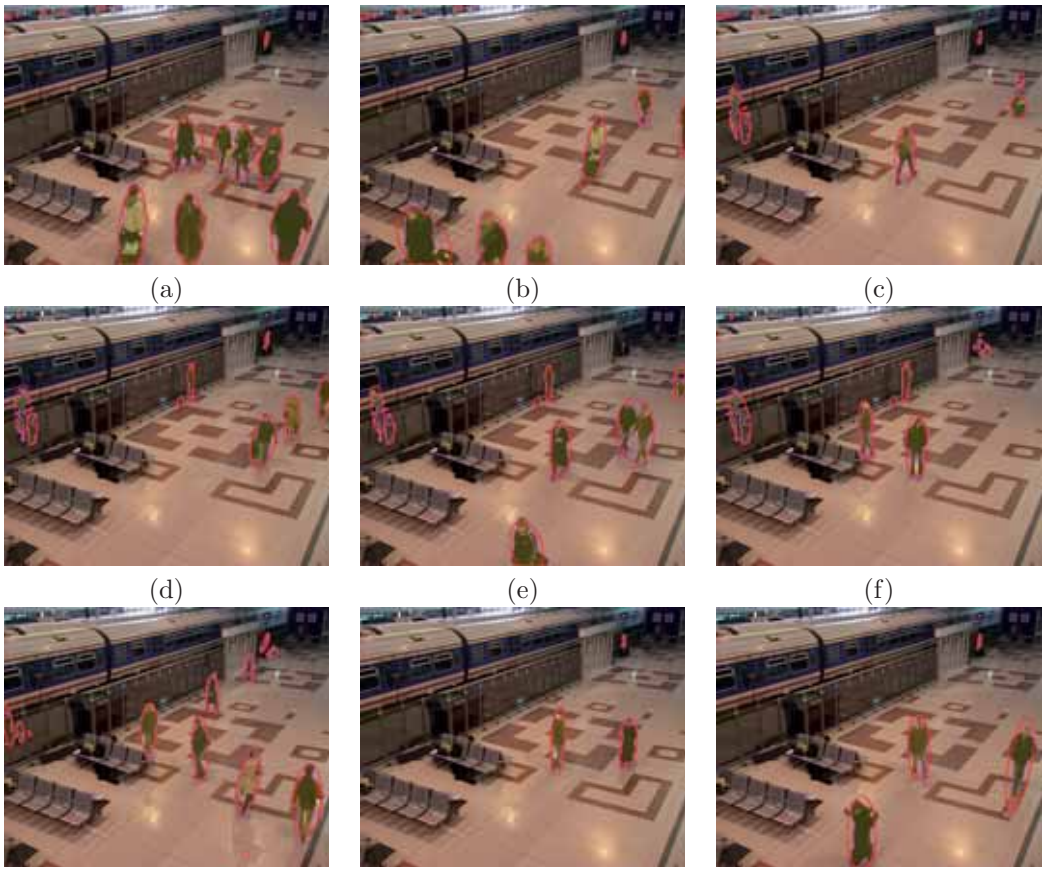
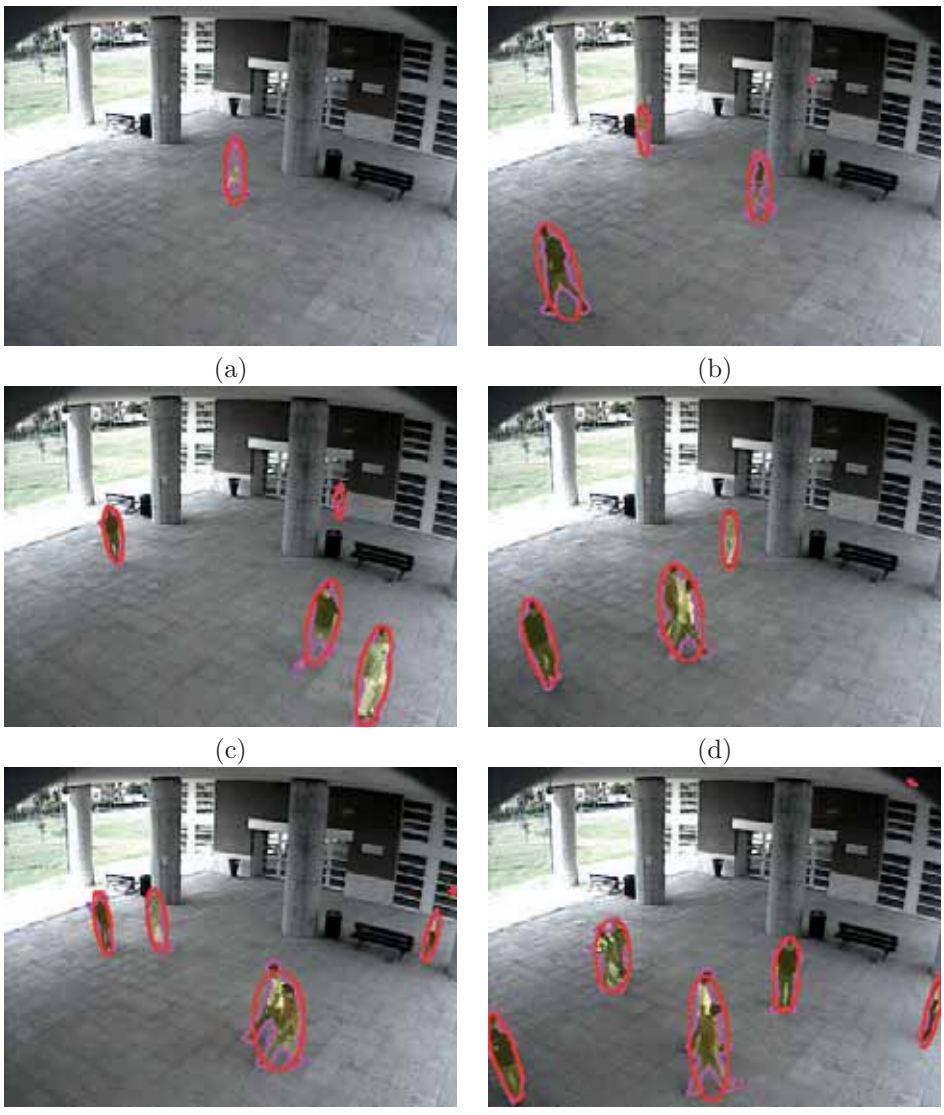


Figure 3.33: Foreground detection results from PETS_2006 sequence S3_T7_A_Cam4 using our final approach. All the agents presented in the scene are correctly detected despite the escene presents multiple problems, saturations, strong and reflected shadows over the floor, and there is several agents which presents different colour apparence allowing all camouflage problems. However, sometime some parts of the agent one are not accurately segmented due to camouflage problem, see (c). Furthermore, some shadows are segmented due to the problem with the floor reflectance and the change in chroma, see (i).

database. In this sequence the HERMES_Outdoor_Cam1 sequence, Fig. 3.19, is virtually augmented with agents and cars. In the Fig. 3.36, it can be seen how the agents and cars from the augmented reality are correctly detected, thereby showing that our approach is also able to detect the real one, but also the virtual ones.

Summarazing all the sequences analysed 3.24, 3.25, 3.19, 3.35, 3.36, 3.26, 3.33, 3.27, 3.28, 3.29, 3.30, 3.31, 3.32, 3.34, 3.21, shows a good detection despite all the problems described in the case analysis.

In spite of the good results achieved in all the database tested, there are some



(a) (b) (c) (d) (e) (f)

Figure 3.34: Foreground detection results from VSSN06_Camera1_070605 sequence using our final approach. The multiple agents in the scene are correctly detected despite the different agents colour appearence and big camouflage that exhibits some agents with the floor and the columns, and the problems with the saturation in the left part of the scene. The approach shows accuratelly segmentation independenty of the number of the agents in the scene.

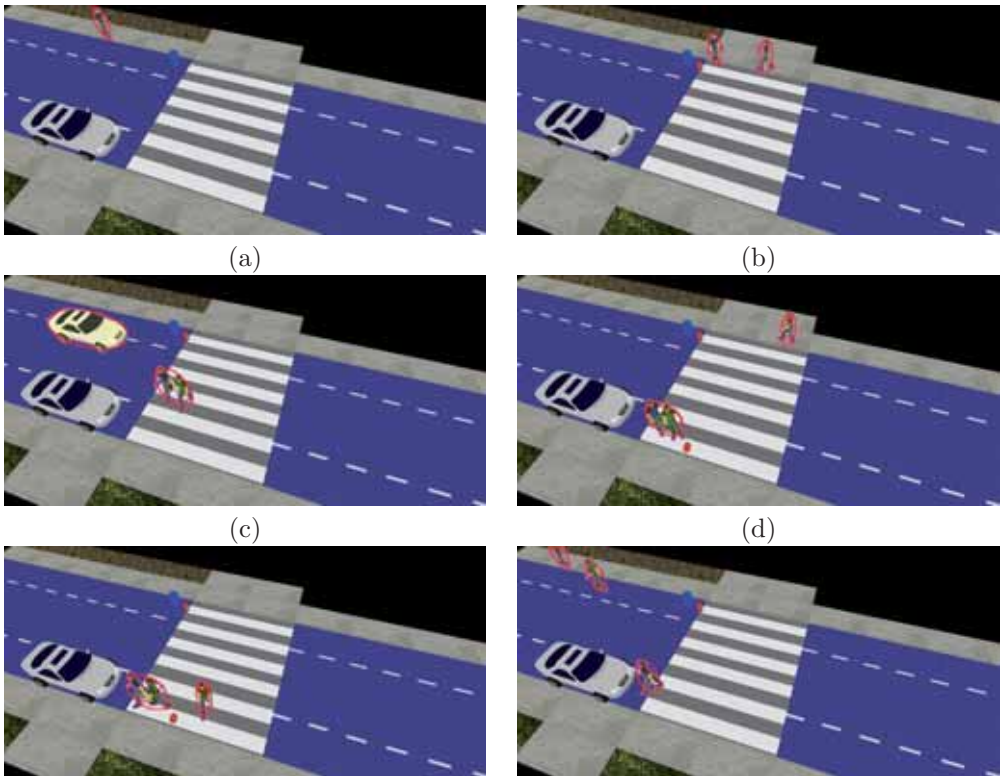


Figure 3.35: Foreground detection results from Synthetic HERMES_Outdoor sequence using our final approach. The cars and the agents are detected thereby showing that our approach is also able to work with synthetic images.

problems not solved yet. Some of them are problems not tackled in this thesis such as the background in motion, see figure 3.37, or when there is a camouflage in intensity and chromaticity at the same time. However, such as it can be seen in the figures 3.37, 3.38, 3.39, and 3.40 the shadow problem is also not solved. The main problem is that the approach is able to cope with achromatic shadow, however it fails when the chroma of the shadow change such as it is explained in the case analysis section 3.1.

Significant processed frames are depicted in Fig. 3.37, showing the detection results using our final approach in the Outdoor_Cam1 sequence from CVC database. This sequence contains most of the problems commented in the case analysis, heavily background in motion due to waving tree, chromatic shadows, saturations and all types of camouflages. Our approach achieves the detection of the agents, however the shadows are also detected because of the change in chroma. Furthermore, some tree pixels are erroneously detected because background in motion (waving tree) is not tackled in our approach.

(a) (b) (c) (d) (e) (f)

Figure 3.36: Foreground detection results from HERMES_Outdoor Augmented Reality sequence using our final approach. In this sequences has been incorporated virtual cars and agents. Our approach is able to detect the real one, but also the virtual ones.

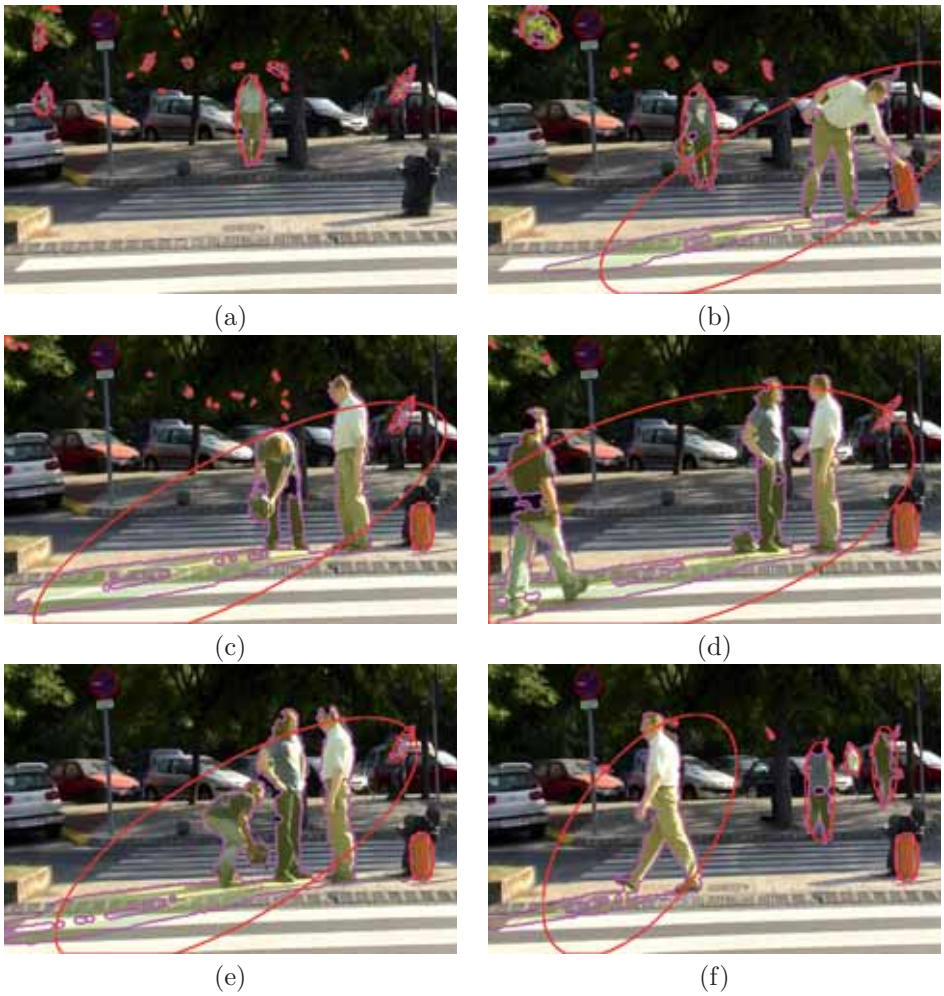


Figure 3.37: Foreground detection results from CVC_Outdoor_Cam1 sequence using our final approach. This sequence contains heavily background in motion due to waving tree, chromatic shadows, saturations and all types of camouflages. Our approach achieves the detection of the agents, however the shadows are also detected because of the change in chroma. Furthermore, some tree pixels are erroneously detected because background in motion (waving tree) is not tackled in our approach.

Fig. 3.38 shows significant processed frames depicting the detection results using our final approach in the HallwayI sequence from LVSN database. This sequence contains moving cast shadows from the agents in the floor and also in the walls which have a change in the chroma. Our approach is able to detect all the agents in the sequence, however most of the shadows are also detected due to this change in the chroma.

Significant processed frames are depicted in Fig. 3.39, showing the detection

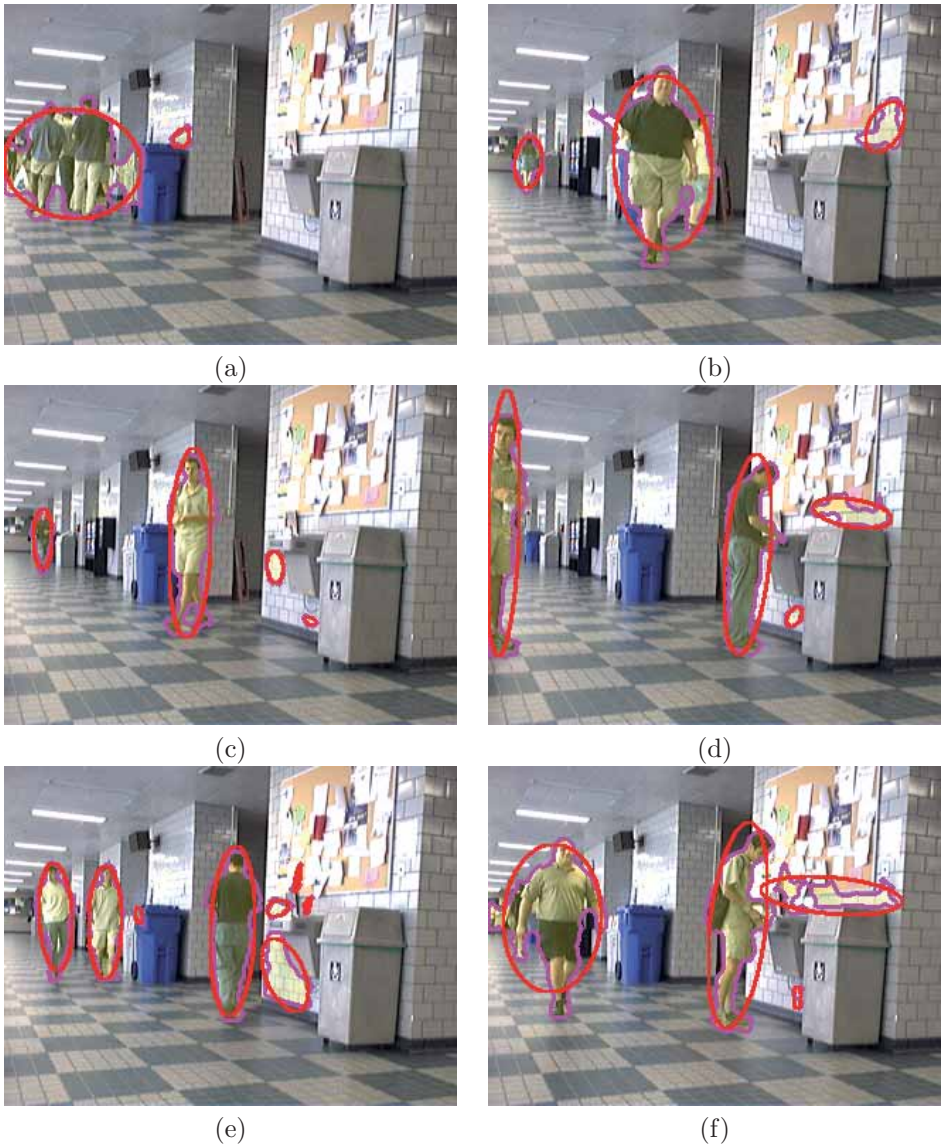


Figure 3.38: Foreground detection results from LVSN_HallwayI sequence using our final approach. This sequence contains moving cast shadows from the agents in the floor and also in the walls which have a change in the chroma. Our approach is able to detect all the agents in the sequence, however most of the shadows are also detected due to this change in the chroma.

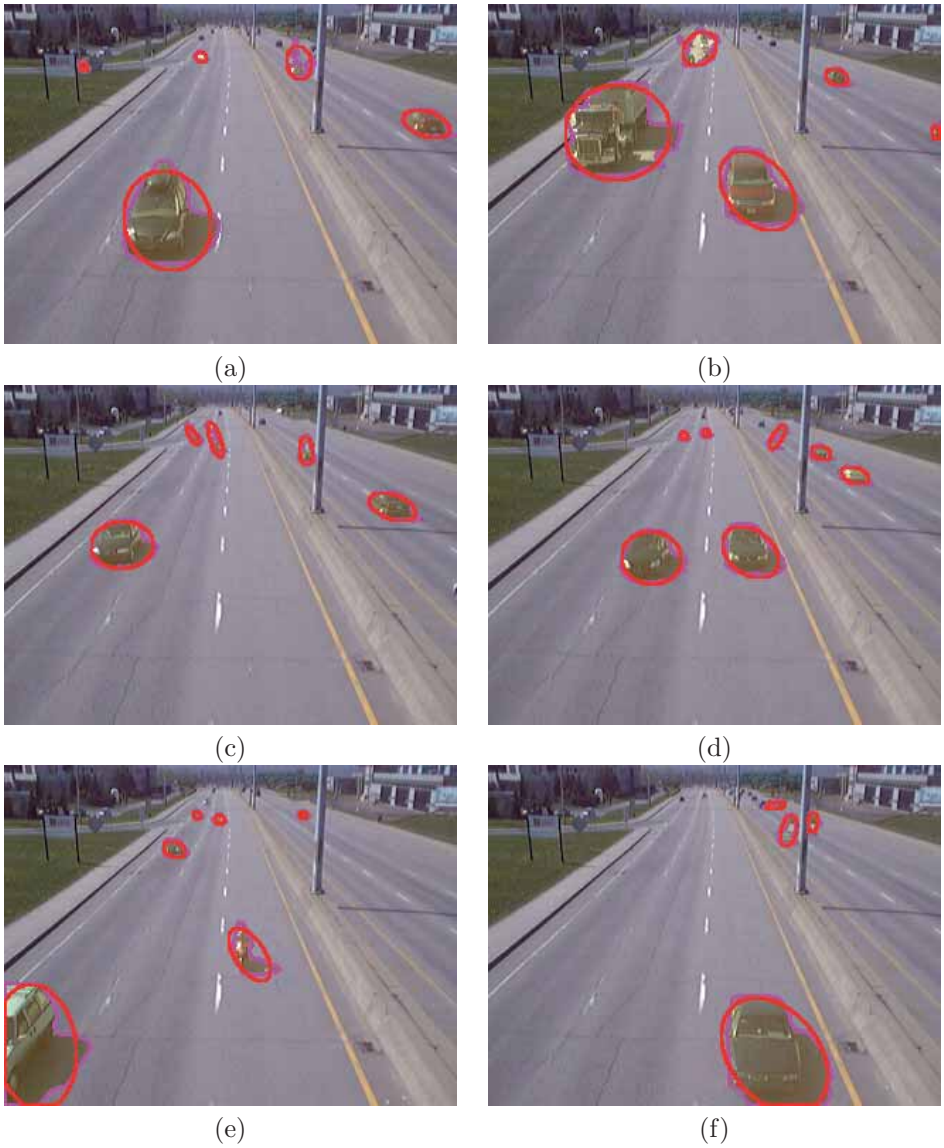


Figure 3.39: Foreground detection results from LVSN_HighwayIII sequence using our final approach. This sequence contains a highway with high density of vehicles. Our approach is able to detect all the vehicles and motorbikes in spite of all of them exhibits different colour appearance and moves fast compared with the frame rate of the image, however the shadows are also detected because the shadows have a change in his chromaticity. Our approach is independently in the size or the velocity of the foreground objects.

results using our final approach in the HighwayIII sequence from LVSN database. This sequence contains a highway with high density of vehicles. Our approach is able to detect all the vehicles and motorbikes spite of all of them exhibits different colour appearance and moves fast compared with the frame rate of the image, however the shadows are also detected because of the change in chroma. The figure shows that our approach is independently in the size or the velocity of the foreground objects.

Fig. 3.40 shows significant processed frames depicting the detection results using our final approach for the HERMES_ETSEdoor_day21sequence from the HERMES database.. This very long sequence contains strong moving cast shadows from the agents which have a change in the chroma. Our approach is able to detect all the agents in the sequence, however the chromatic shadows are also detected due to this change in the chroma.

The depicted frames from all the selected database shows that our approach is able to work in all type of scenes under uncontrolled environments.

Ultimately, some remarks on real-time requirements are here discussed. Significant speed improvements of the previously presented technique can be achieved because of the pixel-based nature of the approach, so the algorithm can be parallelizable. The current system is implemented as a *Matlab* prototype, without a careful code optimisation. Subsequent implementations of bottleneck modules in C++ have yielded speed improvements over 10-100 times the computation time of specific, most time-consuming functions. This would allow the system to process previously described sequences near real time.

3.4 Discussion

A case analysis of motion segmentation has been presented by taking into account the problems associated with the use of different cues such as colour, edge and intensity. This has allowed us to define when to use each model. Then, based on this case analysis, different motion segmentation problems have been solved.

The approach presented in this chapter combines colour, intensity and edge cues, and a temporal differencing technique in a collaborative architecture, in which each model is devoted to a specific task. The background model of each cue has been improved with respect to the current state of the art. A chromatic invariant cone model is used as colour model, and an invariant gradient orientation combined with their magnitudes is used as edge model, which is able to avoid false edges due to intense global illumination changes. These are performed by a particular algorithm, but they can be substituted by enhanced ones without modifying the architecture itself. Hence, this structured framework combines in a principal way the main advantage of each cue. In this way, by taking advantage of several cues, the system is allowed to benefit from all the cues' capabilities.

The proposed hybrid approach can cope with different colour problems as (i) dark and light foreground. Furthermore, it solves problems with (ii) the dynamic range (problems associated with saturation and lack of colour problems) using intensity cues. The approach also tackles (iii) camouflage in intensity and (iv) camouflage in chroma, (v) avoiding global and local (shadows and highlights) illumination problem.

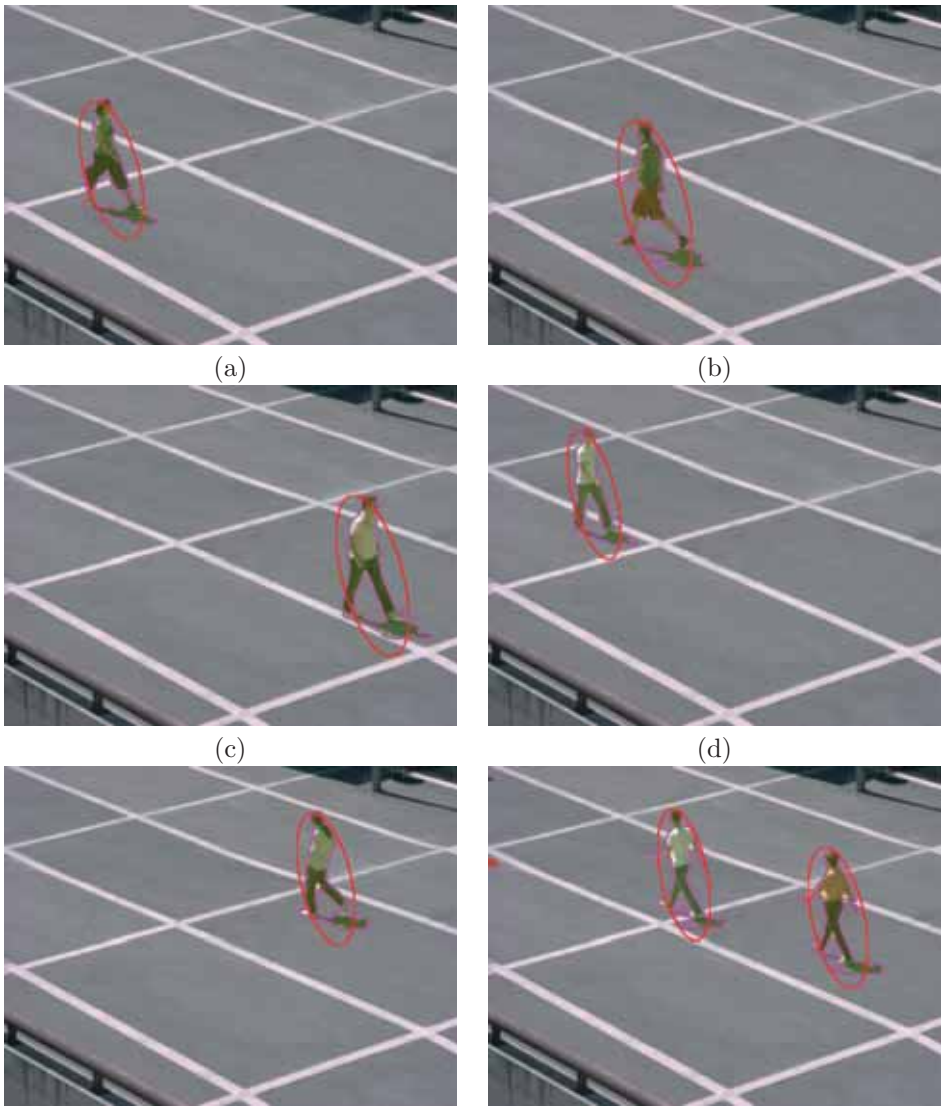


Figure 3.40: Foreground detection results from HERMES_ETSEdoor_day21_I4 sequence using our final approach. This sequence contains moving cast shadows which have a change in the chroma. Our approach is able to detect all the agents in the sequence, however the agent shadows are also detected due to this change in the chroma.

Therefore, it can simultaneously differentiate these camouflages from the illumination changes. In addition, the approach can cope (vi) with bootstrapping and (vii) ghosts problems. But also, it can (viii) reduce the false positives using each cue independently. Therefore, our hybrid approach reduces the number of false negatives and false positives, and increases the detection rate.

Experiments on complex indoor and outdoor scenarios have yielded robust and accurate results, thereby demonstrating the ability of our system to deal with unconstrained and dynamic scenes. Therefore, our approach can work in indoor, outdoor scenes, with high or low resolution, with noise and blurred images, and no need calibrated images. Furthermore, it is also independent on the illumination and the illuminant on the scene. Moreover, size, appearance, number, and velocity of the objects is not important for our motion segmentation approach. This is because it does not make any a-priori assumptions about camera location, surface geometries, surface textures, shape and types of the objects or the background.

Some remarks have to be considered, although it is not needed any calibration for the camera, and no matter where is situated, or the quality of the images from it. In order to use our motion segmentation approach the camera have to be fixed. Since, for modelling the background a static background is needed.

Our approach copes with the non-physical changes in the scene such as local and global illumination problems. Nonetheless, it does not cope with the physical changes in the scene such as when objects are deposited or removed from the scene. Then, in the future work, an updating process should be embedded to the approach in order to incorporate objects to the background model. Furthermore, the use of a pixel-updating process can help to reduce the false positive pixels obtained by using the intensity mask due to drastic illumination changes. In addition, detected motionless objects should be part of a multilayer background model. Furthermore, colour invariant normalisations or colour constancy techniques can be used to improve the colour model. The edge model can be enhanced avoiding false edges due to local intense illumination changes. Further, edge linking or B-spline techniques can be used to avoid the partial loss of foreground borders due to camouflage, thereby improving the edge mask. Lastly, the discrimination between the agents and the local environments can be enhanced by using of new cues such as texture information or high-level information such as tracking.

Chapter 4

Detection and Removal of Chromatic Moving Shadows

A fundamental problem for all automatic video surveillance systems is to detect objects of interest in a given scene. A commonly used technique for segmentation of moving objects is background subtraction [45], as stated in the previous chapters of this thesis. This involves detection of moving regions (i.e., the foreground) in an image by differencing the current image and a reference background image in a pixel-by-pixel manner. Usually, the background image is represented by a statistical background model, which is initialised over some time period.

An important challenge for foreground segmentation is the impact of shadows. Shadows can be divided into two categories: *static shadows* and *dynamic (or moving) shadows*. Static shadows occur due to static background objects (e.g., trees, buildings, parked cars, etc.) blocking the illumination from a light source. Static shadows can be incorporated into the background model, while dynamic shadows have shown to be more problematic. Dynamic shadows are due to moving objects (e.g., people, vehicles, etc.). The impact of dynamic shadows can be crucial for the foreground segmentation, and cause objects to merge, distort their size and shape, or occlude other objects. This results in a reduction of computer vision algorithms' applicability for e.g, scene monitoring, object recognition, target tracking and counting.

Dynamic shadows can take any size and shape, and can be both *umbra* (dark shadow) and *penumbra* (soft shadow) shadows. Penumbra shadows exhibit low values of intensity but similar chromaticity values w.r.t. the background, while umbra shadows can exhibit different chromaticity than the background, and their intensity values can be similar to those of any new object appearing in a scene. When the chromaticity of umbra shadows differs from the chromaticity of the global background illumination, we define this as *chromatic shadow*. Consequently, umbra shadows are significantly more difficult to detect, and therefore usually detected as part of moving objects.

When a shadow has successfully been detected it is usually removed instantly, since it is the object which is of interest for further processing and not the shadow. As a result, the information the shadow brings is lost. An interesting idea is to use this

information to improve other aspects of object and shadow detection and tracking. Concretely, if a detected shadow is tracked over time instead of being discarded, it could be used to improve the shadow detection and possible the object detection and tracking as well.

In this chapter, firstly a bottom up approach for detection and removal of chromatic moving shadows in surveillance scenarios is proposed. Secondly, a top-down approach based on a tracking system has been developed in order to enhance the chromatic shadow detection. In the bottom-up part, we present a novel technique based on gradient and colour models for separating chromatic moving shadows from detected moving objects. Firstly, both a chromatic invariant colour cone model and an invariant gradient model are built to perform automatic segmentation while detecting potential shadows. In a second step, regions corresponding to potential shadows are grouped by considering "a bluish effect" and an edge partitioning. Lastly, (i) temporal similarities between local gradient structures and (ii) spatial similarities between chrominance angle and brightness distortions are analysed for all potential shadow regions in order to finally identify umbra shadows.

In the top-down part, after detection of objects and shadows both are tracked using Kalman filters in order to enhance the chromatic shadow detection. Firstly, this implies a data association between the blobs (foreground and shadows), and Kalman filters. Secondly, an event analysis of the different data association cases are performed, and occlusion handling is managed by a Probabilistic Appearance Model (PAM). Based on this association, temporal consistency is searched in the association between foreground (FG) and shadow (SH) and their respective Kalman Filters, and several FG-SH association cases are studied. As a result, lost chromatic shadows are correctly detected. Finally, the tracking results are used as feedback to improve the shadow and object detection.

The remainder of the chapter is organised as follows. The related methodology in the field of shadow detection and object tracking will be discussed in section 4.1, along with our contributions to this subject. In section 4.2, the theoretical concept of our approach is outlined. The algorithm for foreground segmentation, along with the detection and removal of chromatic moving shadows are described in section 4.3. The top-down process used to enhance the shadow detection is described in section 4.4. Finally, we present experimental results in section 4.5 and concluding remarks in section 4.6.

4.1 Related Methodology

Shadow detection is an extensive field of research within computer vision. Even though many algorithms have been proposed in the literature, the problem of detection and removal of shadows in complex environment is still far from being completely solved.

A common direction is to assume that shadows decrease the luminance of an image, while the chrominance stays relatively unchanged [7, 33]. However, this is not the case in many scenarios, e.g., in outdoor scenes. Other approaches applies geometrical information. Onoguchi [48] uses two cameras to eliminate the shadows of

pedestrians based on object height. However, objects and shadows must be visible to both cameras. Ivanov et al. [28] use a disparity model, which is invariant to arbitrarily rapid changes in illumination, for modelling background. However, to overcome rapid changes in illumination, at least three cameras are required. in [57], Salvador et al. use the fact that a shadow darkens the surfaces, on which it is cast, to identify an initial set of shadowed pixels. This set is then pruned by using colour invariance and geometric properties of shadows. It should be noted that most of the approaches which applies geometrical information normally requires shadows to be on a flat plane.

Another popular approach is to exploit colour differences between shadow and background in different colour spaces. In [8], Cucchiara et al. use the hypothesis that shadows reduce surface brightness and saturation while maintaining hue properties in the HSV colour space. While Schreer et al. [58] adopt the YUV colour space. In [22, 33], Horprasert et al. and Kim et al. build a model in the RGB colour space to express normalised luminance variation and chromaticity distortions. However, these methods require all illumination sources to be white, and assume shadow and non-shadow have similar chrominance. A number of approaches use textures to obtain a segmentation without shadows, such as Heikkila et al. [21] which uses Local Binary Patterns. However, it fails to detect umbra shadows.

To overcome some of these prior mentioned shortcomings, some authors use colour constancy methods, combine different techniques or use multi-stage approaches. In addition to scene brightness properties, [63] uses edge width information to differentiate penumbra regions from the background. In [14], Finlayson et al. use shadow edges along with illuminant invariant images to recover full colour shadow-free images. Nonetheless, a part of the colour information is lost in removing the effect of the scene illumination at each pixel in the image. Weiss [71] uses the reflectance edges of the scene to obtain an intrinsic image without shadows. However, this approach requires significant changes in the scene, and as a result the reflectance image also contains the scene illumination. Martel et al. [41] introduce a nonparametric framework based on the physical properties of light sources and surfaces, and applies spatial gradient information to reinforce the learning of model parameters. Finally, [47] applies a multi-stage approach for outdoor scenes, which is based on a spatio-temporal albedo test and dichromatic reflection model. A comparative and evaluation study of shadow detection techniques can be found in [51].

For chromatic shadow detection we also apply a multi-stage approach inspired by [47] but we use colour and gradient information, together with known shadow properties. The contribution of our approach for chromatic shadow detection is threefold: (i) We combine an invariant colour cone model and an invariant gradient model to improve foreground segmentation and detection of potential shadows. (ii) We extend the shadow detection to cope with chromatic moving cast shadows by grouping potential shadow regions and considering "a bluish effect", edge partitioning, temporal similarities between local gradient, and spatial similarities between chrominance angle and brightness distortions. (iii) Unlike other approaches, our method does not make any assumptions about camera location, surface geometries, surface textures, shapes and types of shadows, objects and background.

In order to enhance the shadow detection a top-down architecture is presented in a second step, where Kalman filters are used for tracking. Shadows can be lost for

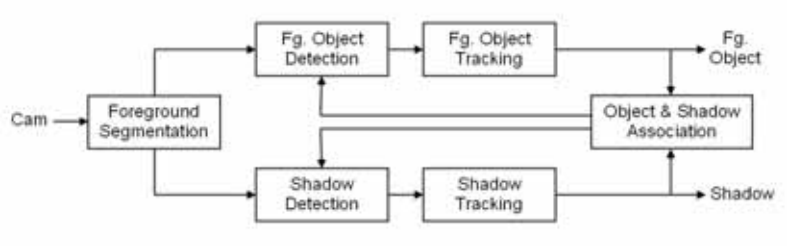


Figure 4.1: Flowchart for the shadow detection and tracking system.

a number of frames of a video sequence, and in these cases the use of Kalman filters to track the shadows can improve the shadow detection. Tracking is an extensive field of research within computer vision. In general tracking can be divided into four main approaches: point [68], kernel [43], silhouette [39] and body model tracking [34]. Body model tracking is computationally expensive and different models are required for different objects. Silhouette tracking suffers from initialization difficulties [23]. Furthermore, compared to point tracking, kernel tracking appears to be more robust, because point tracking has problems with occlusions, missdetection, entries and exits of objects. According to [75] colour is the most popular feature in kernel tracking. Many kernel tracking algorithms use a primitive geometric shape combined with appearance probability densities or appearance templates based on colour features, e.g. [59, 54, 73, 9, 56].

In order to enhance the chromatic shadow detection, a kernel tracking approach is therefore applied, which uses colour and edge information as features, along with shape and appearance descriptions for object representation. A high level scheme for the shadow detection and tracking approach can be seen in Fig. 4.1.

4.2 Analysis of Shadow Properties

Colour information ρ at a given pixel a obtained from a recording camera supposing Lambertian surfaces depends on four components: the Spectral Power Distribution (SPD) of the illuminant denoted $E(\lambda)$, the surface reflectance $R(\lambda)$, the sensor spectral sensitivity $Q(\lambda)$ evaluated at each pixel a and a shading factor σ.

$$\rho_a = \sigma \int E(\lambda)R(\lambda)Q_a(\lambda)d\lambda \tag{4.1}$$

The surface reflectance $R(\lambda)$ depends on the material. Hence, every material have different response to the same illumination change.

4.2.1 Applying the bluish effect

In outdoor scenes, the environment is illuminated by two light sources: a point light source (the sun) and a diffuse source (the sky) with different SPD $E(\lambda)$. Besides a reduction in the intensity, an outdoor cast shadow will result in a change of the

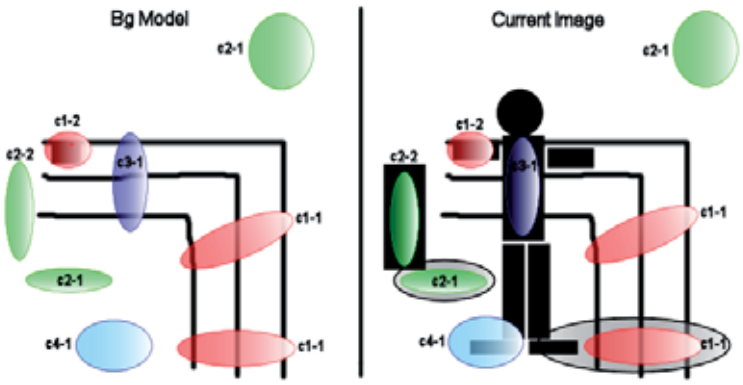


Figure 4.2: A sketch of the four main cases (c1-1 to c4-1) and two anormlies (c2-1 and c2-2) that can occur when performing foreground segmentation with the influence of shadows, and taking the temporal local gradients into account. The ellipses represent detection of potential chromatic shadows. They are grouped by considering an intensity reduction, "the bluish effect", and an edge partition.

chrominance. The illumination of the sky has higher power components in the lower wavelenghts λ (450 - 495 nm) of the visible spectrum, and it is therefore assumed bluish as argued in [47]. When the direct illumination of the sun is blocked and an region is only illuminated by the diffuse ambient light of the sky, materials appears to be more bluish. This "bluish effect" and the chrominance distortion can be exploited for shadow detection and grouping of potential shadow pixels.

4.2.2 Applying temporal local gradient information

By applying gradient information we can obtain knowledge about object boundaries, and thereby improve the foreground segmentation. Additionally, the gradient can also provide textural information of both the background and foreground image. Although shadows will result in a reduction in the intensity of the illumination, and the texture of a given object or the background will have lower gradient magnitude, the structure will still appear the same. Hence, the gradient orientation will be unchanged. This knowledge can be applied to identify shadows.

4.2.3 Shadow scenaria and solutions

When performing foreground segmentation with the influence of shadows, and taking the temporal local gradients into account, four main cases can occur as illustrated in figure 4.2. The ellipses represent detection of potential chromatic shadows. They are grouped by considering an intensity reduction, "the bluish effect", and an edge partition. The entire shadow detection process will be explained in depth in section 4.3.

Case 1: Local gradient structures are present in the background model and in the current image, and they are similar. By examining similarities between the local

gradients, and the fact that there is no foreground object in the current image, potential shadows can be detected and identified as shadow regions (case 1-1). However, if a foreground object is present, it can be miss-classified as shadow if the gradients of the background and the foreground object are similar (case 1-2).

Case 2: There is no available background model nor local gradients in the current image. Since, the change in illumination of all the potential shadow regions has to be similar, temporal and spatial similarities between chrominance angle and brightness distortions within the potential regions are analysed to detect chromatic shadows (case 2-1). However, a foreground object can be miss-classified as shadow if the foreground object has no gradients. Furthermore, the chrominance angle distortion can also be similar among the pixels in the region of the object (case 2-2).

Case 3: Local gradient structure is present in the background model but not in the current image. By examining similarities between temporal gradients, potential shadow can be detected as foreground object if there are background gradients and a new foreground object in the current image.

Case 4: Local gradient structure is present in the current image but not in the background model. Then there must be a new foreground object in the current image. In this case, the gradients in the current image are employed to detect shadow regions. Hence, there is no need to analyse the potential region further.

The described characteristics are not sufficient to address these anomalies in case 1-2 and case 2-2. Therefore, we take further precautions and apply some additional steps, which will be explained in section 4.3. Furthermore, it should be noted that these additional steps also improves the shadow detection in some of the four main cases.

4.3 Chromatic Shadow Detection

The approach, depicted in Fig. 4.3, is a multi-stage approach. The first three stages remove the pixels which cannot be shadow pixels. The fourth step divide the regions of potential shadows. Chromatic shadow detection is realised in stage 5 and 6 based on gradients and chrominance angles, respectively. The last step avoid foreground regions detected erroneously as chromatic shadows.

4.3.1 Moving foreground segmentation

The moving foreground segmentation is obtained from the previous chapter. Where an improved hybrid approach which fuses colour and gradient information is used. Note that the this approach can cope with several motion segmentation anomalies, among them it can cope with penumbra shadows because it is based on a chromatic colour model [22]. It also provides the highest detection rate in comparison to other motion segmentation approaches, such as it is stated in the experimental results from

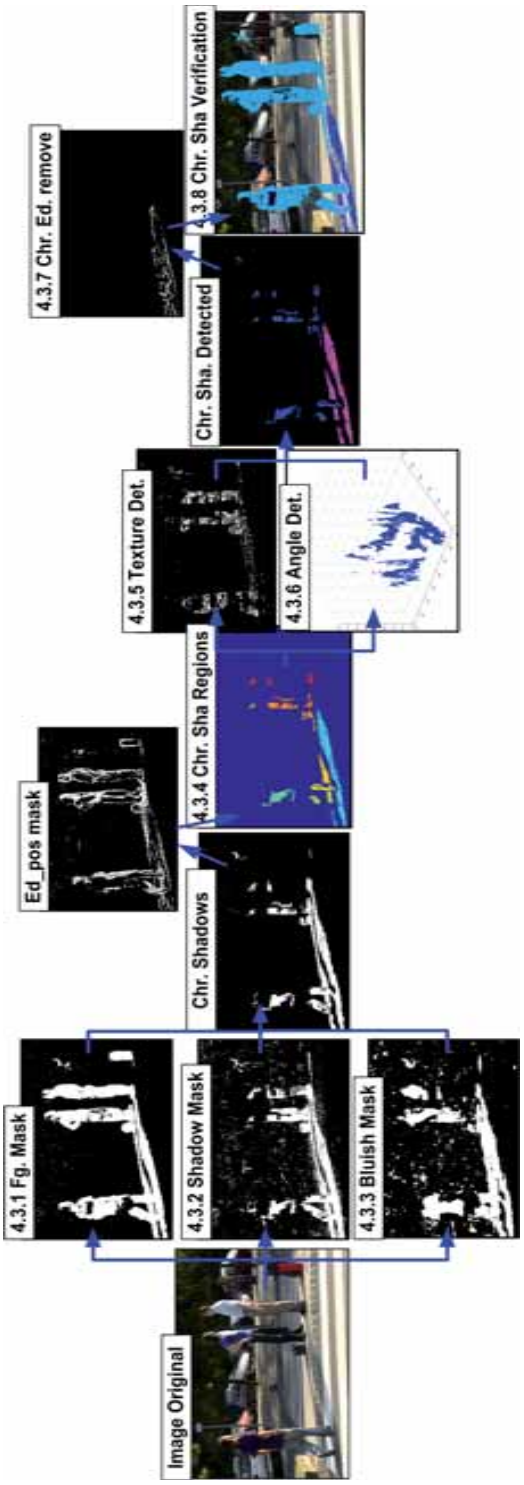


Figure 4.3: An overview of the chromatic shadow detection approach. Each of the numbers added to the image captions corresponds to the sub-sections in section 4.3.

the previous chapter. In order to get more accurate segmentation the approach use an invariant chromatic cone model and invariant gradient model which are perfect for our purposes in the next stages.

The chromatic cylinder model employed in many motion segmentation approaches [25, 22, 33] is changed into a new chromatic cone model. It uses chrominance angle distortion instead of chromatic distortion. For the same chromaticity line the chromatic distortion used in the above mentioned papers depends on the brightness distortion, while the chrominance angle distortion is invariant to the brightness, as it can be seen in Fig. 3.7 (the chromatic distortion δ increases proportional to the brightness distortion α, while the chrominance angle distortion β is equal). The invariant chromatic cone model is more robust towards chromatic shadows because these shadows (umbra shadows) modifies both the brightness and the chromaticity.

As argued in [41, 53], the gradient model has to be invariant towards global and local illuminations changes, such as shadows. The new invariant gradient model presented in the previous chapter uses a new combination of gradient magnitudes and gradient directions which is invariant to illumination changes. Hence, it can be used in order to identify the different local gradient structures of the scene.

The use of the invariant models provides a high detection rate in comparison to other motion segmentation approaches. After the initial detection, moving foreground objects, chromatic shadows and some isolated pixels are contained in a binary mask named *M1*. Furthermore, the mask obtained using the gradient model is divided into two masks, which are used for the next steps. The *Edneg* mask corresponds to the foreground pixels belonging to the background model. While the *Edpos* mask corresponds to the foreground pixels belonging to the current image. A third mask is also created called *Edcom*, which contains the common edges detected in the background model and in the current image.

4.3.2 Shadow intensity reduction

In this step the *M1* mask from step 1 is reduced in order to avoid pixels which cannot be shadows. A foreground pixel cannot be a shadowed pixel if it has a higher intensity than the background model. Then, a new mask for this step is created according the next equation:

$$M2_{a,t} = (I^R_{a,t} < \mu^R) \wedge (I^G_{a,t} < \mu^G) \wedge (I^B_{a,t} < \mu^B) \tag{4.2}$$

where a corresponds to the pixel location in the *M1* mask.

4.3.3 The bluish effect

The effect of illuminants which are different than white lights provokes chromaticity changes because the changes in the intensity are different for every channel. In outdoor sequences the main illuminants are the sky and the sun (any of them white illuminant). The sky is the only source of illumination on shadowed regions, and the sky is assumed to be bluish as argued in [47]. Therefore, the intensity changes in the red and green

channels are bigger than in the blue channel. This knowledge can be used to reduce the shadow region detected in the previous step (*M2*):

$$M2_{a,t} = (I^R_{a,t} - \mu^R) > (I^B_{a,t} - \mu^B) \wedge (I^G_{a,t} - \mu^G) > (I^B_{a,t} - \mu^B) \tag{4.3}$$

where a corresponds to the pixel location in the *M2* mask. Obviously, the bluish effect cannot be applied in indoor sequences.

4.3.4 Potential chromatic shadow regions

It is supposed that shadow regions have the same intensity change for each channel, since the illuminant is similar for all the shadowed region. However, different surfaces have different reflectance characteristics. Hence, the change in intensity depends on the surfaces material for the given shadow pixels. However, edges can show the changes between continuous pixels. Therefore, using the foreground edges detected in the current image, mask *Edpos*, the potential shadow regions can be separated from the moving foreground objects.

$$M3_{a,t} = M2_{a,t} \wedge (\neg Edpos_{a,t}) \tag{4.4}$$

A minimum area morphology is applied in order to avoid smaller regions which do not contain enough information for the subsequent steps of the shadow analysis.

4.3.5 Chromatic shadow gradient detection

In this step the temporal gradients of the regions detected in the previous mask M3 are analysed, in order to identify in which case of the theoretical shadow analysis (see section 4.2) each of the regions complies with. A region will be considered as a shadow if it complies with case 1. Negative foreground edges (Edneg mask) inside of the region are compared to the common foreground edges (Edcom mask), in order to prove if the region is a shadow and avoid the anomaly case 1-2. Furthermore, it also test if the negative edges are noise (larger regions have a higher probability to contain negative edges from noise):

$$\begin{aligned} Tx_b = & \left(\frac{\sum_{a \in R_b} (R_b \wedge Edneg)}{|R_b \wedge Edtot|} \cdot k_n < \frac{\sum_{a \in R_b} (R_b \wedge Edneg)}{|R_b \wedge Edtot|} \right) \\ & \wedge \left(\frac{\sum_{a \in R_b} (R_b \wedge Edneg)}{|R_b|} < k_s \right) \end{aligned} \tag{4.5}$$

where a is the pixel position; R_b is the evaluated region and b is the number of the region; $|R_b|$ denotes the number of pixels of region b; $|R_b \wedge Edtot|$ denotes the number of pixels representing the edges detected in the background model and the current image; k_n corresponds to a confidence region, which is equal to the probability of the region belongs to a shadow or a foreground object; and k_s is used to measure if the negative edges corresponds to noise. Again larger regions have a higher probability to contain negative edges from noise.

4.3.6 Chromatic shadow angle and brightness detection

In this step the temporal and spatial similarities of the chrominance angle and brightness distortion for all pixels belonging to regions, which have so far not been classified as shadow, are analysed. A region will be considered as a shadow if it complies with case 2. The only regions analysed in this section will be those that does not have gradients, neither in the background model nor in the current image. If the pixels do not have gradient, nor similar chrominance angle distortion and do not have a significant brightness distortion, then the region will be classified as shadow.

$$\begin{aligned} &ABd_b = \\ &\left(\frac{\sum_{a \in R_b} (R_b \wedge Edtot)}{|R_b|} < k_t\right) \wedge \left(\frac{\sum_{a \in R_b} (R_b \wedge \breve{\alpha})}{|R_b|} < k_a\right) \\ &\wedge \left(\frac{\sum_{a \in R_b} (R_b \wedge \breve{\beta})}{|R_b|} < k_b\right) \end{aligned} \tag{4.6}$$

where $\breve{\alpha}$ and $\breve{\beta}$ are the chrominance angle and brightness normalised distortions calculated for each pixel in the region number b (R_b), respectively; k_t is a confidence region to avoid noise gradients; k_a and k_b is a minimum threshold used to determine if the angle and brightness distortion are similar among the pixels of the evaluated region.

4.3.7 Chromatic shadow edge removal

Pixels from the potential shadow regions, which were neglected in section 4.3.4 because they were part of the *Edpos* mask, have to be included again in the regions destected as shadow.

4.3.8 Shadow position verification

A moving cast shadow is always caused by a moving foreground object. Therefore, in this section it is tested if a detected shadow has an associated foreground object, in order to avoid the anomaly in case 2-2. Only shadows detected in the chrominance angle and brightness distortion analysis (section 4.3.6) will be tested. During a training period T_2, the chrominance angles between the detected shadows and the foreground objects are calculated. After, the most probable chrominance angle obtained in the training period is used to discard detected shadows, which do not have any foreground object in the direction of the chrominance angle.

4.4 Top-down shadow detection

When a shadow has successfully been detected it is usually removed instantly, since it is the object which is of interest for further processing and not the shadow. As a result, the shadow information is lost. An interesting idea is to use this information a posteriori in order to improve the shadow detection when it fails (e.g., due to

camouflage problems). Concretely, if a detected shadow is tracked over time instead of being discarded, it could be used to improve the shadow detection.

In this section a top-down approach is used in order to enhance the chromatic shadow detection using a Kalman filter based tracking. Fig. 4.4 shows an overview of the top-down shadow detection enhancement process, and the algorithm is listed in Algorithm 2.

Firstly, the tracking module tracks objects and shadows through the scene. As input, the tracking module receives a binary mask from the object and shadow detection described in the previous section, as illustrated in Fig. 4.4. In the following subsections the tracking is explained with special attention on data association, the event analysis and occlusion handling (sections 4.4.1, 4.4.2 and 4.4.3, respectively). The output of the tracking is a list of tracks for each object and shadow and their mutual association, which is used as feedback to improve the object and shadow detection. Secondly, this association between objects and shadows is described and updated for the Kalman Filters (KF), sec. 4.4.4. Thirdly, temporal consistency is investigated for the association between FG and SH blobs and their assigned KF, in order to find the possible lost shadows, sec. 4.4.5. Once the chromatic shadows are detected then they are recovered in the original image, sec. 4.4.6. Finally, the KF and the PAM are updated taking into account the information from the new data association, and used for tracking in the next frames, sec. 4.4.7. An overview of the entire process can be seen in the Fig. 4.4.

4.4.1 Tracking using Kalman Filters

The detected foreground objects and shadows are tracked using first order Kalman filters. The tracking and data association are based on a number of estimated parameters for the detected objects:

- Centroid of an ellipse fitting.
- Major and minor axis length of the ellipse.
- Probabilistic Appearance Model (PAM).

Each track is therefore associated with these parameters, and a Kalman filter is used to predict the object's location using a first order motion model. Refer to appendix A for further deatils on the Kalman filter. Therefore, the target state is defined by $x_t = (posx_t, posy_t, velx_t, vely_t, maj_t, min_t, \theta)$, which establishes a state vector for every observation, and adds the target speed and the size change rate at time t. Where $posx_t$ and $posy_t$ define the position (centroid of the ellipse), $velx_t$ and $vely_t$ the velocity, and maj_t and min_t the major and minor axis, respectively, and θ the orientation.

4.4.2 Data Association between blobs and KF

When performing data association five situations can occur, as shown in fig. 4.5. A new object means that a new track is created. A lost object means that a track is destroyed if the object does not reappear within a certain number of frames (Tdead).

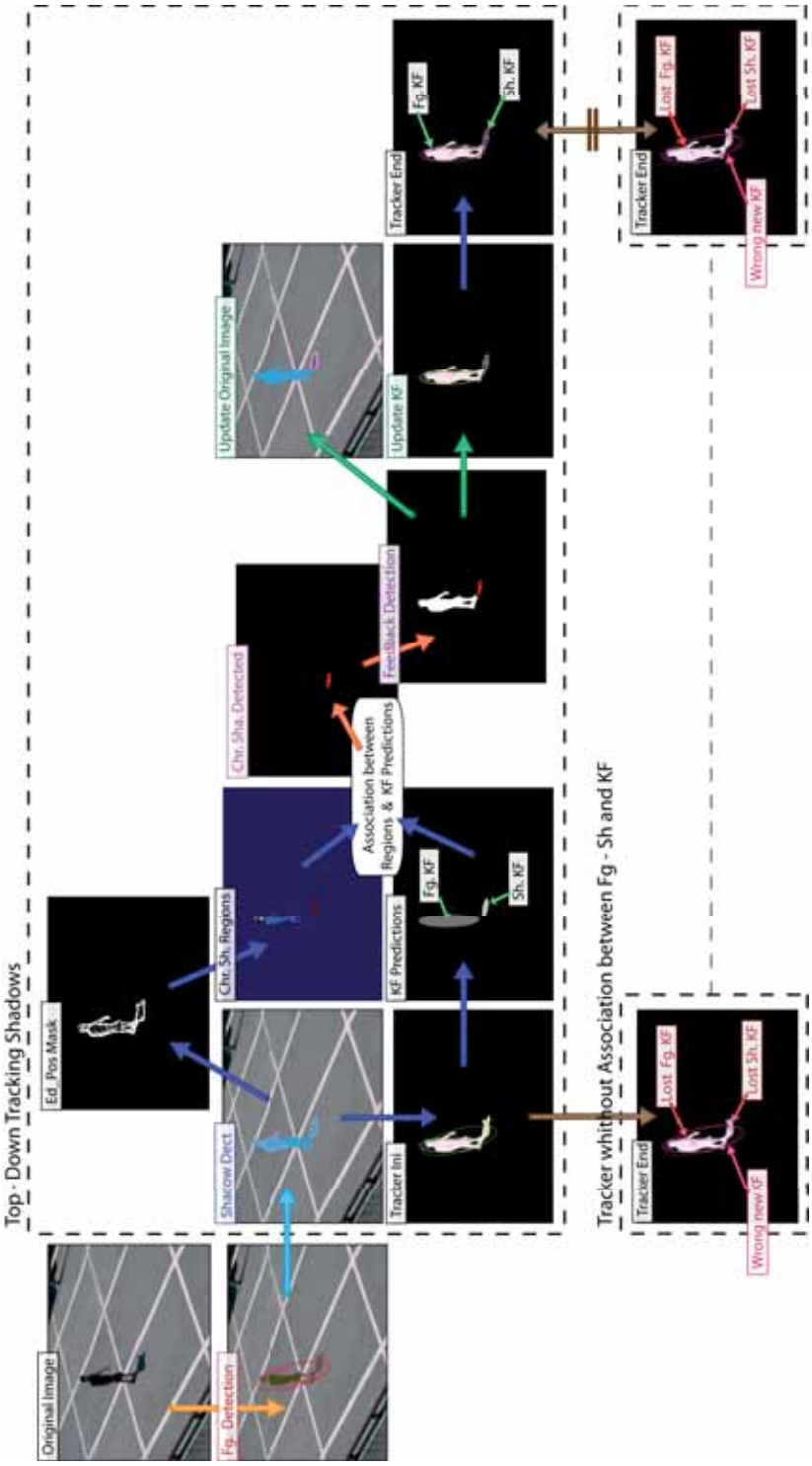


Figure 4.4: An overview of the top-down process to enhance the chromatic shadow detection. The figure illustrates the steps within the tracking and the motion segmentation, in order to enhance the shadow detection when a shadow is lost (lost shadow case). See the main body text for more details.

Algorithm 2 Top-down shadow detection approach

For each blob from the chromatic shadow detection:

- New Kalman Filter (KF) for each new blob, and delete KF for the KF not use in a period of time (Tdead).
- KF Prediction: Time update KF.
- Data Association between blobs (FG and SH blobs) and KFs.
  - Probabilistic Appearance Model (PAM) for each KF.
  - Build weights for the association: two correspondence matrix based on:
    - Euclidean distance based on position (x,y) and size (major and minor axes of an ellipse).
    - Matching of PAM and blob.
  - Case detection, see Fig. 4.5 and Algorithm 3:
    - Five possible cases: object match, new object, lost object, object splitting, and object merging.
  - Association between the blobs (FG and SH blobs) and the KFs.
  - Manage the KFs: updating, creating and deleting the KFs.
- Update KF-info: the association information between the blobs and the KFs.
- Temporal consistency in the data Association between FG and SH and their assigned KFs.
  - Case detection, see Fig. 4.10:
    - Three possible cases: FG and SH match, new shadow (FG-SH splitting), and lost shadow (FG-SH merging).
  - Lost Shadow case:
    - Possible shadow regions from the original FG blob.
    - Build weights for the association: two correspondence matrix based on:
      - Euclidean distance based on position (x,y) and size (major and minor axes of an ellipse).
      - Matching of blobs.
    - Association between KF predictions of FG and SH and the regions extracted from the original FG blob.
- Feedback (top-down) from the tracking to the shadow detection:
  - Classify the original image using the data association and the new FG and SH blob information.
  - Update blob information for the original image.
- Manage the KFs: updating and deleting the KF:
  - Update the KF info related to the new associations between new FG and SH blobs and their correspondent KFs.
  - Delete and create new KF if it is needed.
- KF Prediction of the new KF created: Time update KF.
- KF Correction: Measurement update.

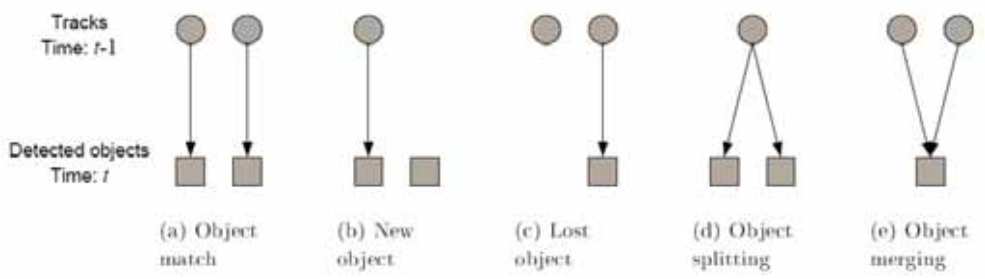


Figure 4.5: The five possible data association situations between objects (blobs) and Kalman Filters (KFs). A circle illustrates a track and a square illustrates a detected object. An arrow indicates association between a track and a detected object.

The object match situation is a one-to-one match, meaning that the track is updated using the detected object assigned to it. Object splitting means that more than one detected object match a track. This is resolved by selecting the object with the highest probability of the matches and creating new KFs for the other objects. Object merging means that a single detected object matches two or more tracks, this is caused by inter-object occlusion, and is handled using probabilistic appearance models as described in section 4.4.3.

Data Association Algorithm

The foreground blobs extracted as described in the previous chapter 3, and classified as foreground or shadow in section 4.3, are associated with a list of possible Kalman filters using Algorithm 3, which is based on the stable marriage algorithm [18].

If the object is identical to the track, then the position and shape of the new blob and the KF have to be similar. Additionally, the colour appearance also has to be similar. In order to test it, the object must have the shortest Euclidean distance between all possible blobs and all possible KFs. Hence, the Euclidean distance is calculated for every blob and KF combinations, in order to compare the position (centroid x,y) and the size (major and minor axis length). The maximum match probability between the PAMs is also computed for the blobs and the KFs, see Algorithm 3.

In order to determine if the object under evaluation is a new object, the best match between the new blob and the KF is compared with a maximum Euclidean distance and a minimum probability for the PAM. When the KF is lost then only the distance can be compared, since the PAM is centred in the object. To know if the blob (FG or SH) corresponds to a previous KF, or if it is a new object, the condition set by equation 4.7 has to be fulfilled:

$$\begin{aligned} \text{if } ((\text{dist}(\text{newobj,KF}) < \tau_{maxD}) \vee (\text{distPAM}(\text{newobj,KF}) > \tau_{minP})) \\ \vee ((\text{dist}(\text{newobj,KF}) < 2 * \tau_{maxD}) \wedge (\text{KF.Tdead } \neq 0)) \end{aligned} \tag{4.7}$$

Where *dist* is the corresponding weighted distance matrix between the blobs and the KFs, and $distPAM$ is the corresponding match probability of the PAM. τ_{minP} is the minimum match probability and τ_{maxD} is the maximum Euclidean distance between the new object and the KF.

If a "newblob" is more similar to a KF than the "oldblob", it has the shortest distance between the centroids and the major and minor axes, respectively, and also the best matching PAM, in comparison to the "oldblob". Hence, the "newblob" and the "oldblob" will be compared with the KF using the following equations:

$$\begin{aligned} &\text{if (dist(newblob,KF)} < \text{dist(oldblob,KF))} \\ \wedge\ &\text{(distPAM(newblob,KF)} > \text{distPAM(oldblob,KF))} \end{aligned} \tag{4.8}$$

Where the first statement determines if the new blob has the shortest distance between the centroids and the axes, respectively, and the second statement determines if the new blob has the highest PAM match probability.

4.4.3 Occlusion Handling using Probabilistic Appearance Model

Probabilistic appearance models inspired by [59] are applied to resolve inter-object occlusion. Each track has its own probabilistic appearance model, which consists of an RGB colour model with an associated probability mask. An example of a probabilistic appearance model is illustrated in figure 4.6. The colour model, which is denoted $M_{RGB}(\mathbf{x})$, shows the appearance of each pixel of an object. $P_c(\mathbf{x})$ denotes the probability mask and represents the probability of the object being observed at that pixel. The use of probabilistic appearance models can be viewed as weighted template matching, where the template is $M_{RGB}(\mathbf{x})$ and the weights are given by $P_c(\mathbf{x})$. The coordinates of $\mathbf{x}$ are expressed using the coordinate system of the model, which is normalized to the object centroid.

Depending on the data association between a track and the detected objects, one of three possible approached is applied as shown in figure 4.7. For each new track, a new probabilistic appearance model is created. In the object match situation, a track refinement step is applied before updating the model by finding the best fit in a small neighbourhood, e.g. 5×5 pixels. Track refinement increases the accuracy of the model; especially the colour model becomes sharper. When updating, the model usually stabilizes after less than 10 frames. Detail on building the model can be found in [59]. In the object merging situation probabilistic appearance models of the tracks are used to assign pixels of the detected object between the tracks using the flow in the bottom line of figure 4.7.

The foundation of the probabilistic appearance model is the ability to estimate the probability that a given pixel x of a detected object belongs to the model $\mathcal{M}_j$ of track j. This is denoted by $P(\mathcal{M}_j|I(\mathbf{x}))$. I is the colour input image and is assumed to be normalized to the centroid of the detected object. The probability is calculated

Algorithm 3 Data Association between blob and KF

- **while** the list of blobs is not empty, **then**:

  - Evaluate the current blob (newblob).
  - **if** there is a KF associated to this blob, **then**:
    - **if** the best KF for this blob is **not used**, **then**:
      - **if** the conditions 4.7 to match KF with the newblob are **valid**, **then**:
        - **Match KF-newblob**, KF Tstable.
      - **else**, KF is **invalid**, **then**:
        - **new KF**.
        - **Match newKF-newblob**.
    - **else** KF is **used**, **then**:
      - Get the blob associated to this KF (oldblob).
      - **if** the newblob is more similar and has a better PAM match than the oldblob, conditions 4.8, **then**:
        - **Match KF-newblob**, KF Tstable.
        - **Free KF-oldblob** and add oldblob to the list of blobs.
      - **else**, **then**:
        - Check **next** best KF for this object.
  - **else**, **no KF** is associated for this blob, **then**:
    - **new KF**.
    - **Match newKF-newblob**.

- **for** KF **not associated**, **then**:

  - **Lost KF**, KF Tdead.

using Bayes' rule:

$$P\left(\mathcal{M}_j|I\left(\mathbf{x}\right)\right) \propto P_{RGB,j}\left(I\left(\mathbf{x}\right)|\mathcal{M}_j\right) \cdot P_{c,j}\left(\mathbf{x}\right) \tag{4.9}$$

The a priori probability is given by the probability mask of model $\mathcal{M}_j$, $P_{c,j}\left(\mathbf{x}\right)$. $P_{RGB,j}\left(I\left(\mathbf{x}\right)|\mathcal{M}_j\right)$ is the color appearance likelihood, and this is approximated using a Gaussian color distribution:

$$P_{RGB,j}\left(I\left(\mathbf{x}\right)|\mathcal{M}_j\right) = \frac{1}{\left(2\pi\right)^{3/2}\left|\Sigma\right|^{1/2}} \cdot \tag{4.10}$$
$$\exp\left(-\frac{1}{2}\left(I\left(\mathbf{x}\right) - M_{RGB,j}\left(\mathbf{x}\right)\right)^T \Sigma^{-1}\left(I\left(\mathbf{x}\right) - M_{RGB,j}\left(\mathbf{x}\right)\right)\right)$$

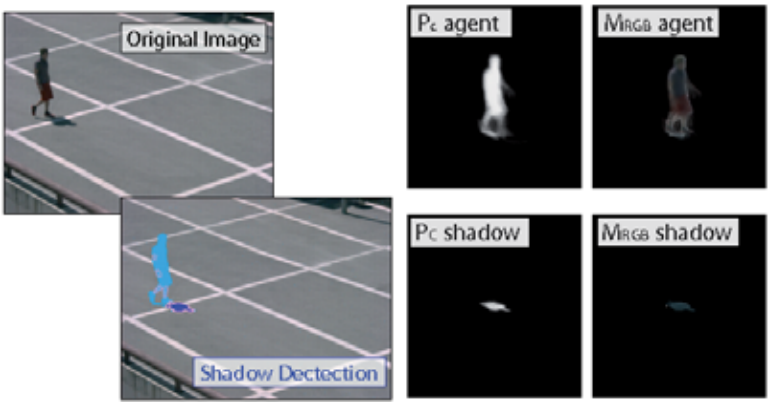


Figure 4.6: An example of a probabilistic appearance model, where the input Image, the shadow detection Image, the probability mask P_c and the color model M_{RGB} for the detected agent and shadow are shown.

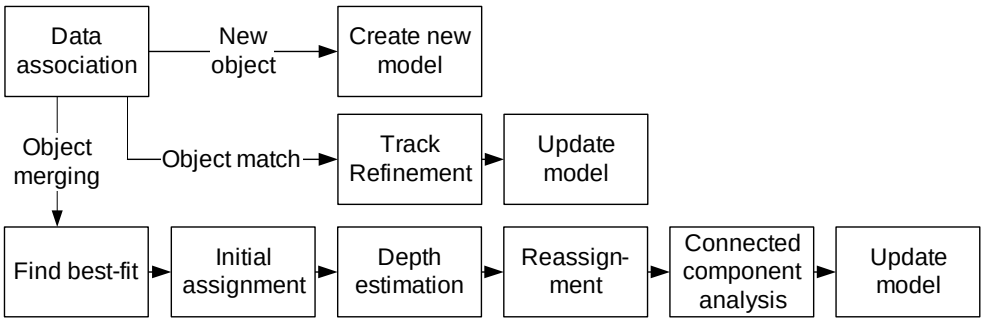
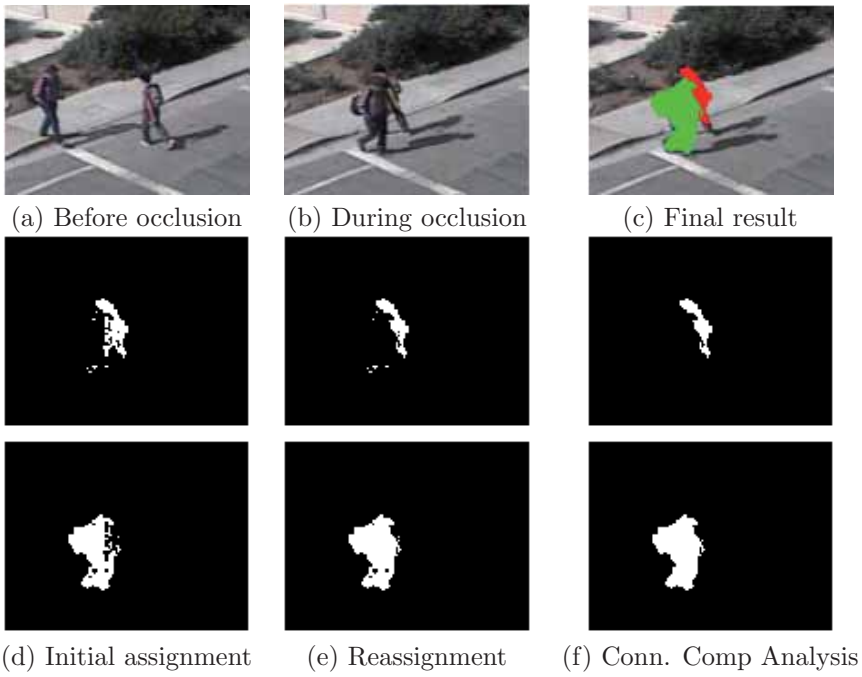


Figure 4.7: Flow related to the use of probabilistic appearance model.

The colour model for track j, $M_{RGB,j}$, represents the mean colour for each pixel. To reduce the complexity, the covariance matrix Σ can be assumed to be a diagonal matrix with identical variance σ in each colour channel. Given these assumptions, Equation 4.4.3 reduces to:

$$P_{RGB,j}\left(I\left(\mathbf{x}\right)|\mathcal{M}_j\right) = \left(2\pi\sigma^2\right)^{-3/2} \cdot \exp\left(\frac{-\left\|I\left(\mathbf{x}\right) - M_{RGB}\left(\mathbf{x}\right)\right\|^2}{2\sigma^2}\right) \tag{4.11}$$

where σ is selected empirically. The algorithm in 4 explains the procedure for segmenting objects under occlusion using probabilistic appearance models. An example is given in figure 4.8.



(a) Before occlusion (b) During occlusion (c) Final result

(d) Initial assignment (e) Reassignment (f) Conn. Comp Analysis

Figure 4.8: Example of occlusion handling using probabilistic appearance models. (d), (e) and (f) show the intermediate steps for resolving the occlusion in (b).

Algorithm 4 Occlusion Handling

1. The centroids are predicted for each track using a first order kalman filter, as described in appendix A
2. Tracks are fitted to the foreground pixels, to find the best-fit location. If depth order is available the foremost pixels are fitted first, and the pixels where this track's probability mask has non-zero probability are not used for fitting the tracks with greater depth.
3. Pixels with non-zero probability belonging to more than one track are identified as disputed pixels. Each disputed pixel is assigned to the track with the highest probability based on Equation 4.9. See Fig. 4.8.(d).
4. Tracks are ordered so that tracks assigned fewer disputed pixels are given greater depth, and all disputed pixels are reassigned to the foremost track that overlaps the pixels. See Fig. 4.8.(e).
5. Connected component analysis is performed to clean up the segmentation, as seen in Fig. 4.8.(f).

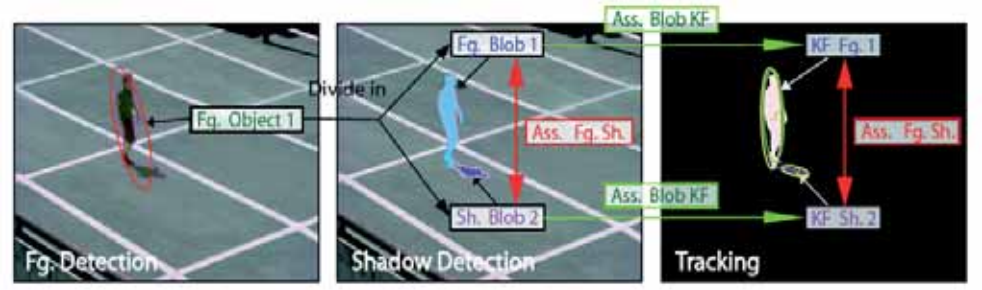


Figure 4.9: An example of data association between FG and SH and the assigned KFs. First image represents the FG detection provided in chapter 3. Second image shows the shadow detection presented in sec. 4.4.2, and how the analysed FG is divided into FG and SH blobs. These blobs are associated because both come from the same FG object. In the third image the tracking system has assigned one KF to each blob, then the data association info between the FG and the SH blobs is added to the KF info for each blob.

4.4.4 Update FG-SH association in KF info

After the blobs (belonging to a FG or a SH) has been assigned to the KF, as described in sec. 4.4.2, the association between which shadow belongs to which FG and vice versa is saved in the KF info for use in the next frames. This info will be used to identify the possible cases in the association between FG and SH. An example showing the data association between the blobs and the Kalman Filters, and how the data association between FG and SH is later saved in the KF info, can be seen in Fig. 4.9.

The first image of Fig. 4.9 represents the FG detection provided in chapter 3. The second image shows the shadow detection presented in sec. 4.4.2, and how the FG segmentation is further analysed and divided into FG and SH blobs. These blobs are associated because both come from the same FG object. In the third image the tracking system has assigned one KF to each blob, and the data association info between the FG and the SH blobs is added to the KF info for each blob.

4.4.5 Temporal consistency in the Data Association

The information related to the association between FG and SH saved in the KF has to be analysed in order to check the possible cases, e.g., if a shadow has been lost. In Fig. 4.4 it can be seen how the approach works in the case of the detection of a lost shadow. The images showed in the figure are explained further in the following subsections.

Cases

When performing temporal consistency in the data association between the FG and SH with their respective KF, three situations can occur:

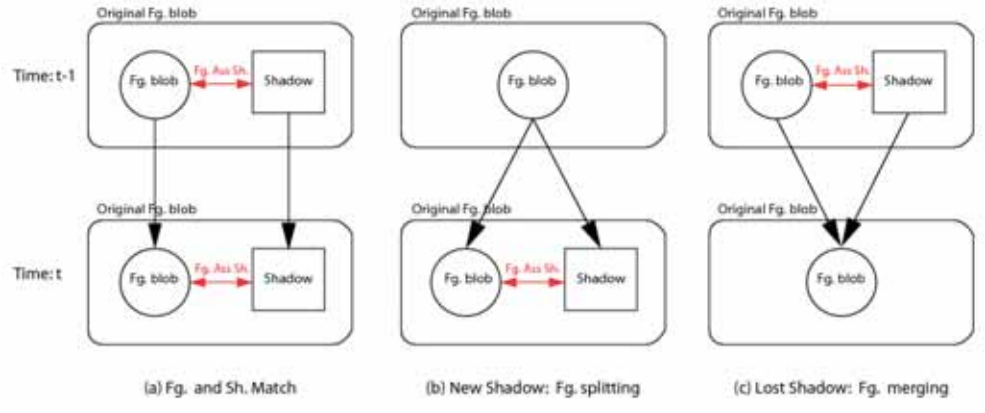


Figure 4.10: The three posible data association situations between FG and SH and their KFs. A rounded rectangle illustrates an original FG blob before shadow detection, a circle illustrates a FG and a square illustrates a SH from the shadow detection. A double red arrow indicates an association between FG and SH And a black arrow indicates an association between FG and SH in the next frame.

- FG and SH match:
  The association created at time t-1 continues at time t, which is the ideal case.

- New shadow: object splitting in FG/SH.
  A new association between the FG and SH is created at time t.

  This can be a problem if the shadow is erroneously detected because it can also be falsely detected as lost shadow in the next frame. If it is a new shadow, it is not Tstable, and it is not detected as shadow over a time period (Tdead), it will not be considered as lost shadow in the next frame, and later it will be discarded as shadow. Problems exist when objects move quickly combined with low frame rate, since in one frame the shadow is correctly detected, in the next frame it is not detected, and in the following frames it is lost.

- Lost shadow: object merging with shadow.
  The association between the FG and SH at time t-1 has been lost at time t because the shadow is lost. See Fig. 4.4for an example of this case.

The three possible cases are illustrated in the Fig. 4.10. It is possible that a new shadows appears or a shadow is lost without a splitting or merging in the FG object. However, these cases are not of interest because they do not have any data association, and they will be tracked in an usual manner by the Kalman filters.

Lost Shadow: object merging with shadow

A shadow is considered lost when the blob (the KF that is associated with this blob) fulfil a set of conditions: it was classified as SH at time t-1 (the previous frame), and

it had a FG associated, the FG also had this SH associated. At time t (the current frame), this FG has no shadow associated and the SH has also lost the association with this FG, then this shadow is considered lost.

The shadow region can be recovered by evaluating the FG blob (which contains the FG and the shadow), the blob prediction for the FG KF, and the blob prediction for the lost SH KF.

Possible regions from FG object

The FG blob which belongs to the FG KF associated in the previous frame with the shadow considered lost, is analysed in order to find the possible shadow region. Therefore, the mask of the positives edges (ed_pos mask) plus morphological operators are applied for the FG blob to divide it into FGs and possible shadow regions. Multiple regions can be found but theoretically only one is the shadow. This happens because the positive edges are used to divide the image, and these edges come from the current image. Such as explained in sec. 4.2, one of the characteristics of shadows is that they can only have negative edges, e.i., the edges from the background image. Therefore, theoretically several FG regions can be found but only one SH region. In Fig. 4.4 it can be seen how the original FG blob detected, as described in section 4.3, is subdivided into the possible chromatic shadow regions (image Chr.Sh.Regions in the figure; the regions in the image are shown in different colours) using the ed_pos mask. These blobs from the possible regions will later be associated with the predictions of the Kalman Filters in order to detect the chromatic shadow.

Correspondence matrix between the new blobs and the KF predictions

The weights for the blob prediction for the FG KF and the SH KF are calculated w.r.t. all the possible regions found in the previous step. Therefore, two correspondence matrix are calculated, where one contains the euclidean distance between the new blobs and the FG and SH KF predictions, and the other the overlapping (matching) between the new blobs and the FG and SH KF predictions. These weights are used to associate the SH and FG KF predictions with the blobs.

Association between FG and SH KF predictions and the blobs

The best match (shortest distance and best overlap) between the SH KF predictions and the blob will be considered as the shadow region, while the other blobs will be considered as FG blobs, since only one region can be shadow. Hence, the other blobs have to be FGs.

In this way, by using the tracking information, the original FG blob can be segmented into FG and SH regions, thereby detecting additional chromatic shadows, which are not detected in the previous section 4.3. This information is used as a feedback from the tracking to the shadow detection step.

In Fig. 4.4 it is shown how the blobs extracted from the possible regions are associated with the prediction of the Kalman Filters in order to detect the chromatic shadow.

4.4.6 Feedback from tracking to the original image

Once the chromatic shadows are detected, the original image (original FG blob) is divided into only one blob for the FGs and only one blob for the SHs. Hence, the FG blob will be associated with the FG KF and the SH blob will be associated with the lost SH KF.

However, the association can result in multiple FG blobs. In order to get only one blob for the FG without splitting the original FG blob into multiple blobs, we only have to use the positive edges calculated in the previous frame to divide the shadow blob in the original blob image.

Once the original blob is divided into two blobs, one corresponding to a FG and the other to a shadow, the original image has to be updated so the chromatic shadows, which were not detected before, are now marked as detected.

In Fig. 4.4 it is shown how the detected chromatic shadow is correctly updated according to the original blob, after the feedback from the association between the tracks and the possible regions from the original FG blob.

4.4.7 Manage and Update KF info and PAM

The info related to the new association between the new FG and SH blob and their respective KFs have to be updated. The KFs also have to be updated with the new associated blobs, and the PAMs have to be updated considering the new blobs.

It is possible that a new KF was erroneously created because one object together with its shadow can be considered as a new object. Therefore, the new KFs created in the data association between the blobs and the KFs have to be checked. If there is a KF which was created but not used, since it was assigned to the blob currently considered as lost shadow, then it is deleted.

Consequently, thanks to the data association between FG and SH we have achieved: (i) enhancing the chromatic shadow detection by detecting shadows which were not possible to detect before. (ii) Improving the segmentation for high processes, such as detection and tracking, by avoiding shadows. (iii) A more robust tracking, since (1) erroneously created KFs are deleted and (2) the PAM and the KF tracker are more robust and correctly updated. In Fig. 4.4 an example is given, where the output of the tracking process is shown with and without the Top-down approach (see the last two figures called Tracker End). The figure of the Top-down approach shows how the system is correctly detecting the chromatic shadow, therefore the FG KF and SH KF are also correctly updated. On the other hand, in the image from the tracker without taking into account the association between the FG and the SH and their assigned KFs, it can be seen how the FG and SH KF are lost but also a new false KF is created.

4.5 Experimental Results

The results presented in this section are all from tests conducted on datasets selected from well-known databases. Our approach is tested on sequences of outdoor and indoor scenarios, and compared to other statistical approaches when results are avail-

Figure 4.11: An original image from the Outdoor_Cam1 sequence, and foreground results after shadow removal using the Huerta et al. approach [25], the Zivkovic et al. approach [78] using a shadow detector [51], and our approach, respectively (read row-wise).

able. The chosen test sequences are relatively long and umbra and penumbra shadows are cast by multiple foreground objects. The sequences analysed are Outdoor_Cam1 (800 frames, 607x387 PX), HigwayIII (2227 frames, 320x240 PX), HallwayI[1] (1800 frames, 320x240 PX), and HERMES_ETSEdoor_day21_I4 (6500 frames, 640x480 PX).

Figure 4.11, 4.12, and 4.13 show the results when comparing our shadow detector with other approaches from the state-of-the-art [25, 33, 78, 51, 41]. As it can be seen in these figures our approach outperforms the other analysed methods. However, in a few cases the gradient masks cannot be accurately build due to camouflage and noise problems. Thus, the separation of a foreground object and a shadow region can fail. Occasionally, when the anomaly in case 2-2 (see sec. 4.2.3) occurs and a part of the foreground object or the shadow is not segmented due to segmentation problems, the shadow position can miss-classify the shadow as a foreground object. The top-down approach can solve part of this problems, as it can be seen in figure 4.19 and 4.18.

To evaluate our approach in a quantitative way, it has been compared with the approaches [40, 41] using the most employed quantitative expressions utilized to evaluate the shadow detection performance: the Shadow Detection Rate (SR) and the Shadow Discriminate Rate (DR). Refer to [51] for the exact equations. The results in the table 4.1 shows that our method outperforms both the parametric approach based on Gaussian mixtures GMSM [40] and the nonparametric physical model [41]. Note that the results for the GMSM [40] and the physical model [41] on the Hallway sequence have been obtained directly from [41].

It should be noted that our approach needs a reasonable resolution to work cor-

[1] http://vision.gel.ulaval.ca/ CastShadows/

Figure 4.12: An original image from the HallwayI sequence, and foreground results after shadow removal using the Huerta et al. approach [25], the Kim et al. approach [33], the Zivkovic et al. approach [78] using a shadow detector [51], the Martel et al. approach [41], and our approach, respectively (read row-wise).

Method	HallwayI	
	SR	SD
GMSM	0.605	0.870
Physical model	0.724	0.867
Our approach	0.807	0.907

Table 4.1
SR and SD results for our approach and two of the most successful methods: Gaussian Mixture Shadow Models (GMSM) [40] and a physical model of light and surfaces [41].

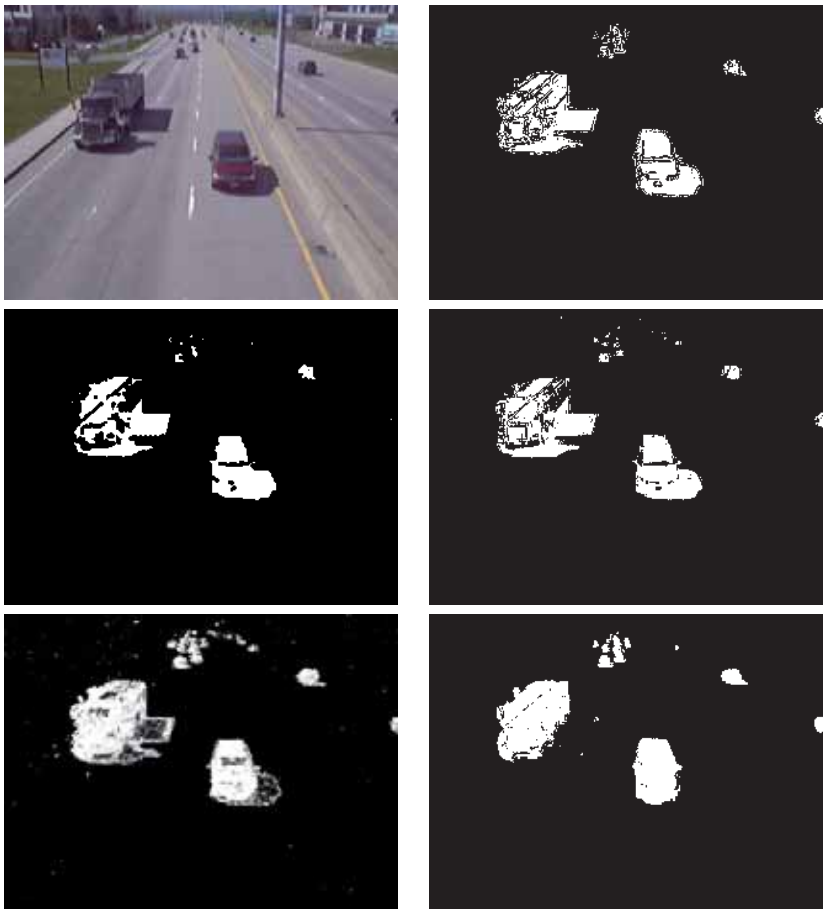


Figure 4.13: An original image from the HighwayIII sequence, and foreground results after shadow removal using the Huerta et al. approach [25], the Kim et al. approach [33], the Zivkovic et al. approach [78] using a shadow detector [51], the Martel et al. approach [41], and our approach, respectively (read row-wise).

rectly. Futhermore, shadow regions need to have a minimum area for analysis or there might not be enough information for a proper shadow detection and classification.

A number of processed frames of significance are depicted in Fig. 4.14, showing the shadow detection results using our final approach for the Outdoor_Cam1 sequence from the CVC database. This sequence contains most of the problems commented in the case analysis, heavy background motion due to waving trees, chromatic shadows, saturations and all types of camouflages. The approach described in the previous chapter achieved the detection of the agents, however the shadows were also detected as parts of the agents because of their change in chroma. However, in the figure it can be observed that our shadow detector can detect the chromatic shadows in this sequence.

Fig. 4.15 shows processed frames of significance, depicting the shadow detection results using our final approach for the HallwayI sequence from the LVSN database. This sequence contains moving cast shadows of the agents on the floor and the walls, which have a change in the chroma. Our last approach was able to detect all the agents in the sequence, however most of the shadows were also detected as parts of the objects, due to this change in the chroma. As it can be seen in the figure, our shadow detector detects most of the shadows. However, the smaller of them are not detected, since our approach need a minimum area to process the applied statistics.

Another set of processed frames are depicted in Fig. 4.16, showing the shadow detection results using our final approach for the HighwayIII sequence from the LVSN database. The shadow detector can cope with the chromatic shadows of the vehicles such as it can be seen in the figure. However, the detection can sometimes fail because the regions are not well defined. This happens when a part of a vehicle is considered as a shadow. Due to camouflage problems, it fails to divide the vehicle region from the shadow region, see Fig. 4.16.(c). The shadow detection can also fail when a region does not have any local gradient structure, and all the pixels in this region exhibit similar chromaticity and brightness. For example the front car in Fig. 4.16.(f).

Fig. 4.17 shows processed frames of significance, depicting the shadow detection results using our final approach for the HERMES_ETSEdoor_day21sequence from the HERMES database. This outdoor surveillance sequence is a long sequence with several agents walking through the scene. The shadow detector can cope with the chromatic shadows of the agents, as it can be seen in the figure. However, the detection can sometimes fail. This can happen when regions are not well defined. For instance, in depicted case it happens when the shadow is not correctly split up from the agent due to camouflage problems, see Fig. 4.17.(e).

The top-down process assists the chromatic shadow detector when it fails to detect shadows, as shown in figure 4.16.(c) and 4.17.(e). The tracking system is able to track the shadows and use this information as feedback to the chromatic shadow detector. Hence, the miss-detected shadows can be recovered and correctly detected. Figures 4.18 and 4.19 show the results of the top-down approach. The figures are an example of shadow recovery using our top-down approach for the HERMES_ETSEdoor_day21 sequence and the LVSN_HighwayIII sequence.

Fig.4.19.(a) shows the fg. detection results obtained in the previous chapter. In Fig.4.19.(b) the chromatic shadow detection results of our detector are shown. Note that the shadows are not correctly detected. Fig.4.19.(c) shows the output of the

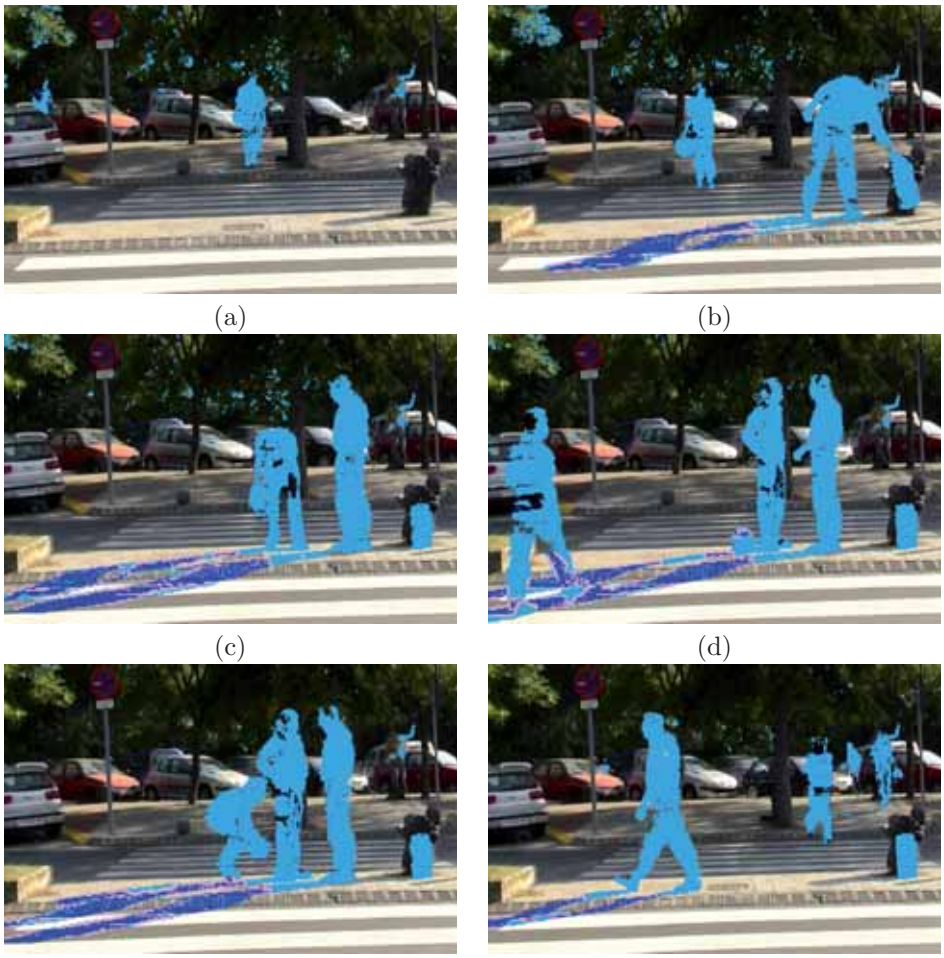


Figure 4.14: Chromatic Shadow detection results for the CVC_Outdoor_Cam1 sequence using our shadow detector. The sequence contains several shadows that exhibit a change in their chroma. They are detected using our approach, thereby showing that it is able to tackle these shadow problems. Previous segmenation results are coloured in cyan, and the shadow detection results are coloured in blue.

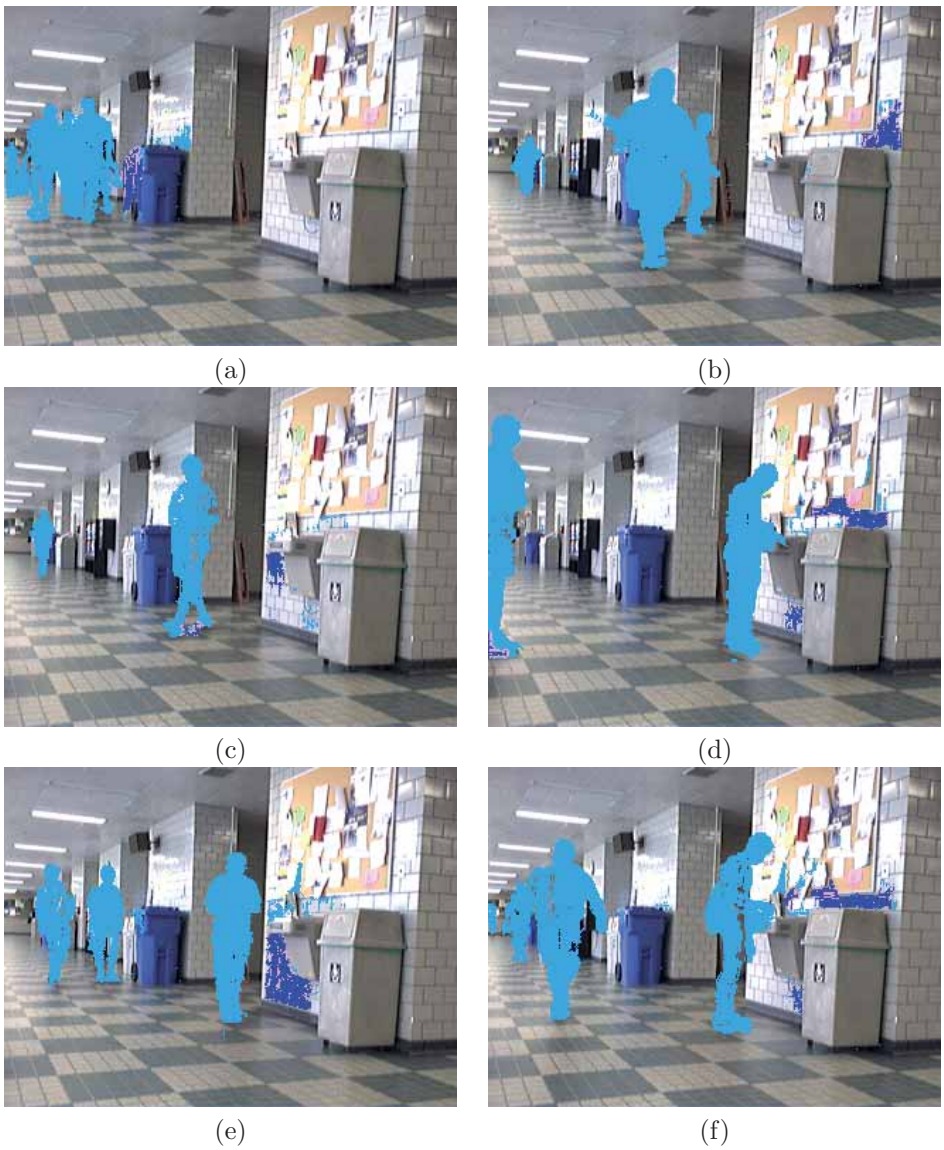


Figure 4.15: Chromatic Shadow detection results for the LVSN_HallwayI sequence using our shadow detector. Our shadow detector is able to detect the chromatic shadows cast by the agents on the floor and the walls. Previous segmenation results are coloured in cyan, and the shadow detection results are coloured in blue.

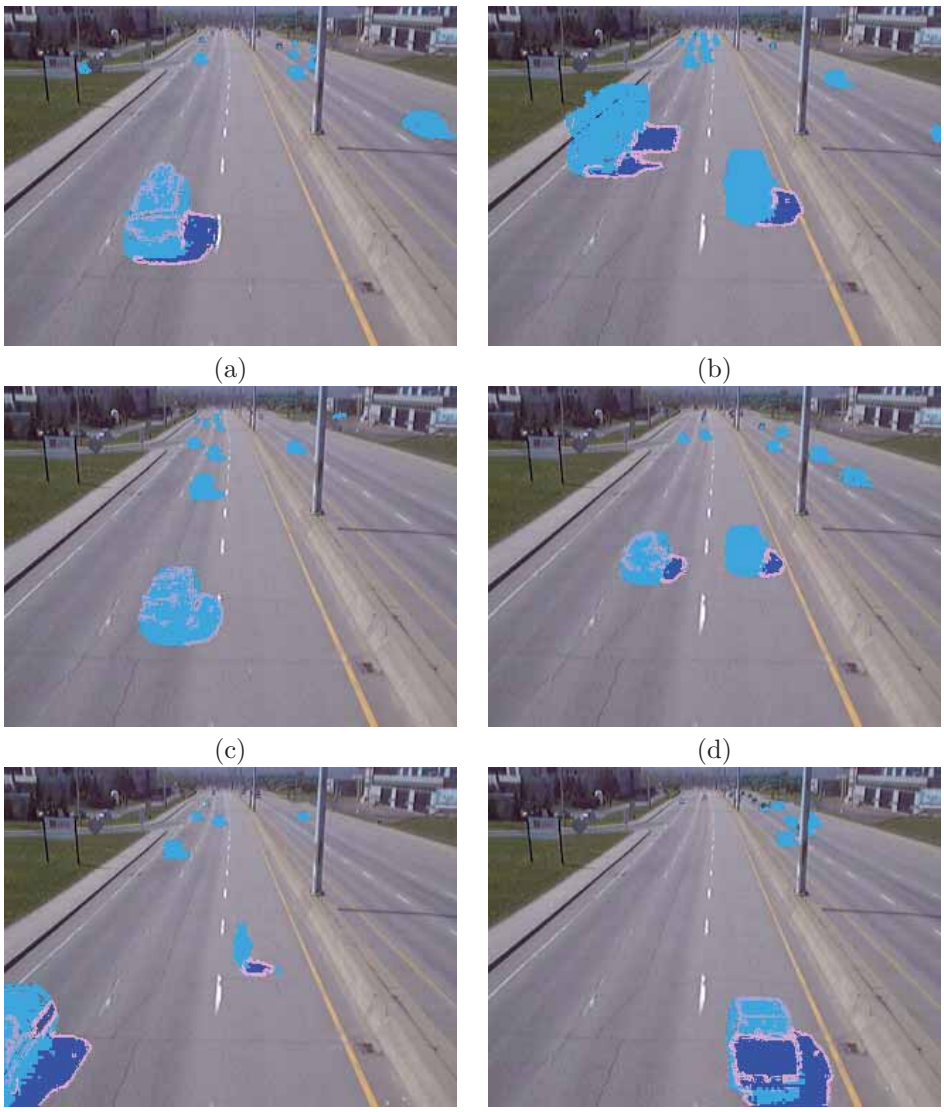


Figure 4.16: Chromatic Shadow detection results for the LVSN_HighwayIII sequence using our shadow detector. Our shadow detector is able to accurately detect most of the larger chromatic shadows of the vehicles. However, it can fail due to camouflage problems dividing the shadow regions, see (c). Previous segmenation results are coloured in cyan, and the shadow detection results are coloured in blue. See the main body text for further details.

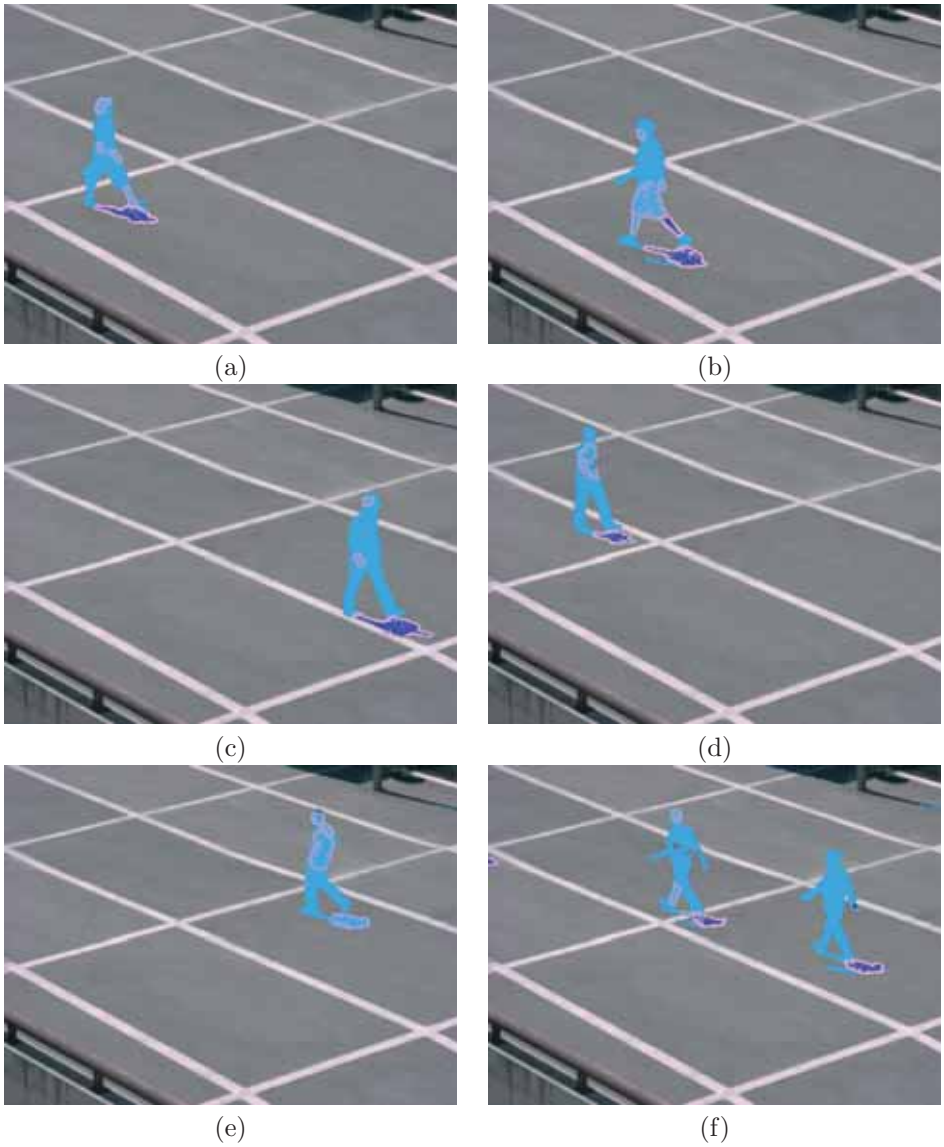
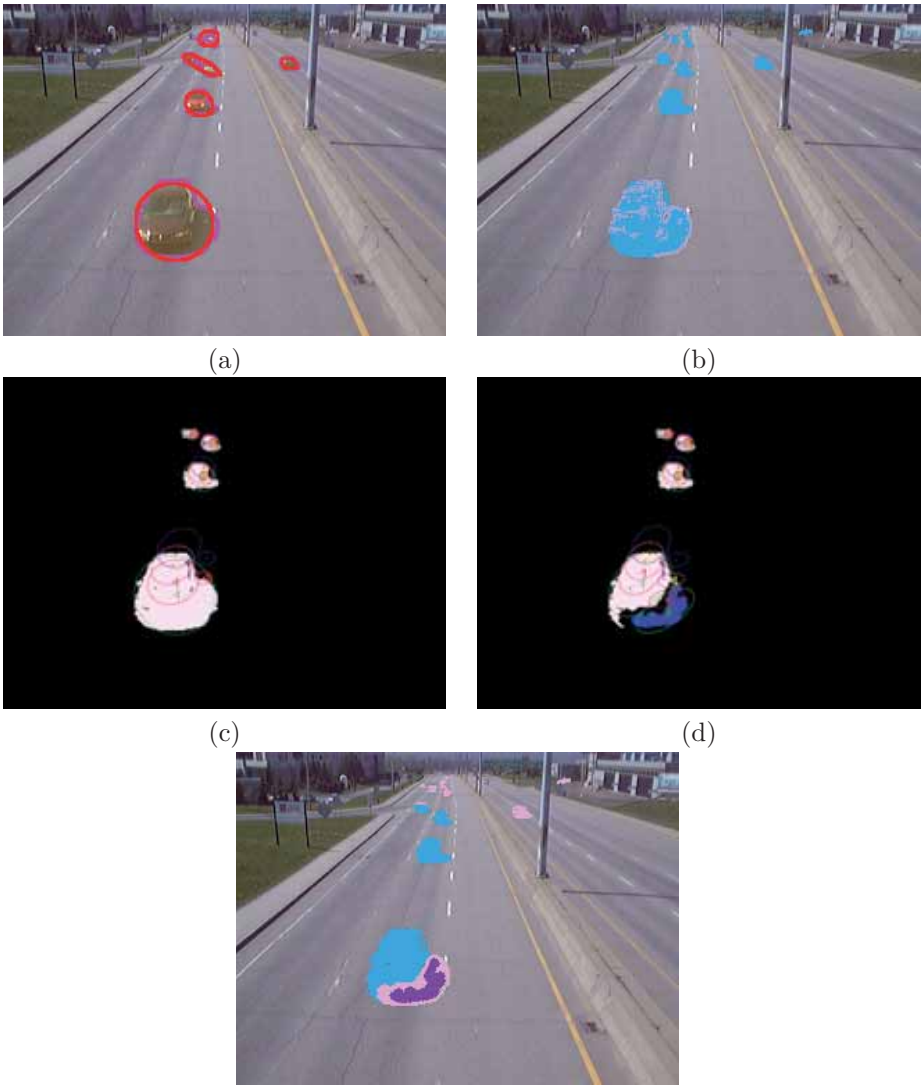


Figure 4.17: Chromatic Shadow detection results for the HERMES_ETSEdoor_day21 sequence using our shadow detector. All the shadows in te sequence exhibit a change in their chroma. Our shadow detector is able to accurately detect most of them. However, it can fail sometimes, such as it can be seen in (b) and (e). In (b) a part of the leg is erroneous detected as shadow, and in (e) the shadow is not detected due to camouflage problems. Previous segmenation results are coloured in cyan, and the shadow detection results are coloured in blue. See the main body text for further details.



(a) (b) (c) (d) (e) (f)

Figure 4.18: An example of shadow recovery using our top-down approach for the LVSN_HighwayIII sequence. Image (a) is the fg. detection image. (b) is the chromatic shadow detection results of our detector. Note that the shadow is not correctly detected. (c) is the output of the tracker without applying our top-down approach. The KF representing the shadow is lost and therefore the KF will be falsely updated. (d) shows results of our top-down approach: the output of the tracker, after the chromatic shadow is recovered, using our top-down process. In this image the shadow is accurately detected, the FG KF and the SH KF are correctly updated, and none of them are lost. The a posteriori state of the tracker is depicted with a red ellipse. Image (e) shows the final chromatic shadow detection results in the original image. See the main body text for further details.

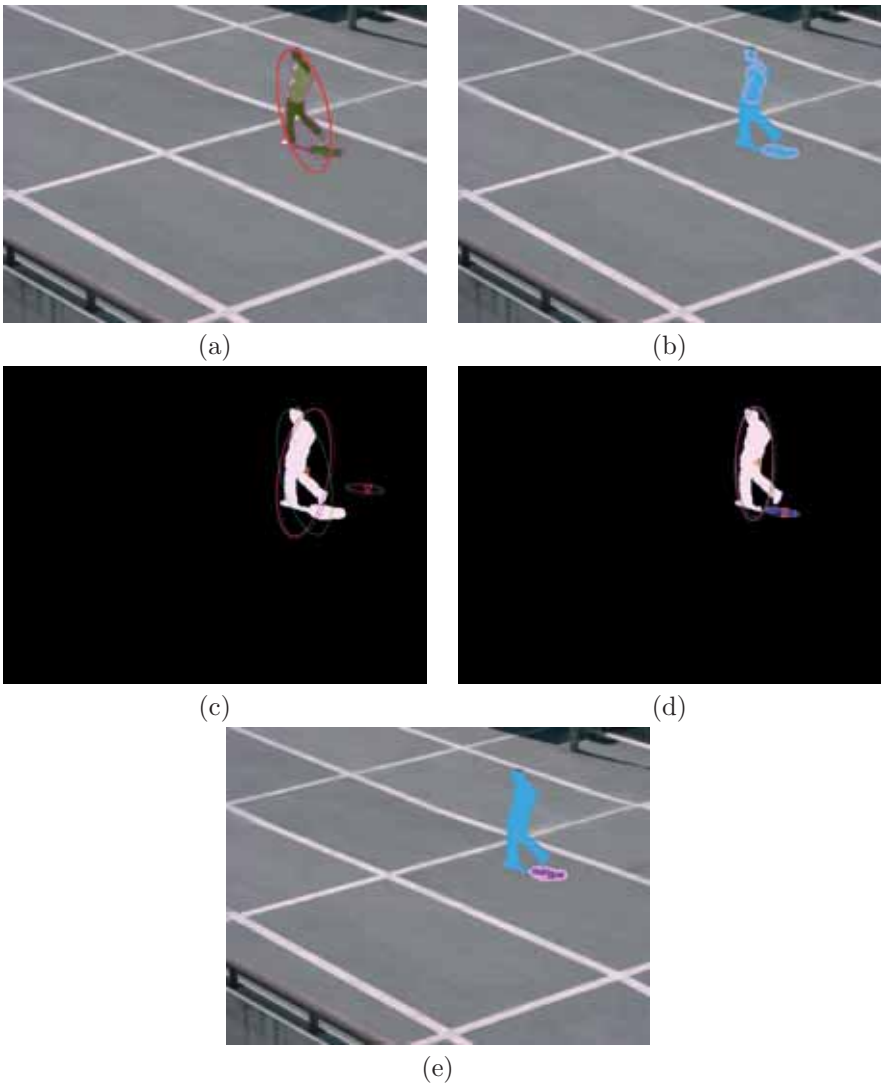


Figure 4.19: En example of shadow recovery using our top-down approach for the HERMES_ETSEdoor_day21_I4 sequence. Image (a) is the fg. detection image. (b) is the chromatic shadow detection results of our detector. Note that the shadow is not correctly detected. (c) is the output of the tracker without using the top-down approach. The shown image is after 10 frames without a detected shadow, therefore the KF associated to the shadow is lost (Tdead) and falsely updated. (d) shows results of our top-down approach: the output of the tracker, after the chromatic shadow is recovered, using our top-down process. In this image the shadow is accurately detected, the FG KF and the SH KF are also correctly updated, and non of them are lost. The a posteriori state of the tracker is depicted with a red ellipse. Image (e) shows the final chromatic shadow detection results in the original image. See the main body text for further details.

tracker without applying our top-down approach, while Fig.4.19.(d) shows the results of our top-down approach. Finally, Fig.4.19.(e) shows how the chromatic shadow has been accurately detected after the feedback from the tracker to the chromatic shadow detector. The input image shown in this figure is captured without a detected shadow for 10 frames. Hence, this example illustrates how the tracker's KF assigned to the shadow is completely lost without the top-down approach (Fig.4.19.(c)), while the other tracker's KF is tracking the combined fg. and sh. blob (a Red ellipse depicts the a posteriori state of the KF). In some cases the tracker will create a new KF, since the combined fg. and sh. blob is so different that the system thinks it is a new object. In contrast, Fig. 4.19.(d) shows the output of the tracker using our top-down process. In this image the shadow is accurately detected, and the KFs are correctly updated. This is illustrated by the red ellipses in the image.

Fig. 4.18 shows similar results in comparison to Fig. 4.18 but using the LVSN_-HighwayIII Sequence. Our top-down approach achieves to detect the chromatic shadows, as it can seen in figure 4.18.(d) and 4.18.(e). However, this scenario is very difficult to track, since the fg. blobs move very fast compared to the frame rate of the sequence. Additionally, the appearance of the objects changes very quickly. In this case the tracks are sometimes lost, and therefore it is not possible to run the top-down process throughout all of the sequence.

Fig. 4.20 shows a number of processed frames of significance, depicting the results using our top-down approach. In the figure it can be seen how our approach is able to track the objects and the shadows, and when the chromatic shadow is lost, the system is able to recover it. In this way the feedback from the tracking to the segmentation process is assisting the chromatic shadow detector, and thereby achieves to detect shadows which were miss-detected before. Thus, by updating the KFs in order to get an accurate a posterior state for the image, the segmentation process take advantage of the recovered information about objects and shadows through the feedback from the tracking. This is shown in the second and third column of Fig. 4.20, where the tracking results using our top-down approach (third column) achieves a correct tracking compared to the falsified tracking results, when it is not applied (second column).

4.6 Discussion

In this chapter, we have presented two main novelties: (i) a bottom-up approach for detection and removal of chromatic moving shadows in surveillance scenarios [26], and (ii) a top-down approach based on Kalman filters to detect and track shadows.

In the Bottom-up part the shadow detection approach apply a novel technique based on gradient and colour models for separating chromatic moving shadows from moving objects.

Firstly, we extend and improve well-known colour and gradient models into an invariant colour cone model and an invariant gradient model, respectively, to perform automatic segmentation while detecting potential shadows. Hereafter, the regions corresponding to potential shadows are grouped by considering "a bluish effect" and an edge partitioning. Lastly, (i) temporal similarities between local gradient structures

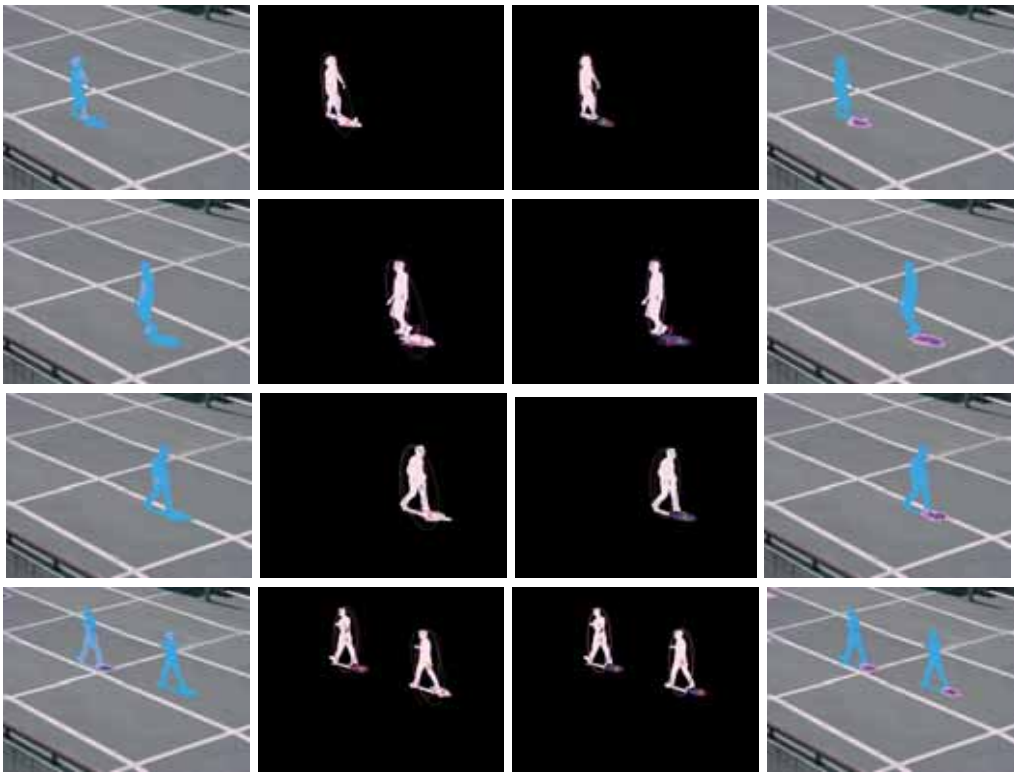


Figure 4.20: Additional chromatic Shadows detection results using our top-down approach for the HERMES_ETSEdoor_day21_I4 sequence. First column shows the chromatic shadow detection results. Note that the shadows are not correctly detected. Second column shows the output of the tracker without the association between the FG-SH, where the tracks for the shadows are lost. Third column shows the tracker output using our top-down approach. The chromatic shadows are detected and the tracker are correctly updated for the FG and the SH. The a posteriori state of the tracker is depicted with a red ellipse. The last column shows how the chromatic shadow is recoverd and correctly detected in the original image.

and (ii) spatial similarities between chrominance angle and brightness distortions are analysed for all potential shadow regions, in order to finally identify umbra shadows.

The resulting shadow detection can (1) detect and remove chromatic moving shadows (umbra shadows) and (2) penumbra shadows, while several other methods are restricted to the latter.

However, in some cases the separation between a foreground object and a shadow region can fail. Occasionally, a part of the foreground object or the shadow is not accurately segmented due to segmentation problems, e.g. camouflage. Therefore, the shadow detection can miss-classify a shadow as being a part of a foreground object.

In order to solve this problem a top-down approach has been developed, as outlined in this chapter. After detection of the objects and shadows both are tracked using

Kalman filters, in order to enhance the chromatic shadow detection, when it fails to detect a shadow. Firstly, a data association between the blobs (FG and SH blobs) and the Kalman filters is performed. Secondly, an event analysis is carried out in order to detect the different cases: object match, new object, lost object, object splitting and object merging. Taking this information into account, the Kalman filters are managed. Furthermore, occlusion handling is managed based on a Probabilistic Based Model (PAM). Temporal consistency is evaluated in the association between FGs and SHs and their respective Kalman Filters over time. Consequently, a number of cases are studied: FG and SH match, new shadow and lost shadow. Finally, the tracking results are feedback to the chromatic shadow detector to improve the object and shadow detection. Thus, chromatic shadows are correctly detected in cases with the mentioned segmentation problems.

Consequently, thanks to the data association between FG and SH we have achieved: (i) enhancement of the chromatic shadow detection by detecting shadows which were not possible to detect before. (ii) improvement the segmentation for high level processes, such as detection and tracking, by avoiding shadows. (iii) a more robust tracking, since (1) the PAM and the KF tracker are more robust and correctly updated, and (2) erroneous created KFs are deleted.

Qualitative and quantitative results of tests for both outdoor and indoor sequences from well-known databases validate the presented approach. Overall, our approach gives a more robust and accurate shadow detection and foreground segmentation compared to the state-of-the-art methods.

Unlike other approaches, our method does not make any a-priori assumptions about camera location, surface geometries, surface textures, shapes and types of shadows, objects, and background. Experimental results show the performance and accuracy of our approach in different shadowed materials and illumination conditions.

However, some remarks have to be said with respect to the bottom-up part (chromatic shadow detector) and the top-down part (shadow tracking). The chromatic shadow detector needs a reasonable resolution to work correctly, and noisy and blurred images intensify the camouflage problems. Furthermore, shadow regions need to have a minimum area for analysis, or there might not be enough information for a proper shadow detection and classification. The "bluish effect" gives very good results for some outdoor sequences. However, sometimes it does not work as defined theoretically, since it is affected by external factors, such as the sensibility of the camera and image compression. For the tracking process, targets are assumed to move with a reasonable velocity compared to the frame rate. Since, objects which move quickly and change their appearance suddenly are difficult to track.

In future work, edge-linking or B-spline techniques can be used to avoid the partial loss of foreground borders due to camouflage, and thereby improve the edge model. Thus, avoiding shadows miss-classified as foreground objects, when shadows and objects are not split correctly.

Another interesting aspect is applying the direction of penumbra to umbra for a cast shadow to discriminate between foregrounds and shadows, when the image region does not have gradient nor similar temporal and spatial chrominance angle and brightness distortions. Additionally, fg. detection could be improved using a combination of a physical colour model and a camera model. Furthermore, a probabilistic

scheme based on the feedback from the tracking shadows could be taking into account in the shadow detection process.

In spite of the main subject of this thesis is motion segmentation and shadow detection, for future work the tracking system has to be improved, in order to be able to test the approach on more complex scenario, such as crowded scenes or situations with multiple grouping and splitting processes. The proposed tracking system performs well for basic scene situations. However, for more complex scenario the tracks are sometimes lost, and the tracker information becomes corrupted. Thus, in future work a high level tracker is needed, which can manage the low level trackers in a top-down architecture.

Furthermore, detection of gradient changes can be applied in the occlusion handling to improve the splitting process of occluded objects. Moreover, the target representation can be refined by including structure components and shape cues. E.g., body-part histograms and salient points could enhance agent tracking during long-term partial occlusions, while SIFT descriptors could provide new ways of target discrimination.

On the other hand, more complex cases of the temporal consistency of the data association between FG and SH with their respectively assigned Kalman Filter have to be studied. E.g., a possible shadow without any associated FG, because it has been split up before or the positive edge mask (ed_pos mask) has divided it erroneously. In order to solve these cases a high level event analysis has to be applied.

In order to solve the problem with fast moving objects relatively to the frame rate, with sudden changes in their appearance, in future work the momentum can be applied for the tracker information. In this way the tracker could improve the accuracy through the historical map of the previous trackers.

Finally, high-level information from the Human Sequence Evaluation structure, see appendix B for more information, can be applied to enhance the detection process in subsequent frames.

Chapter 5

Concluding Remarks

This Thesis is mainly divided in two parts. In the first one, a study of motion segmentation problems is firstly presented. Based on this study, a novel algorithm for mobile-object segmentation from a static background scene is developed. This approach is demonstrated robust and accurate under most of the common problems in motion segmentation. The second one tackles the problem of shadows in depth. Firstly, a bottom-up approach based on a chromatic shadow detector is presented to deal with umbra shadows. Secondly, a top-down approach based on a tracking system has been developed in order to enhance the chromatic shadow detection.

5.1 Discussion and contributions

In the first part, a case analysis of motion segmentation has been presented by taking into account the problems associated with the use of different cues such as colour, edge and intensity. This has allowed us to define when to use each model. Then, based on this case analysis, different motion segmentation problems have been solved.

The approach presented in this first part of the Thesis combines colour, intensity and edge cues, and a temporal differencing technique in a collaborative architecture, in which each model is devoted to a specific task. The background model of each cue has been improved with respect to the current state of the art. A chromatic invariant cone model is used as colour model, and an invariant gradient orientation combined with their magnitudes is used as edge model, which can avoid false edges due to intense global illumination changes. These are performed by a particular algorithm, but they can be substituted by enhanced ones without modifying the architecture itself. Hence, this structured framework combines in a principal way the main advantage of each cue. In this way, by taking advantage of several cues, the system is allowed to benefit from all the cues' capabilities.

The proposed hybrid approach can cope with different colour problems as (i) dark and light foreground. Furthermore, it solves problems with (ii) the dynamic range (problems associated with saturation and lack of colour problems) using intensity cues. The approach also tackles (iii) camouflage in intensity and (iv) camouflage in chroma, (v) avoiding the global and local (shadows and highlights) illumination

problems. Therefore, it can simultaneously differentiate these camouflages from the illumination changes. In addition, the approach can cope (vi) with bootstrapping and (vii) ghosts problems. But also, it can (viii) reduce the false positives using each cue independently. Therefore, our hybrid approach reduces the number of false negatives and false positives, and increases the detection rate.

Experiments on complex indoor and outdoor scenarios have yielded robust and accurate results, thereby demonstrating the ability of our system to deal with unconstrained and dynamic scenes. Therefore, our approach can work in indoor, outdoor scenes, with high or low resolution, with noise and blurred images, and no need calibrated images. Furthermore, it is also independent on the illumination and the illuminant on the scene. Moreover, size, appearance, number, and velocity of the objects are not important for our motion segmentation approach. This is because it does not make any a-priori assumptions about camera location, surface geometries, surface textures, shape and types of the objects or the background.

Some remarks have to be considered, although it is not needed any calibration for the camera, and no matter where it is situated, or the quality of the images from it. In order to use our motion segmentation approach the camera has to be a static camera.

In the second part of the Thesis, we have presented two main novelties: (i) a bottom-up approach for detection and removal of chromatic moving shadows in surveillance scenarios, and (ii) a top-down approach based on Kalman filters to detect and track shadows.

In the Bottom-up part, the shadow detection approach applies a novel technique based on gradient and colour models for separating chromatic moving shadows from moving objects.

Firstly, we extend and improve well-known colour and gradient models into an invariant colour cone model and an invariant gradient model, respectively, to perform automatic segmentation while detecting potential shadows. Hereafter, the regions corresponding to potential shadows are grouped by considering "a bluish effect" and an edge partitioning. Lastly, (i) temporal similarities between local gradient structures and (ii) spatial similarities between chrominance angle and brightness distortions are analysed for all potential shadow regions, in order to finally identify umbra shadows.

The resulting shadow detection can (1) detect and remove chromatic moving shadows (umbra shadows) and (2) penumbra shadows, while several other methods are restricted to the latter.

However, in some cases the separation between a foreground object and a shadow region can fail. Occasionally, a part of the foreground object or the shadow is not accurately segmented due to segmentation problems, e.g. camouflage. Therefore, the shadow detection can miss-classify a shadow as being a part of a foreground object.

In order to solve this problem a top-down approach has been developed, as outlined in this chapter. After detection of the objects and shadows both are tracked using Kalman filters, in order to enhance the chromatic shadow detection, when it fails to detect a shadow. Firstly, a data association between the blobs (FG and SH blobs) and the Kalman filters is performed. Secondly, an event analysis is carried out in order to detect the different cases: object match, new object, lost object, object splitting and object merging. Taking this information into account, the Kalman filters

are managed. Furthermore, occlusion handling is managed based on a Probabilistic Based Model (PAM). Temporal consistency is evaluated in the association between FGs and SHs and their respective Kalman Filters over time. Consequently, a number of cases are studied: FG and SH match, new shadow and lost shadow. Finally, the tracking results are feedback to the chromatic shadow detector to improve the object and shadow detection. Thus, chromatic shadows are correctly detected in cases with the mentioned segmentation problems.

Consequently, thanks to the data association between FG and SH we have achieved: (i) enhancement of the chromatic shadow detection by detecting shadows, which were not possible to detect before. (ii) improvement of the segmentation for high level processes, such as detection and tracking, by avoiding shadows. (iii) a more robust tracking, since (1) the PAM and the KF tracker are more robust and correctly updated, and (2) erroneous created KFs are deleted.

Qualitative and quantitative results of tests for both outdoor and indoor sequences from well-known databases validate the presented approach. Overall, our approach gives a more robust and accurate shadow detection and foreground segmentation in comparison to the state-of-the-art methods.

Unlike other approaches, our method does not make any a-priori assumptions about camera location, surface geometries, surface textures, shapes and types of shadows, objects, and background. Experimental results show the performance and accuracy of our approach in different shadowed materials and illumination conditions.

Nevertheless, some remarks have to be said with respect to the bottom-up part (chromatic shadow detector) and the top-down part (shadow tracking). The chromatic shadow detector needs a reasonable resolution to work correctly, and noisy and blurred images intensify the camouflage problems. Furthermore, shadow regions need to have a minimum area for analysis, or there might not be enough information for a proper shadow detection and classification. The "bluish effect" gives very good results for some outdoor sequences. However, sometimes it does not work as defined theoretically, since it is affected by external factors, such as the sensibility of the camera and image compression. For the tracking process, targets are assumed to move with a reasonable velocity in comparison to the frame rate. Since, objects which move quickly and change their appearance suddenly are difficult to track.

5.2 Aplications

The work presented in this Thesis has been already used for higher levels on the Human Sequence Evaluation (HSE), see appendix B and [16] for more information about HSE, and applications giving satisfactory results.

The approach has been used in the behaviour analysis field. The motion segmentation is the first procedure to be applied in a cognitive vision system (CVS) devoted to analyse behaviour in image sequences. The correct separation of the moving objects from the background allows a better interpretation of their motion at higher level procedures [2].

It has also been used in the Augmented Reality field. The segmentation has been tested on several application domains: in [3], a combination of real-time human

agents and behaviour-based virtual agents allowed to generate augmented image sequences where real and virtual agents show a certain level of interaction. For this application, a good segmentation is crucial in order to get a good estimation of real agent's silhouette and therefore perform a realistic image composition. Additionally, these augmented image sequences can be later used to measure the performance of segmentation and tracking algorithms [4], since the number of virtual agents can be incremented gradually, thus increasing the segmentation complexity of the resulting sequence."

The segmentation also has been applied to a series of applications, including: (i) interpretation and indexing of video events and behaviours [11], (ii) generation of multilingual NL descriptions of videos [12], and (iii) authoring tools for component-performance evaluation [13].

5.3 Open Issues and Future work

Our motion segmentation approach copes with the non-physical changes in the scene such as local and global illumination problems. Nonetheless, it does not cope with the physical changes in the scene such as when objects are deposited or removed from the scene. Then, in the future work, an updating process should be embedded to the approach in order to incorporate objects to the background model. Furthermore, the use of a pixel-updating process can help to reduce the false positive pixels obtained by using the intensity mask due to drastic illumination changes. In addition, detected motionless objects should be part of a multilayer background model. Moreover, colour invariant normalisations or colour constancy techniques can be used to improve the colour model. The edge model can be enhanced avoiding false edges due to local intense illumination changes. The discrimination between the agents and the local environments can be enhanced by using of new cues such as texture information.

In future work, edge-linking or B-spline techniques can be used to avoid the partial loss of foreground borders due to camouflage, and thereby improve the edge model. Thus, avoiding shadows miss-classified as foreground objects, when shadows and objects are not split correctly.

Another interesting aspect is applying the direction of penumbra to umbra for a cast shadow to discriminate between foregrounds and shadows, when the image region does not have gradient nor similar temporal and spatial chrominance angle and brightness distortions. Additionally, fg. detection could be improved using a combination of a physics colour model and a camera model. Furthermore, a probabilistic scheme based on the feedback from the tracking shadows could be taking into account in the shadow detection process.

In spite of the main subject of this thesis is motion segmentation and shadow detection, for future work the tracking system has to be improved, in order to be able to test the approach on a more complex scenario, such as crowded scenes or situations with multiple grouping and splitting processes. The proposed tracking system performs well for basic scene situations. However, for more complex scenario the tracks are sometimes lost, and the tracker information becomes corrupted. Thus, in future work a high level tracker is needed, which can manage the low level trackers

in a top-down architecture.

Furthermore, detection of gradient changes can be applied in the occlusion handling to improve the splitting process of occluded objects. Moreover, the target representation can be refined by including structure components and shape cues. E.g., body-part histograms and salient points could enhance agent tracking during long-term partial occlusions, while SIFT descriptors could provide new ways of target discrimination.

On the other hand, more complex cases of the temporal consistency of the data association between FG and SH with their respectively assigned Kalman Filter have to be studied. E.g., a possible shadow without any associated FG, because it has been split up before or the positive edge mask (ed_pos mask) has divided it erroneously. In order to solve these cases a high level event analysis has to be applied.

In order to solve the problem with fast moving objects relatively to the frame rate, with sudden changes in their appearance, in future work the momentum can be applied for the tracker information. In this way, the tracker could improve the accuracy through the historical map of the previous trackers.

Finally, high-level information from the Human Sequence Evaluation structure, see appendix B for more information, can be applied to enhance the detection process in subsequent frames.

Appendix A

Kalman Filter

The Kalman filter [31] is a stochastic state estimator developed by Rudolph E. Kalman in 1960. It implements a recursive algorithm which works in a prediction-correction way, estimating the system state from noisy measures. The estimator is optimal in the sense that it minimises the steady-state error covariance:

$$\mathbf{P} = \lim_{t \to \infty} \mathbb{E}\left[(\mathbf{x} - \hat{\mathbf{x}}) (\mathbf{x} - \hat{\mathbf{x}})^T \right]. \tag{A.1}$$

However, strong assumptions are required: the transition model must be linear Gaussian, and the sensor model must be Gaussian. Nevertheless, albeit these conditions rarely exist, the filter still works reasonably well for many applications, and it has been widely used.

It works as follows. The process is assumed to be governed by a linear stochastic difference equation:

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \boldsymbol{\omega}_t, \tag{A.2}$$

where

- $\mathbf{x}_t \in \mathcal{R}^n$ is the system state, n the state-space dimension, and t a discrete time index,
- $\mathbf{A}$ is a $n\,x\,n$ matrix describing the linear transition model,
- $\boldsymbol{\omega}_t \sim \mathcal{N}(0, \mathbf{Q})$ is the process noise, and $\mathbf{Q}$ the noise covariance. Hereby, zero-mean white additive Gaussian noise is assumed to represent modelling uncertainties and disturbances.

The measure process is assumed to be governed by the next equation:

$$\mathbf{z}_t = \mathbf{C}\mathbf{x}_t + \boldsymbol{\nu}_t, \tag{A.3}$$

where,

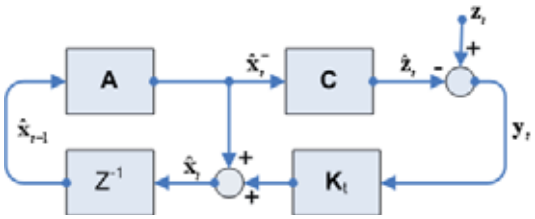


Figure A.1: Diagram block of a Kalman state estimator. See text for details.

- $\mathbf{z}_t \in \mathcal{R}^m$ is the measure vector, and m the measure-space dimension,
- $\mathbf{C}$ is a $m\,x\,n$ matrix relating the state to measure,
- $\boldsymbol{\nu}_t \sim \mathcal{N}(0, \mathbf{R})$ is the sensor noise, and $\mathbf{R}$ the noise covariance. Hereby, zero-mean white additive Gaussian noise is assumed to represent measurement noise.

It is also assumed that both process and measurement noise are uncorrelated:

$$Cov\left(\boldsymbol{\nu}_t \boldsymbol{\omega}_t^T\right) = 0. \tag{A.4}$$

The initial state is unknown, but it is assumed that it follows a normal law:

$$\mathbf{x}_0 \sim \mathcal{N}(\mu_0, \mathbf{P}_0), \tag{A.5}$$

where

- $\mathbf{x}_0$ is the system initial state,
- μ_0 is the initial distribution mean,
- $\mathbf{P}_0$ is the initial distribution covariance.

Independence of process noises $\boldsymbol{\omega}_t, \boldsymbol{\nu}_t$ and initial state $\mathbf{x}_0$ is assumed.

The filter works in two steps which are recursively performed —a block diagram is shown in Fig. A.1. In the first one, a prediction is made: the expectation and covariance are propagated according to the the dynamic model, thereby obtaining the temporal prior:

$$\begin{aligned} \hat{\mathbf{x}}_t^- &= \mathbb{E}\left[\mathbf{A}\mathbf{x}_{t-1} + \boldsymbol{\omega}_t\right] \\ &= \mathbf{A}\hat{\mathbf{x}}_{t-1}, \end{aligned} \tag{A.6}$$

and the prior covariance matrix:

$$\begin{aligned}\mathbf{P}_t^- &= \mathbb{E}\left[\left(\mathbf{x}_t - \mathbb{E}\left[\mathbf{x}_t\right]\right)\left(\mathbf{x}_t - \mathbb{E}\left[\mathbf{x}_t\right]\right)^T\right] \\ &= \mathbb{E}\left[\left(\mathbf{A}\left(\mathbf{x}_{t-1} - \mathbb{E}\left[\mathbf{x}_{t-1}\right]\right) + \mathbf{w}_t\right)\left(\mathbf{A}\left(\mathbf{x}_{t-1} - \mathbb{E}\left[\mathbf{x}_{t-1}\right]\right) + \boldsymbol{\omega}_t\right)^T\right] \\ &= \mathbf{A}\mathbf{P}_{t-1}\mathbf{A}^T + \mathbf{Q}. \end{aligned} \tag{A.7}$$

After obtaining the new measurement $\mathbf{z}_t$, the second step is carried out, and values are updated according to the observation likelihood:

$$\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_t^- + \mathbf{K}_t\mathbf{y}_t, \tag{A.8}$$

$$\mathbf{P}_t = \mathbf{I} - \mathbf{K}_t\mathbf{C}\mathbf{P}_t^-, \tag{A.9}$$

where:

$$\mathbf{y}_t = \mathbf{z}_t - \mathbf{C}\hat{\mathbf{x}}_t^-, \tag{A.10}$$

is called the *innovation* or the *residual,*

$$\mathbf{S}_t = \mathbf{C}\mathbf{P}_t^-\mathbf{C}^T + \mathbf{R}, \tag{A.11}$$

is called the *innovation covariance*, and

$$\mathbf{K}_t = \mathbf{P}_t^-\mathbf{C}^T\mathbf{S}_t^{-1}, \tag{A.12}$$

is known as the *Kalman gain.*

Appendix B

A Framework to Human-Sequence Evaluation

Accomplishing Human Sequence Evaluation (HSE) [16] involves not only human motion analysis, but also behaviour understanding. Therefore, the proposed framework must include the different required system functionalities while making use of cognitive processes. It should not be restricted to Image Processing and Analysis, or Pattern Recognition techniques, but it should also comprehend topics related to Artificial Intelligence, Computational Linguistics, Computer Animation, and Automatic Control. For instance, Computer Animation techniques are taken into account in order to provide graphical information and simulations about the situation which is taking place, as well as predictions about potential future ones; Automatic Control can come into scene to allow machine responses to recognised behaviours, and to operate PTZ cameras.

In this section, the HSE framework presented in [17] is reviewed. HSE defines a complete Cognitive Vision System which transforms image values into semantic descriptions of human behaviour by performing multiple bottom-up and top-down processes. Thus, its aim goes far beyond detecting, tracking and identifying the actions being performed. Its goal is to apply cognition methodologies to understand human behaviour, thereby being able to provide Natural Language (NL) descriptions of what is taking place within the scene, and generating synthetic visual representations of the scene and agents.

Mainly, the implementation of HSE involves three cooperating tasks: (i) the acquisition of a dynamic description of the observed human motion; (ii) the transformation of these quantitative parameters into logic predicates; and (iii) the communication of the obtained results to an human user. The third task can be achieved by means of NL text generation —by applying syntax rules to those instantiated conceptual primitives— and by the synthetisation of virtual environments from this conceptual information.

Therefore, multiple issues are demanded in order to accomplish HSE. At the very least, these include (i) active video camera control, (ii) target segmentation, (iii) robust and accurate multiple-target tracking, (iv) target classification, (v) posture and

action recognition, (vi) facial expression analysis, (vii) behaviour understanding, and (viii) communication of those inferred conceptual interpretations to human operators. Thus, the computational knowledge of the three different channels of human motion, namely the motion of agents —trajectories, bodies— postures and actions, and faces —expressions and emotions, is linked together in the same discourse domain.

Unfortunately, adversities common to other Computer Vision areas could cause system failures, for instance due to acquisition conditions, uncontrolled illumination, shadows, cluttered backgrounds —-possibly in motion—- etc. In addition, dealing with people entails numerous special difficulties such as posture changes, huge appearance variability, or unforeseeable motion changes. Moreover, conceptual interpretations of motion may include uncertainty due to the inaccuracy of the semantic terms used to explain human behaviour.

Due to this complexity, a HSE system is here presented as a highly modularised and hierarchically organised framework, see Fig. B.1. Thus, multiple co-operating modules are defined through the different levels. They work following both top-down and bottom-up approaches, thereby defining the interactions of different Computer Vision algorithms with other components, such as human behaviour modelling and NL text generation. This is done while taking into account the uncertainty generated during motion naming, i.e. the textual explanation of perceived motion. HSE requires intermediate models of human motion to associate geometric knowledge with conceptual statements. Thus, each level exploits the a-priori knowledge provided by models and context.

Levels are defined according to main functionalities. Thus, each level performs some general task such as providing a machine interface $-$*Active Sensor Level (ASL), User Interaction Level (UIL)*$-$ processing and analysing the image sequence $-$*Image Signal Level (ISL), Picture Domain Level (PDL), Scene Domain Level (SDL)*$-$ and describing and reasoning over the obtained quantitative results $-$*Conceptual Integration Level (CIL), Behaviour Interpretation Level (BIL).*

The whole structure is highly interconnected, and each level receives inputs from higher and lower ones, providing the system with redundancy. The inter-level communication can be seen in three different ways: first of all, a data stream is provided to the higher levels by lower ones including all the results obtained in the bottom-up process; secondly, higher levels feed back the lower ones in a top-down process; at the same time, higher levels can act on the lower ones by tuning the parameters, and selecting different operation modes, models or approaches depending on what is known about the current scene, and what goals are pursued.

Visual sensors provide information to the system about the real world at the ASL. Pieces of reality are captured by the cameras according to the kind of sensor used and the visual field. Thus, this level includes hardware devices, such as the camera itself and the acquisition cards, and models to deal with these devices. Further, being the sensors active, the system is allowed to modify the camera parameters depending on the task and scene conditions. At the ISL, the sequence of image data is processed by segmenting potential targets. The resulting foreground regions are the basis for the following level, the PDL. Possible segmentation errors generated at the ISL are handled here by means of representation, classification, and tracking techniques. At the SDL, the 3D configuration of the scene is used to compute the parameters of each

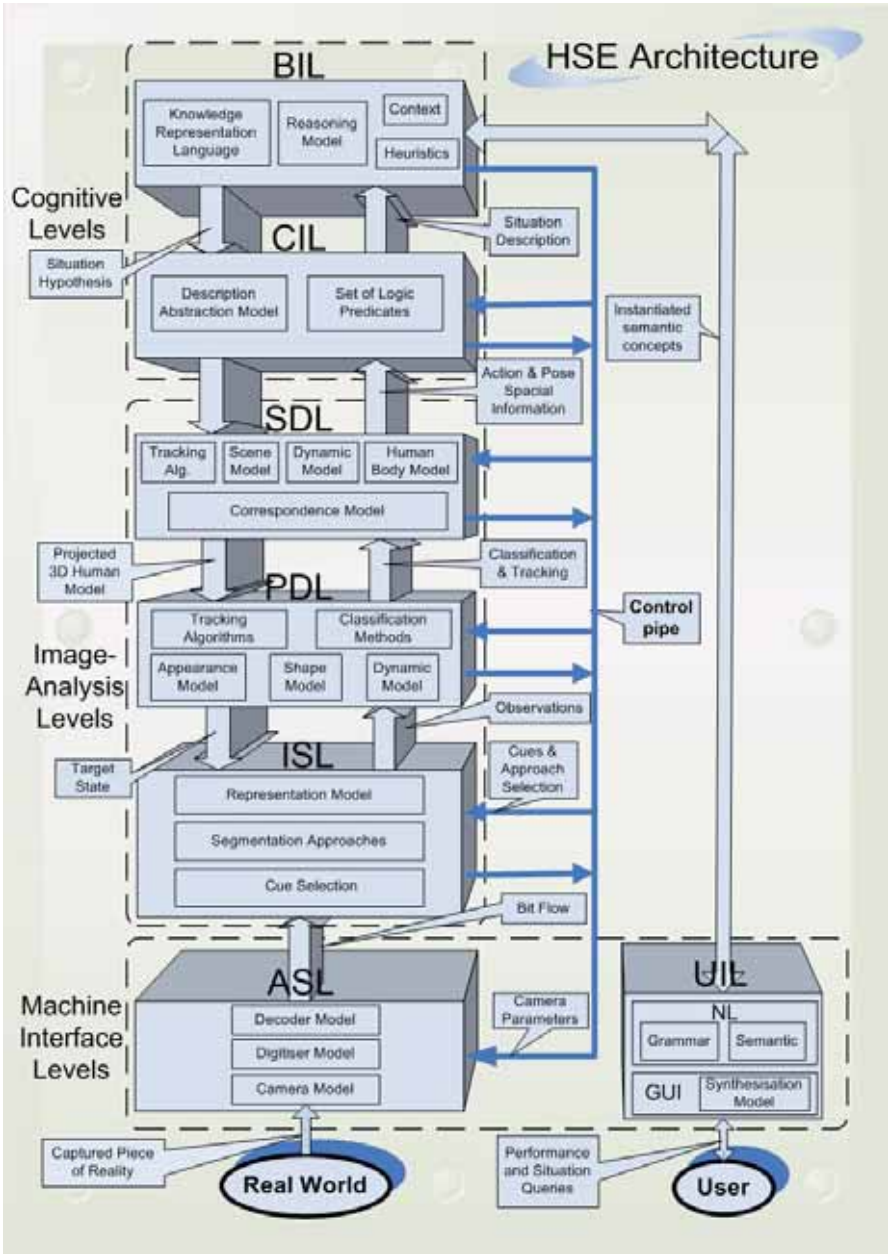


Figure B.1: Human-Sequence Evaluation framework from [55].

agent within its 3D environment.

Results obtained at either the PDL or the SDL are forwarded to the CIL to instantiate semantic predicates for a given agent and time step. These qualitative descriptions are used to generate interpretations of its motion, as well as conceptual relationships of the agent and its environment. Instantiated predicates are fed forward to the BIL, where the expected temporal evolution of descriptions is a-priori modelled in order to generate coherent spatio-temporal interpretations.

The UIL attempts to provide a Natural-Language description of what is actually happening within the scene. The quantitative information generated at lower levels is associated with qualitative semantic terms such as verbs, nouns, adverbs and adjectives, and it is used to generate natural sentences by means of syntactical, morphological, and orthographic rules. Finally, an interactive, Graphical User Interface (GUI) allows a single human operator to monitor a significant area of interest. Thus, the GUI automatically places virtual agents representing people and vehicles into a synthetic view of the environment. This approach has the benefit that visualisation of scene events is no longer tied to the original resolution and viewpoint of a single video sensor. Through this interface, the user can act on individual sensor units, modify the system parameters, select one particular approach, and ask for situation descriptions, behaviour explanations, and synthetic simulations. An audio-based interactive environment can also be here considered to enhance the user interaction.

Appendix C

Publications

- Ivan Huerta, Xavier Roca and Jordi Gonzalez. Multiple Cues Fusion for Robust Motion Segmentation using Background Subtraction. *Journal of Transactions on Image Processing (Under review process).*

- Ivan Huerta, Michael Holte, Thomas Moeslund, and Jordi Gonzàlez. Detection and Removal of Chromatic Moving Shadows in Surveillance Scenarios. In *12th International Conference on Computer Vision (ICCV2009).* Kyoto, Japan, October, 2009

- Ariel Amato, Mikhail Mozerov, Iván Huerta, Jordi Gonzàlez, Juan José Villanueva. Background Subtraction Technique Based on Chromaticity and Intensity Patterns. In *19th International Conference on Pattern Recognition (ICPR'2008).* Tampa, Florida, USA, December, 2008

- Ivan Huerta, Ariel Amato, Jordi Gonzàlez, and Juan José Villanueva. Fusing Edge Cues to handle Colour Problems in Image Segmentation. In *5th International Workshop on Articulated Motion and Deformable Objects (AMDO'2008).* Andratx, Mallorca, Spain, July, 2008

- Ivan Huerta, Daniel Rowe, Mikhail Mozerov, Jordi Gonzàlez. Improving Background Subtraction based on a Casuistry of Colour-Motion Segmentation Problems. In *3rd Iberian Conference on Pattern Recognition and Image Analysis (ibPRIA'2007).* Girona, Spain, June, 2007

- Daniel Rowe, Ivan Huerta, Jordi Gonzàlez, Juan J. Villanueva. Robust Multiple-People Tracking Using Colour-Based Particle Filters. In *3rd Iberian Conference on Pattern Recognition and Image Analysis (ibPRIA'2007).* Girona, Spain, June, 2007

- Daniel Rowe, Jordi González, Ivan Huerta, Juan José Villanueva. On Reasoning over Tracking Events. In *15th Scandinavian Conference on Image Analysis (SCIA'2007).* Aalborg, Denmark, June, 2007

- Iván Huerta, Daniel Rowe, Jordi Gonzàlez, Juan José Villanueva. Efficient Incorporation of Motionless Foreground Objects for Adaptive Background Segmentation. In *4th International Workshop on Articulated Motion and Deformable Objects (AMDO-e'2006).* Andratx, Mallorca, Spain, July, 2006

- Iván Huerta, Ariel Amato, Marco Pedersoli, Jordi González. Motion Segmentation of Image Sequences: A survey. In *3rd International Workshop on Current Challenges in Computer Vision(CVCRD'08).* Bellaterra, Spain, 2008

- Marco Pedersoli, Iván Huerta, Jordi González, Juan J. Villanueva. Fast Human Detection using Multiresolution Cascade. In *3rd International Workshop on Current Challenges in Computer Vision(CVCRD'08).* Bellaterra, Spain, 2008

- Iván Huerta Casado, Daniel Rowe, Miguel Viñas, Mikhail Mozerov, Jordi Gonzàlez. Background Subtraction Fusing Colour, Intensity and Edge Cues. In *2nd CVC Workshop on Computer Vision: Progress of Research and Development (CVCRD' 07).* Bellaterra, Spain,2007

- Daniel Rowe, Iván Huerta, Jordi Gonzàlez, Juan J. Villanueva. A Hierarchical Architecture to Multiple Target Tracking. In *2nd CVC Workshop on Computer Vision: Progress of Research and Development (CVCRD'07).* Bellaterra, Spain, 2007

- Iván Huerta Casado, Daniel Rowe, Jordi Gonzàlez, Juan J. Villanueva. Improving Foreground Detection for Adaptive Background Segmentation. In *1st CVC Workshop on Computer Vision: Progress of Research and Development (CVCRD'06).* Bellaterra, Spain,2006

- Daniel Rowe, Iván Huerta, Jordi Gonzàlez, Juan J. Villanueva. Detection and Tracking of Multiple Agents in Unconstrained Environments. In *1st CVC Workshop on Computer Vision: Progress of Research and Development (CVCRD'06).* Bellaterra, Spain, 2006

- Iván Huerta. Image-Sequence Segmentation in Uncontrolled Environments. CVC Technical Report 115, CVC (UAB) , July, 2007

Bibliography

[1] A. Amato, M. Mozerov, I. Huerta, J. Gonzàlez, and J.J. Villanueva. Background subtraction technique based on chromaticity and intensity patterns. In *19th ICPR'2008*, December 2008.

[2] P. Baiget, C. Fernández, X. Roca, and J. Gonzàlez. Automatic learning of conceptual knowledge for the interpretation of human behavior in video sequences. In *Ibpria 2007*, Girona, Spain, 2007. Springer LNCS.

[3] P. Baiget, C. Fernández, X. Roca, and J. Gonzàlez. Generation of augmented video sequences combining behavioral animation and multi-object tracking. *CAVW*, 20(4):473–489, 2009.

[4] P. Baiget, X. Roca, and J. Gonzàlez. Autonomous virtual agents for performance evaluation of tracking algorithms. In *AMDO'2008*, Andratx, Mallorca, Spain, 2008.

[5] A. Bugeau and P. Perez. Detection and segmentation of moving objects in highly dynamic scenes. In *IEEE CVPR'07*, pages 1–6, June 2008.

[6] Y. Chen, C. Chen, C. Huang, and Y. Hung. Efficient hierarchical method for background subtraction. *Pattern Recognition*, 40(10):2706–2715, October 2007.

[7] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Detecting moving objects, ghosts, and shadows in video streams. *IEEE TPAMI*, 25(10):1337–1342, October 2003.

[8] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti. Improving shadow suppression in moving object detection with hsv color information. In *Proceedings. IEEE Intelligent Transportation Systems*, pages 334–339, Oakland, USA, 2001.

[9] R. Cucchiara, C. Grana, and G. Tardini. Track-based and object-based occlusion for people tracking refinement in indoor surveillance. In *ACM 2nd International Workshop on Video Surveillance and Sensor Networks*, pages 388–393, New York, NY, USA, 2004.

[10] A. Elgammal, D. Harwood, and L. S. Davis. Nonparametric background model for background subtraction. In *ECCV'00*, pages 751–767, Dublin, 2000.

[11] C. Fernández, P. Baiget, X. Roca, and J. Gonzàlez. Interpretation of complex situations in a semantic-based surveillance framework. *SPIC*, 2008.

[12] C. Fernández, P. Baiget, X. Roca, and J. Gonzàlez. *Exploiting Natural Language Generation in Scene Interpretation*. Elsevier Science and Technology Book Group, 2009.

[13] C. Fernández, P. Baiget, X. Roca, and J. Gonzàlez. Augmenting video surveillance footage with virtual agents for incremental event evaluation. *PRL*, 2010.

[14] G.D. Finlayson, S.D. Hordley, C. Lu, and M.S. Drew. On the removal of shadows from images. *IEEE TPAMI*, 28(1):59–68, January 2006.

[15] D.M. Gavrila. The visual analysis of human movement: A survey. *CVIU*, 73(1):82–98, 1999.

[16] J. Gonzàlez. *Human Sequence Evaluation: the Key-frame Approach*. PhD thesis, Barcelona, Spain, May 2004.

[17] J. Gonzàlez, D. Rowe, J. Varona, and F.Xavier Roca. Understanding dynamic scenes based on human sequence evaluation. *Image and Vision Computing, doi: 10.1016/j.imavis.2008.02.004*, February 2008.

[18] D. Gusfield. The stable marriage problem: structure and algorithms. *MIT Press*, 1989.

[19] I. Haritaoglu, D. Harwood, and L.S. Davis. W4: Real-time surveillance of people and their activities. *IEEE TPAMI*, 22(8):809–830, 2000.

[20] 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.

[21] M. Heikkila and M. Pietikainen. A texture-based method for modeling the background and detecting moving objects. *IEEE TPAMI*, 28(4):657–662, 2006.

[22] 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.

[23] W. Hu, T. Tan, L. Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE TSMC*, 34(3):334–352, 2004.

[24] S. Huang, L. Fu, and P. Hsiao. Region-level motion-based background modeling and subtraction using mrfs. *IEEE TIP*, 16(5):1446–1456, May 2007.

[25] I. Huerta, A. Amato, J. Gonzàlez, and J.J. Villanueva. Fusing edge cues to handle colour problems in image segmentation. In *Proc. AMDO'08*, volume 5098, pages 279–288, Andratx, Mallorca, Spain, 2008. Springer LNCS.

[26] I. Huerta, M. Holte, T.B. Moeslund, and J. Gonzàlez. Detection and removal of chromatic moving shadows in surveillance scenarios. In *ICCV2009*, Kyoto, Japan, 2009.

[27] I. Huerta, D. Rowe, M. Mozerov, and J. Gonzàlez. Improving background subtraction based on a casuistry of colour-motion segmentation problems. In *Ibpria'07*, volume 2, pages 475–482, Girona, Spain, 2007. Springer LNCS.

[28] Y. Ivanov, A. Bobick, and J. Liu. Fast lighting independent background subtraction. *IJCV*, 37(2):199–207, June 2000.

[29] 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.

[30] 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.

[31] R.E. Kalman. A new approach to linear filtering and prediction problems. *Trans. ASME J.of Basic Engineering*, 1960.

[32] M. Karaman, L. Goldmann, D. Yu, and T. Sikora. Comparison of static background segmentation methods. In *VCIP '05*, Beijing, China, July 2005.

[33] 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.

[34] D. Koller, K. Daniilidis, and H.-H. Nagel. Model-based object tracking in monocular image sequences. *Kluwer Academic Publishers*, 1993.

[35] A. Leone and C. Distante. Shadow detection for moving objects based on texture analysis. *Pattern Recognition*, 40(4):1222–1233, April 2007.

[36] 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.

[37] L. Maddalena and A. Petrosino. A self-organizing approach to background subtraction for visual surveillance applications. *IEEE TIP*, 17(7):1168–1177, July 2008.

[38] V. Mahadevan and N. Vasconcelos. Background subtraction in highly dynamic scenes. In *IEEE CVPR'08*, pages 1–6, June 2008.

[39] A.-R. Mansouri. Region tracking via level set pdes without motion computation. *IEEE TPAMI*, 24(7):947–961, 2002.

[40] N. Martel-Brisson and A. Zaccarin. Learning and removing cast shadows through a multidistribution approach. *IEEE TPAMI*, 29(7):1133–1146, 2007.

[41] N. Martel-Brisson and A. Zaccarin. Kernel-based learning of cast shadows from a physical model of light sources and surfaces for low-level segmentation. In *IEEE CVPR'08*, pages 1–8, June 2008.

[42] A. McIvor. Background subtraction techniques. In *In Proc. of Image and Vision Computing*, Auckland, New Zealand, 2000.

[43] S. J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler. Tracking groups of people. *CVIU*, 80(1):42–56, 2000.

[44] 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.

[45] T. B. Moeslund, A. Hilton, and V. Kruger. A survey of advances in vision-based human motion capture and analysis. *CVIU*, 104:90–126, November-December 2006.

[46] T.B. Moeslund and E. Granum. A survey of computer vision based human motion capture. *CVIU*, 81(3):231–268, March 2001.

[47] S. Nadimi and B. Bhanu. Physical models for moving shadow and object detection in video. *IEEE TPAMI*, 26(8):1079–1087, August 2004.

[48] K. Onoguchi. Shadow elimination method for moving object detection. In *ICPR*, volume 1, pages 583–587, 1998.

[49] K. A. Patwardhan, G. Sapiro, and V. Morellas. Robust foreground detection in video using pixel layers. *IEEE TPAMI*, 30(4):746–751, April 2008.

[50] 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.

[51] A. Prati, I. Mikic, M. Trivedi, and R. Cucchiara. Detecting moving shadows: Algorithms and evaluation. *IEEE TPAMI*, 25(7):918–923, July 2003.

[52] R.J. Radke, S.Andra, O. Al-Kofahi, and B.Roysam. Image change detection algorithms: a systematic survey. *IEEE TIP*, 14(3):294–307, March 2005.

[53] R.O'Callaghan and T. Haga. Robust change-detection by normalised gradient-correlation. In *IEEE CVPR'07*, pages 1–8, June 2007.

[54] D. Roth, P. Doubek, and L. V. Gool. Bayesian pixel classification for human tracking. In *Proceedings of the IEEE Workshop on Motion and Video Computing*, pages 78–83, Breckenridge, CO, USA, 2005.

[55] D. Rowe. *Towards Robust Multiple-Target Tracking in Unconstrained Human-Populated Environments*. PhD thesis, Barcelona, Spain, 2008.

[56] D. Rowe, J. Gonzàlez, M. Pedersoli, and J. Villanueva. On tracking inside groups. *MVA*, 2010.

[57] E. Salvador, A. Cavallaro, and T. Ebrahimi. Cast shadow segmentation using invariant color features. *CVIU*, 95(2):238–259, August 2004.

[58] O. Schreer, I. Feldmann, U. Goelz, and P. Kauff. Fast and robust shadow detection in videoconference applications. In *Proc. IEEE VIPromCom*, pages 371–375, 2002.

[59] A. Senior, A. Hampapur, Y.-L. Tian, L. Brown, S. Pankanti, and R. Bolle. Appearance models for occlusion handlinge. *Image and Vision Computing*, 24(11):1233–1243, 2006.

[60] Y. Sheikh and M. Shah. Bayesian modeling of dynamic scenes for object detection. *IEEE TPAMI*, 27(11):1778–1792, November 2005.

[61] J. Shen. Motion detection in color image sequence and shadow elimination. *Visual Communications and Image Processing*, 5308:731–740, January 2004.

[62] P. Spagnolo, T.D Orazio, M. Leo, and A. Distante. Moving object segmentation by background subtraction and temporal analysis. *Image and Vision Computing*, 24(5):411–423, May 2006.

[63] J. Stauder, R. Mech, and J. Ostermann. Detection of moving cast shadows for object segmentation. *IEEE Trans. Multimedia*, 1(1):65–76, March 1999.

[64] C. Stauffer, W. Eric, and L. Grimson. Learning patterns of activity using real-time tracking. *IEEE TPAMI*, 22(8):747–757, 2000.

[65] 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.

[66] 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.

[67] M. Vanrell, F. Lumbreras, A. Pujol, R. Baldrich, J. Llados, and J.J. Villanueva. Colour normalisation based on background information. In *Proceedings of IEEE International Conference on Image Processing*, volume 1, pages 874–877, 2001.

[68] J. J. Veenman, M. J. T. Reinders, and E. Backer. Resolving motion correspondence for densely moving points. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 23(1):54–72, 2001.

[69] L. Wang, W. Hu, and T. Tan. Recent developments in human motion analysis. *Pattern Recognition*, 36(3):585–601, 2003.

[70] L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette analysis-based gait recognition for human identification. *IEEE TPAMI*, 25(12):1505–1518, Dec 2003.

[71] Yair Weiss. Deriving intrinsic images from image sequences. In *Proc. ICCV'01*, volume 02, pages 68–75, Vancouver, Canada, 2001.

[72] C.R. Wren, A. Azarbayejani, T. Darrell, and A.P. Pentland. Pfinder: Real-time tracking of the human body. *IEEE TPAMI*, 19(7):780–785, 1997.

[73] L.Q. Xu and P. Puig. A hybrid blob- and appearance-based framework for multi-object tracking through complex occlusions. In *Proceedings 2nd Joint IEEE International Workshop on VS-PETS*, Beijing, China, 2005.

[74] 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.

[75] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *ACM Computing Surveys*, 38(4), 2006.

[76] J. Zhong and S. Sclaroff. Segmenting foreground objects from a dynamic textured background via a robust kalman filter. In *IEEE ICCV'03*, pages 44–50, October 2003.

[77] Z. Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *Proc. ICPR'04*, volume 2, pages 23–26, August 2004.

[78] 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.