Background Subtraction in Real Applications: Challenges, Current Models and Future Directions

Belmar Garcia-Garcia · Thierry Bouwmans · Alberto Jorge Rosales Silva

Received: date / Accepted: date

Abstract Computer vision applications based on videos often require the detection of moving objects in their first step. Background subtraction is then applied in order to separate the background and the foreground. In literature, background subtraction is surely among the most investigated field in computer vision providing a big amount of publications. Most of them concern the application of mathematical and machine learning models to be more robust to the challenges met in videos. However, the ultimate goal is that the background subtraction methods developed in research could be employed in real applications like traffic surveillance. But looking at the literature, we can remark that there is often a gap between the current methods used in real applications and the current methods in fundamental research. In addition, the videos evaluated in large-scale datasets are not exhaustive in the way that they only covered a part of the complete spectrum of the challenges met in real applications. In this context, we attempt to provide the most exhaustive survey as possible on real applications that used background subtraction in order to identify the real challenges met in practice, the current used background models and to provide future directions. Thus, challenges are investigated in terms of camera (i.e CCD cameras, omnidirectional cameras,...), foreground objects and environments. In addition, we identify the background models that are effectively used in these applications in order to find potential usable recent background models in terms of robustness, time and memory requirements.

Keywords Background Subtraction · Background Initialization · Foreground Detection · Visual Surveillance

T. Bouwmans
Lab. MIA, Univ. La Rochelle, France
Tel.: +05.46.45.72.02
E-mail: tbouwman@univ-lr.fr,
B. Garcia-Garcia
Instituto Politecnico Nacional, Ciudad de México, México
A. Rosales Silva
Instituto Politecnico Nacional, Ciudad de México, México

1 Introduction

With the rise of the different sensors, background initialization and background subtraction are widely employed in different computer vision applications based on video taken by fixed cameras. These applications involve a big variety of environments with different challenges and different kinds of moving foreground objects of interest. The most well-known and oldest applications are surely intelligent visual surveillance systems of human activities such as traffic surveillance of road, airport and maritime surveillance [180]. But, detection of moving objects are also required for intelligent visual observation systems of animals and insects in order to study the behavior of the observed animals in their environment. However, it requires visual observation in natural environments such as forest, river, ocean and submarine environments with specific challenges. Other miscellaneous applications like optical motion capture, human-machine interaction system, vision-based hand gesture recognition, content-based video coding and background substitution also need in their first step either background initialization and background subtraction. Even if detection of moving objects is widely employed in these real application cases, no full survey can be found in literature that identifies, reviews and groups in one paper the current models and the challenges met in videos taken with fixed cameras for these applications. Furthermore, most of the research are done on large-scale datasets that often consist of videos which are taken in the aim of evaluation by researchers, and thus several challenging situations that appear in real cases are not covered. This context generates a conflict both in terms of used models and addressed challenges between the sophistication of research-grade solutions and real applied problems. More specifically, research focus on future directions with advanced mathematical models, machine learning models, signal processing models and classification models whereas real applications still employ models such as mean, median and MOG models [379] that were developed 20 years ago.

However, background subtraction methods can be classified as 1) statistical models [317,269,259,8], fuzzy models [27,28] and Dempster-schafer models [286] for mathematical concepts; 2) subspace learning models either reconstructive [295,109,204], discriminative [267,121] and mixed [268], robust subspace learning via matrix decomposition [185,186,191,182] or tensor decomposition [367,183,370]), robust subspace tracking [405], support vector machines [248,198,411,389,388], neural networks [261,264,68,67,69], and deep learning [56,416,246,244,279,280], for machine learning concepts; 3) Wiener filter [397], Kalman filter [275], correntropy filter [88], and Chebychev filter [76] for signal processing models; and 4) clustering algorithms [62,210,296,460,428] for classification models. Statistical, fuzzy and Dempster-Schafer models allow to handle imprecision, uncertainty and incompleteness in the data due the different challenges while machine learning concepts allow to learn the background pixel representation in an supervised or unsupervised manner. Signal processing models allow to estimate the background value and the classification models attempt to classify pixels as background or foreground. But, in practice, most of the authors in real applications employed basic techniques (temporal median [361,176,242], temporal histogram [472,224,454,436,265] and filter [207,216,41,275,76]) and relative old techniques like MOG [379] published in 1999, codebook

[210] in 2004 and ViBe [30] in 2009. This fact is due to two main reasons: 1) most of the time the recent advances can not be currently employed in real application cases due their time and memory requirements, and 2) there is also an ignorance in the communities of traffic surveillance and animals surveillance about the recent background subtraction methods with direct real-time ability with low computation and memory requirements. In 2017, Weinstein [422] highlighted the need of collaboration between computer vision researchers and ecologists for both the design of computer vision algorithms and large-scale datasets.

To address the previous mentioned issues, we first attempt in this review to survey most of the real applications that used background initialization and subtraction in their process by classifying them in terms of aims, environments and objects of interests. We reviewed only the publications that specifically address the problem of background subtraction in real applications with experiments on corresponding videos. To have an overview about the fundamental research in this field, the reader can refer to numerous surveys on background initialization [262,263,52,194] and background subtraction methods [49,46,144,48,44,45,55,47,54,51,50]. Furthermore, we also highlight recent background subtraction models that can be directly used in real applications. Finally, this paper is intended for researchers and engineers in the field of computer vision (i.e visual surveillance of human activities), and biologist/ethologist (i.e visual surveillance of animals and insects).

The rest of this paper is as follows. First, we provide in Section 2 a short reminder on the different key points in background subtraction for novices. In Section 3, we provide a preliminary overview of the different real application cases in which background initialization and background subtraction are required. In Section 4, we review the background models and the challenges met in current intelligent visual surveillance systems of human activities such as road, airport and maritime surveillance. In Section 5, intelligent visual observation systems for animals and insects are reviewed in terms of the challenges related to the behavior of the observed animals and its environment. Then, in Section 6, we specifically investigate the challenges met in visual observation of natural environments such as forest, river, ocean and submarine environments. From Section 8 to 10, we review the applications of automated visual analysis of human activities, optical motion capture, human-machine interaction system, vision-based hand gesture recognition, content-based video coding. In Section 11, we survey other miscellaneous applications like background substitution, carried baggage detection, fire detection and OLED defect detection. In Section 12, we provide a discussion identifying the solved and unsolved challenges in these different application cases as well as proposing prospective solutions to address them. Finally, in Section 13, concluding remarks are given.

2 Background Subtraction: A Short Overview

In this section, we remain briefly the aim of background subtraction for segmentation of static and moving foreground objects from a video stream. This task is the fundamental step in many visual surveillance applications for which background subtraction oers a suitable solution which provide a good compromise in terms of quality

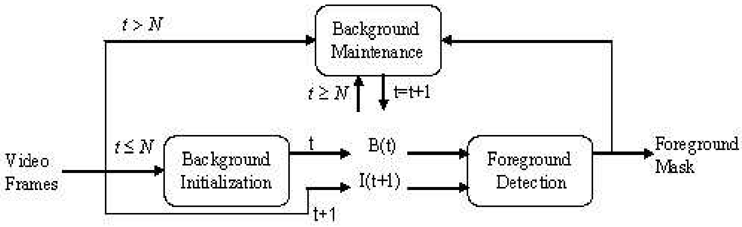

Fig. 1 Background Subtraction Process. N is the number of frames that is used for the background initialization. B_t and I_t are the background and the current image at time t, respectively.

of detection and computation time. The different steps of background subtraction methods as follows:

1. **Background initialization** (also called *background generation*, *background extraction* and *background reconstruction*) consists in computing the first background image.
2. **Background Modeling** (also called *Background Representation*) describes the model use to represent the background.
3. **Background Maintenance** concerns the mechanism of update for the model to adapt itself to the changes which appear over time.
4. **Classification of pixels in background/moving objects** (also called *Foreground Detection*) consists in classifying pixels in the class "background" or the class "moving objects".

These different steps employ methods which have different aims and constraints. Thus, they need algorithms with different features. Background initialization requires *"off-line"* algorithms which are *"batch"* by taking all the data at one time. On the other hand, background maintenance needs *"on-line"* algorithms which are *"incremental"* algorithms by taking the incoming data one by one. Background initialization, modeling and maintenance require reconstructive algorithms while foreground detection needs discriminative algorithms.

Figure 1 shows an overview of a background subtraction process which includes the following stages: **(1)** the background initialization module provides the first background image from N training frames, **(2)** Foreground detection that consists in classifying pixels as foreground or background, is achieved by comparing the background image and the current image. **(3)** Background maintenance module updates the background image by using the previous background, the current image and the foreground detection mask. The steps **(2)** and **(3)** are executed repeatedly as time progresses. It is important to see that two images are compared and that for it, the methods compare a sub-entity of the entity background image with its corresponding sub-entity in the current image. This sub-entity can be of the size of a pixel, a region or a cluster. Furthermore, this sub-entity is characterized by a *"feature"* which can be either directly obtained by the sensors (color features, stereo features), hand-crafted made (edge features, texture features and motion feature), or (deep) learned features. For more details about these features, the reader is refer to the exhaustive survey of

Bouwmans et al. [53]. Developing a background subtraction method, researchers and engineers must design each step and choose the features in relation to the challenges they want to handle in the concerned applications.

3 Real Applications: A Preliminary Overview

In real application cases, either background initialization and background subtraction are required in video taken by a static camera to generate a clean background image of the filmed scene or to detect static or moving foreground objects.

3.1 Background Initialization based Applications

Background initialization provides a a clean background from video sequence, and thus it is required for several applications as developed by Bouwmans et al. [52]:

1. **Video inpainting:** Video inpainting (also called video completion) tries to fill-in user defined spatio-temporal holes in a video sequence using information extracted in the existent spatio-temporal volume, according to consistency criteria evaluated both in time and space as in Colombari et al. [94]. Note that video inpainting is not necessarily used defined. For example, if a landscape is filmed and there are people in front of it, you may want to remove them automatically.
2. **Privacy protection:** Privacy protection for videos aims to avoid the infringement on the privacy right of people taken in the many videos uploaded to video sharing services, that may contain privacy sensitive information of the people as in Nakashima et al. [287].
3. **Computational photography:** It concerns the case where the user wants to obtain a clean background plate from a set of input images containing cluttering foreground objects.

These real application cases only need a clean background without detection of moving object. As the reader can found surveys for these applications in literature, we do not review them in this paper.

3.2 Background Subtraction based Applications

Segmentation of static and moving foreground objects from a video stream is the fundamental step in many computer vision applications for which background subtraction offers a suitable solution which provide a good compromise in terms of quality of detection and computation time.

1. **Visual Surveillance of Human Activities:** The aim is to identify and track objects of interests in several environments. The most frequent environments are traffic scenes (also called road or highway scenes) for their analysis in order to detect incidents such as stopped vehicles on highways [284, 283, 282, 281] or to traffic density estimation on highways which can be then categorized as empty,

fluid, heavy and jam. Thus, it is needed to detect and track vehicles [146,32] or to count the number of vehicles [406]. Background subtraction can be also used for congestion detection [249,285] in urban traffic surveillance, for illegal parking detection [230,470,339,87,408] and for the detection of free parking places [312,290,85]. It is also important for security in train stations and airports, where unattended luggage can be a main goal. Human activities can be also monitored in maritime scenes to count the number of ships which circulated in a marina or in a harbor [223,325,465,125], and to detect and track ships in fluvial canals. Other environments are store scenes for the detection and tracking of consumers [237, 236,235,17,294].

2. **Visual Observation of Animals and Insects Behaviors:** The system required for intelligent visual observation of animals and insects need to be simple and non-invasive. In this context, a video-based system is suitable to detect and track animals and insects in order to analyze their behavior which is needed **(1)** to evaluate their interaction inside their group such as in the case of honeybees which are social insects and interact extensively with their mates [213] and in the case of mice [327,16]; **(2)** to have a fine-scale analysis of the interaction of the animals with their environment (plants,...) such as in the case of birds in order to have information about the impact of climate change on the ecosystem [214,215, 105,106]; **(3)** to study the behavior of animals in different weather conditions such as in the case of fish in presence of typhoons, storms or sea currents for the Fish4Knowledge (F4K[1]) [374,377,375,376]; **(4)** for census of either endangered or threatened species like various species of fox, jaguar, mountain beaver, and wolf [91,90]. In this case, the location and movements of these animals must be acquired, recorded, and made available for review; **(5)** for livestock and pigs [399] surveillance in farms; and **(6)** to design robots that mimics an animals locomotion such as in Iwatani et al. [179]. Observing nature is the fundamental way to collect data on biodiversity. The increased use of sensor-based observation increases the need for methods to distinguish biological foreground objects from video backgrounds. By using automated observation systems, scientists can better understand and protect the natural world.
3. **Visual Observation of Natural Environments:** The aim is to detect foreign objects in natural environments such as forest, ocean and river to protect the biodiversity in terms of fauna and flora. For example, foreign objects in river and ocean can be floating bottles [475], floating wood [6] [7] or mines [42].
4. **Visual Analysis of Human Activities:** Background subtraction is also used in sport **(1)** when important decisions need to be made quickly as in soccer and in tennis with "Hawk-Eye[2]". It has become a key part of the game; **(2)** for precise analysis of athletic performance, since it has no physical effect on the athlete as in Tamas et al. [385] for rowing motion and in John et al. [196] for aerobic routines; and **(3)** for surveillance as in Bastos [31] for surfers activities. However, elderly fall detection can also be made using background subtraction.
5. **Visual Hull Computation:** Visual hull is used for image synthesis to obtain an approximated geometric model of an object which can be static or not. In the first case, it allows a realistic model by provided an image-based model of objects. In the second case, it allows an image-based model of human which can be used

for optical motion capture. Practically, visual hull is a geometric entity obtained with shape-from-silhouette 3D reconstruction technique introduced by Laurentini [227]. First, it uses background subtraction to separate the foreground object from the background to obtain the foreground mask known as a silhouette which is then considered as the 2D projection of the corresponding 3D foreground object. Based on the camera viewing parameters, the silhouette extracted in each view defines a back-projected generalized cone that contains the actual object. This cone is called a silhouette cone. The intersection of all the cones is called a visual hull, which is a bounding geometry of the actual 3D object. In practice, visual hull computation is employed for the following tasks:

- **Image-based Modeling:** Traditional modeling in image synthesis is made by a modeler but traditional modeling present several disadvantages: **(1)** it is a complex task and it requires time, **(2)** real data of objects are often not known, **(3)** specifying a practical scene is tedious, and **(4)** it presents lack of realism. To address this problem, visual hull is used to extract the model of an object from different images.
- **Optical Motion Capture:** Optical motion capture systems are used **(1)** to give a character life and personality, and **(2)** to allow interactions between the real world and virtual world as in games or virtual reality. In the first application, an actor is filmed on up to 200 cameras which monitor his movements precisely. Then, by using background subtraction, these movements are extracted and translated onto the character. In the second application, the gamer is filmed by a conventional camera or a RGB-D camera such as Microsoft's Kinect. His movements are tracked to interact with virtual objects.

6. **Human-Machine Interaction (HMI):** In several applications, it requires human-machine interaction such as in arts [234], games and ludo-applications [20,22, 304]. In the case of games, the gamer can observe his own image or silhouette composed into a virtual scene as in PlayStation Eye-Toy. In the case of ludo-multimedia applications, several ones concern the detection of a selected moving object in a video by an user as in the project Aqu@theque [20,22,304] which requires to detect the selected fish in video and to recognize it in terms of species.
7. **Vision-based Hand Gesture Recognition:** This application requires to detect, track and recognize hand gesture for several applications such as human-computer interface, behavior studies, sign language interpretation and learning, teleconferencing, distance learning, robotics, games selection and object manipulation in virtual environments.
8. **Content-based Video Coding:** In video coding for transmission such as in teleconferencing, digital movies and video phones, only the key frames are transmitted with the moving objects. For example, the MPEG-4 multimedia communication standard enables the content-based functionality by using the video object plane (VOP) as the basic coding element. Each VOP includes the shape and texture information of a semantically meaningful object in the scene. New functionality like object manipulation and scene composition can be achieved because the video bit stream contains the object shape information. Thus, background subtraction can be used in content-based video coding.

9. **Background Substitution:** The aim of background substitution (also called background cut and video matting) is to extract the foreground from the input video and then combine it with a new background. Thus, background subtraction can be used in the first step as in Huang et al. [169].
10. **Miscellaneous applications:** Other applications used background subtraction such as carried baggage detection as in Tzanidou [401], fire detection as in Toreyin et al. [395], and OLED defect detection as in Wang et al. [417].

All these applications require the detection of moving objects in their first step, and possess their own characteristics in terms of challenges due to the location of the camera, the environment and the type of the moving objects. Background subtraction can be applied with one view or a multi-view as in Diaz et al. [103]. In addition, background subtraction can be also used in applications in which cameras are slowly moving [357,111,381]. For example, Taneja et al. [386] proposed to model dynamic scenes recorded with freely moving cameras. Extensions of background subtraction to moving cameras are presented in Yazdi and Bouwmans [439]. But, real applications with moving camera is out of the scope of this review as we limited this paper to applications with fixed camera. Table 1, Table 2, Table 3, Table 4 and Table 5 show an overview of the different applications, the corresponding types of moving objects of interest, and specific characteristics.

4 Intelligent Visual Surveillance of Human Activities

Visual surveillance is the main application of background modeling and foreground detection, and in this section we only reviewed papers that appeared after 1997 because before the techniques used to detect static or moving objects were based on two or three frames differences due to the limitations of the computer. Practically, the goal in visual surveillance is to automatically detect static or moving foreground objects as follows:

- **Static Foreground Objects (SFO):** Detection of abandoned objects is needed to assure the security of the concerned area. A representative approach can be found in Porikli et al. [310] and a full survey on background model to detect stationary objects can be found in Cuevas et al. [96].
- **Moving Foreground Objects (MFO):** Detection of moving objects is needed to compute statistics on the traffic such as in road [473,402,250,391], airport [38] or maritime surveillance [40,254]. The objects of interest are very different such as vehicles, airplanes, boats, persons and luggages. Surveillance can be more specific as in the case of the study for consumer behavior [236,235,17,232,479] or product amount estimation [159,160] in stores.

Figure 2, Figure 3 and Figure 4 show sample of current images and their corresponding foreground detection mask images or ground-truth images in the applications of Intelligent Visual Surveillance. The reader can see that the objects of interest to detect

[1] http://groups.inf.ed.ac.uk/f4k/

[2] http://www.hawkeyeinnovations.co.uk/

Sub-categories-Aims	Objects of interest-Scenes	Authors-Dates
1) Road Surveillance	**1-Cars**	
1.1) Vehicles Detection		
Vehicles Detection	Road Traffic	Zheng et al. (2006) [473]
Vehicles Detection	Urban Traffic (Korea)	Hwang et al. (2009) [177]
Vehicles Detection	Highways Traffic (ATON Project)	Wang and Song (2011) [409]
Vehicles Detection	Aerial Videos (USA)	Reilly et al. (2012) [326]
Vehicles Detection	Intersection (CPS) (China)	Ding et al. (2012) [107]
Vehicles Detection	Intersection (USA)	Hao et al. (2013) [153]
Vehicles Detection	Intersection (Spain)	Milla et al. (2013) [278]
Vehicles Detection	Aerial Videos (VIVID Dataset [93])	Teutsch et al. (2014) [391]
Vehicles Detection	Road Traffic (CCTV cameras)(Korea)	Lee et al. (2014) [229]
Vehicles Detection	CD.net Dataset 2012 [134]	Hadi et al. (2014) [147]
Vehicles Detection	Northern Jutland (Danemark)	Alldieck (2015) [9]
Vehicles Detection	Road Traffic (Weather) (Croatia)	Vujovic et al. (2014) [407]
Vehicles Detection	Road Traffic (Night) (Hungary)	Lipovac et al. (2014) [251]
Vehicles Detection	CD.net Dataset 2012 [134]	Aqel et al. (2015) [12]
Vehicles Detection	CD.net Dataset 2012 [134]	Aqel et al. (2016) [11]
Vehicles Detection	CD.net Dataset 2012 [134]	Wang et al. (2016) [412]
Vehicles Detection	Urban Traffic (China)	Zhang et al. (2016) [462]
Vehicles Detection	CCTV cameras (India)	Hargude and Idate (2016) [156]
Vehicles Detection	Intersection (USA)	Li et al. (2016) [241]
Vehicles Detection	Dhaka city (Bangladesh)	Hadiuzzaman et al. (2017) [148]
Vehicles Detection	Road Traffic (Weather) (Madrid/Tehran)	Ershadi et al. (2018) [120]
1.2) Vehicles Detection/Tracking		
Vehicles Detection/Tracking	Urban Traffic/Highways Traffic (Portugal)	Batista et al. (2008) [32]
Vehicles Detection/Tracking	Intersection (China)	Qu et al. (2010) [320]
Vehicles Detection/Tracking	Downtown Traffic (Night) (China)	Tian et al. (2013) [393]
Vehicles Detection/Tracking	Urban Traffic (China)	Ling et al. (2014) [250]
Vehicles Detection/Tracking	Highways Traffic (India)	Sawalakhe and Metkar (2015) [346]
Vehicles Detection/Tracking	Highways Traffic (India)	Dey and Praveen (2016) [102]
Multi-Vehicles Detection/Tracking	CD.net Dataset 2012 [134]	Hadi et al. (2016) [146]
Vehicles Tracking	NYCDOT video/NGSIM US-101 highway dataset(USA)	Li et al. (2016) [238]
1.3) Vehicles Counting		
Vehicles Counting/Classification	Donostia-San Sebastian (Spain)	Unzueta et al. (2012) [402]
Vehicles Detection/Counting	Road (Portugal)	Toropov et al. (2015) [396]
Vehicles Counting	Lankershim Boulevard dataset (USA)	Quesada and Rodriguez (2016) [322]
1.4) Stopped Vehicles		
Stopped Vehicles	Portuguese Highways Traffic (24/7)	Monteiro et al. (2008) [284]
Stopped Vehicles	Portuguese Highways Traffic (24/7)	Monteiro et al. (2008) [283]
Stopped Vehicles	Portuguese Highways Traffic (24/7)	Monteiro et al. (2008) [282]
Stopped Vehicles	Portuguese Highways Traffic (24/7)	Monteiro (2009) [281]
1.5) Congestion Detection		
	1-Cars	
Congestion Detection	Aerial Videos	Lin et al. (2009) [249]
Free-Flow/Congestion Detection	Urban Traffic (India)	Muniruzzaman et al. (2016) [285]
	2-Motorcycles (Motorbikes)	
Helmet Detection	Public Roads (Brazil)	Silva et al. (2013) [418]
Helmet Detection	Naresuan University Campus (Thailand)	Waranusast et al. (2013) [418]
Helmet Detection	Indian Institute of Technology Hyderabad (India)	Dahiya et al. (2016) [100]
	3-Pedestrians	
Pedestrian Abnormal Behavior	Public Roads (China)	Jiang et al. (2015) [193]

Table 1 Intelligent Visual Surveillance of Human Activities: An Overview (Part I)

are very various in shape and in color such as humans, cars, airplanes, and boats in various challenging environments and situations.

4.1 Traffic surveillance

Traffic surveillance videos present their own characteristics in terms of locations of the camera, environments and types of the moving objects as follows:

– **Location of the cameras:** There are three kinds of videos in traffic video surveillance: **(1)** videos taken by a fixed camera as in most of the cases, **(2)** aerial videos

Categories-Aims	Objects of interest-Scenes	Authors-Dates
Airport Surveillance	**1-Airplane**	
AVITRACK Project (European)	Airports Apron	Blauensteiner and Kampel (2004) [38]
AVITRACK Project (European)	Airports Apron	Aguilera (2005) [5]
AVITRACK Project (European)	Airports Apron	Thirde (2006) [392]
AVITRACK Project (European)	Airports Apron	Aguilera (2006) [4]
	2-Ground Vehicles (Fueling vehicles/Baggage cars)	
AVITRACK Project (European)	Airports Apron	Blauensteiner and Kampel (2004) [38]
AVITRACK Project (European)	Airports Apron	Aguilera (2005) [5]
AVITRACK Project (European)	Airports Apron	Thirde (2006) [392]
AVITRACK Project (European)	Airports Apron	Aguilera (2006) [4]
	3-People (Workers)	
AVITRACK Project (European)	Airports Apron	Blauensteiner and Kampel (2004) [38]
AVITRACK Project (European)	Airports Apron	Aguilera (2005) [5]
AVITRACK Project (European)	Airports Apron	Thirde (2006) [392]
AVITRACK Project (European)	Airports Apron	Aguilera (2006) [4]
Maritime Surveillance	**1-Cargos**	
	Ocean at Miami (USA)	Culibrk et al. (2006) [97]
	Harbor Scenes (Ireland)	Zhang et al. (2012) [450]
	2-Boats	
Stationary Camera	Miami Canals (USA)	Socek et al. (2005) [372]
Dock Inspecting Event	Harbor Scenes (China)	Ju et al. (2008) [197]
Different Kinds of Targets	Baichay Beach (Vietnam)	Tran and Le (2016) [398]
Salient Events (Coastal environments)	Nantucket Island (USA)	Cullen et al. (2012) [99]
Salient Events (Coastal environments)	Nantucket Island (USA)	Cullen (2012) [98]
Boat ramps surveillance	Boat ramps (New Zealand)	Pang et al. (2016) [297]
Ship lock surveillance	Ship lock (China)	Chen et al. (2018) [82]
	3-Sailboats	
Sailboats Detection	UCSD Background Subtraction Dataset	Sobral et al. (2015) [369]
Sailboats Detection	Sailboats videos (China)	Chen et al. (2018) [82]
	4-Ships	
	Italy	Bloisi et al. (2014) [40]
	Fixed ship-borne camera (China) (IR)	Liu et al. (2014) [254]
Different Kinds of Targets	Ocean (South Africa)	Szpak and Tapamo (2011) [382]
Cage Aquaculture	Ocean (Taiwan)	Hu et al. (2011) [167]
	Ocean (Korea)	Arshad et al. (2010) [13]
	Ocean (Korea)	Arshad et al. (2011) [14]
	Ocean (Korea)	Arshad et al. (2014) [15]
	Ocean (Korea)	Saghafi et al. (2012) [338]
Overloaded Ship Identification	Ocean (China)	Xie et al. (2012) [429]
Ship-Bridge Collision	Wuhan Yangtze River (China)	Zheng et al. (2013) [474]
	Wuhan Yangtze River (China)	Mei et al. (2017) [273]
	5-Motor Vehicles	
Salient Events (Coastal environments)	Nantucket Island (USA)	Cullen et al. (2012) [99]
Salient Events (Coastal environments)	Nantucket Island (USA)	Cullen (2012) [98]
	6-People (Shoreline)	
Salient Events (Coastal environments)	Nantucket Island (USA)	Cullen et al. (2012) [99]
Salient Events (Coastal environments)	Nantucket Island (USA)	Cullen (2012) [98]
	7-Floating Objects	
Detection of Drifting Mines	Floating Test Targets (IR)	Borghgraef et al. (2010) [42]
Store Surveillance	**1-People**	
Apparel Retail Store	Panoramic Camera	Leykin and Tuceryan (2005) [236]
Apparel Retail Store	Panoramic Camera	Leykin and Tuceryan (2005) [235]
Apparel Retail Store	Panoramic Camera	Leykin and Tuceryan (2007) [237]
Retail Store Statistics	Top View Camera	Avinash et al. (2012) [17]
	2-Products	
Product Amount Estimation	Camera (Ceiling)	Higa and Iwamoto (2017) [159]
Product Amount Estimation	Camera (Ceiling)	Higa and Iwamoto (2018) [160]

Table 2 Intelligent Visual Surveillance of Human Activities: An Overview (Part II)

as in Reilly [326], in Teutsch et al. [391], and in ElTantawy and Shehata [115, 113, 114, 117, 116], and **(3)** very high resolution satellite videos as in Kopsiaftis and Karantzalos [217]. In the first case, the camera can be highly-mounted [127] or not as in most of the cases. Furthermore, the camera can be mono-directional or omnidirectional [43, 435].

Categories-Aims	Objects of interest-Scenes	Authors-Dates
Birds Surveillance	**Birds**	
Feeder Stations in natural habitats	Feeder Station Webcam/Camcorder Datasets	Ko et al. (2008) [214]
Feeder Stations in natural habitats	Feeder Station Webcam/Camcorder Datasets	Ko et al. (2010) [215]
Seabirds	Cliff Face Nesting Sites	Dickinson et al. (2008) [105]
Seabirds	Cliff Face Nesting Sites	Dickinson et al. (2010) [106]
Observation in the air	Lakes in Northern Alberta (Canada)	Shakeri and Zhang (2012) [354]
Wildlife@Home	Natural Nesting Stations	Goehner et al. (2015) [132]
Fish Surveillance	**Fish**	
1-Tank		
1.1-Ethology		
Aqu@theque Project	Aquarium of La Rochelle	Penciuc et al. (2006) [304]
Aqu@theque Project	Aquarium of La Rochelle	Baf et al. (2007) [22]
Aqu@theque Project	Aquarium of La Rochelle	Baf et al. (2007) [20]
1.2-Fishing		
Fish Farming	Japanese rice fish	Abe et al. (2016) [2]
2-Open Sea		
2.1-Census/Ethology		
EcoGrid Project (Taiwan)	Ken-Ding sub-tropical coral reef	Spampinato et al. (2008) [374]
EcoGrid Project (Taiwan)	Ken-Ding sub-tropical coral reef	Spampinato et al. (2010) [377]
Fish4Knowledge (European)	Taiwans coral reefs	Kavasidis and Palazzo (2012) [202]
Fish4Knowledge (European)	Taiwans coral reefs	Spampinato et al. (2014) [375]
Fish4Knowledge (European)	Taiwans coral reefs	Spampinato et al. (2014) [376]
UnderwaterChangeDetection (European)	Underwater Scenes (Germany)	Radolko et al. (2016) [323]
-	Simulated Underwater Environment	Liu et al. (2016) [252]
Fish4Knowledge (European)	Taiwans coral reefs	Seese et al. (2016) [347]
Fish4Knowledge (European)	Taiwans coral reefs	Rout et al. (2017) [337]
2.2-Fishing		
Rail-based fish catching	Open Sea Environment	Huang et al. (2016) [174]
Fish Length Measurement	Chute Multi-Spectral Dataset [170]	Huang et al. (2016) [170]
Fine-Grained Fish Recognition	Cam-Trawl Dataset [425]/Chute Multi-Spectral Dataset [170]	Wang et al. (2016) [410]
Dolphins Surveillance	**Dolphins**	
Social marine mammals	Open sea environments	Karnowski et al. (2015) [201]
Lizards Surveillance	**Lizards**	
Endangered lizard species	Natural environments	Nguwi et al. (2016) [291]
Mice Surveillance	**Mice**	
Social behavior	Caltech mice dataset [61]	Rezaei and Ostadabbas (2017) [327]
Social behavior	Caltech mice dataset [61]	Rezaei and Ostadabbas (2018) [16]
Pigs Surveillance	**Pigs**	
Farming	Farming box (piglets)	Mc Farlane and Schofield (1995) [271]
Farming	Farming box	Guo et al. (2014) [142]
Farming	Farming box	Tu et al. (2014) [399]
Farming	Farming box	Tu et al. (2015) [400]
Hinds Surveillance	**Hinds**	
Animal Species Detection	Forest Environment	Khorrami et al. (2012) [206]
Insects Surveillance	**1) Honeybees**	
Hygienic Bees	Institute of Apiculture in Hohen-Neuendorf (Germany)	Knauer et al. (2005) [213]
Honeybee Colonies	Hive Entrance	Campbell et al. (2008) [63]
Honeybees Behaviors	Flat Surface - Karl-Franzens-Universitt in Graz (Austria)	Kimura et al. (2012) [211]
Pollen Bearing Honeybees	Hive Entrance	Babic et al. (2016) [19]
Honeybees Detection	Hive Entrance	Pilipovic et al. (2016) [307]
	2) Spiders	
Spiders Detection	Observation Box	Iwatani et al. (2016) [179]
	3) Insects	
Insects Detection	Container (Bulk maize)	De Geus et al. (2019) [101]

Table 3 Intelligent Visual Observation of Animal/Insect Behaviors: An Overview (Part III)

- **Quality of the cameras:** Most of the time CCTV cameras are used but the quality of the cameras can varied from low-quality to high quality (HD Cameras). Low quality cameras which generate one and two frames per second and 100k pixels/frame are used to reduce both the data and the price. Indeed, a video generated by a low quality city camera is roughly estimated to be about 1GB to 10GB of data per day.
- **Environments:** Traffic scenes present highways, roads and urban traffic environments with their different challenges. In highways scenes, there are often shadows and illumination changes. Road scenes often present environments with trees, and their foliage moves with the wind. In urban traffic scenes, there are often illumination changes such as highlight.

Categories-Aims	Objects of interest-Scenes	Authors-Dates
1-Forest Environments	**Woods**	
Human Detection	Omnidirectional cameras	Boult et al. (2003) [43]
Animal Detection	Illumination change dataset [355]	Shakeri and Zhang (2017) [355]
Animal Detection	Camera-trap dataset [442]	Yousif et al. (2017) [442]
2-River Environments	**Woods**	
Floating Bottles Detection	Dynamic Texture Videos	Zhong et al. (2003) [475]
Floating wood detection	River Videos	Ali et al. (2012) [6]
Floating wood detection	River Videos	Ali et al. (2013) [7]
3-Ocean Environments		
Mine Detection	Open Sea Environments	Borghgraef et al. (2010) [42]
Intruders Detection	Open Sea Environments	Szpak and Tapamo (2011) [382]
Boats Detection	Singapore Marine dataset	Prasad et al. (2016) [315]
Boats Detection	Singapore Marine dataset	Prasad et al. (2017) [314]
4-Submarine Environments		
4.1- Swimming Pools Surveillance		
Human Detection	Public Swimming Pool	Eng et al. (2003) [118]
Human Detection	Public Swimming Pool	Eng et al. (2004) [119]
Human Detection	Public Swimming Pool	Lei and Zhao (2010) [233]
Human Detection	Public Swimming Pool	Fei et al. (2009) [124]
Human Detection	Public Swimming Pool	Chan (2011) [73]
Human Detection	Public Swimming Pool	Chan (2013) [74]
Human Detection	Private Swimming Pool	Peixoto al. (2012) [303]
4.2- Tank Environments		
Fish Detection	Aquarium of La Rochelle	Penciuc et al. (2006) [304]
Fish Detection	Aquarium of La Rochelle	Baf et al. (2007) [22]
Fish Detection	Aquarium of La Rochelle	Baf et al. (2007) [20]
Fish Detection	Fish Farming	Abe et al. (2016) [2]
4.3- Open Sea Environments		
Fish Detection	Taiwans coral reefs	Kasavidis and Palazzo (2012) [202]
Fish Detection	Taiwans coral reefs	Spampinato et al. (2014) [376]
Fish Detection	UnderwaterChangeDetection Dataset	Radolko et al. (2017) [323]

Table 4 Intelligent Visual Observation of Natural Environments: An Overview (Part IV)

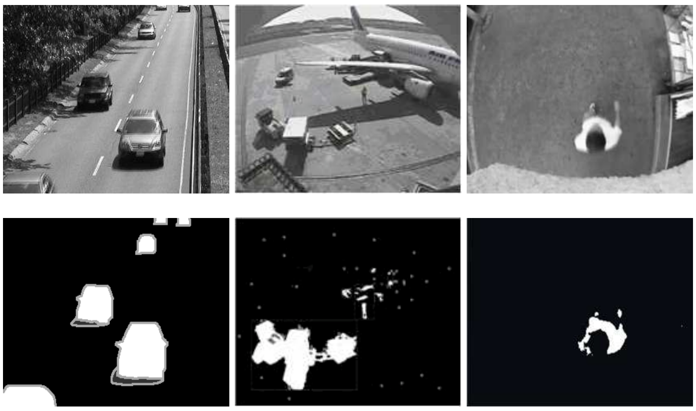

Fig. 2 The first row presents original frames in the following applications: Traffic Surveillance (Goyette et al. [23]), Airport Surveillance (Blauensteiner and Kampel[38]) and Store Surveillance (Avinash et al. [17]). The second row shows the ground truth (GT) or the segmentation results.

Categories	Sub-categories-Aims	Objects of interest	Authors-Dates
Visual Hull Computing	**Image-based Modeling**	**Object**	
	Marker Free	Indoor Scenes	Matusik et al. (2000) [270]
	Optical Motion Capture	**People**	
	Marker Free	Indoor Scenes	Wren et al. (1997) [427]
	Marker Free	Indoor Scenes	Horprasert et al. (1998) [163]
	Marker Free	Indoor Scenes	Horprasert et al. (1999) [164]
	Marker Free	Indoor Scenes	Horprasert et al. (2000) [166]
	Marker Free	Indoor Scenes	Mikic et al. (2002) [276]
	Marker Free	Indoor Scenes	Mikic et al. (2003) [277]
	Marker Free	Indoor Scenes	Chu et al. (2003) [86]
	Marker Free	Indoor Scenes	Carranza et al. (2003) [66]
	Marker Detection	Indoor Scenes	Guerra-Filho (2005) [136]
	Marker Free	Indoor Scenes	Kim et al. (2007) [208]
	Marker Free	Indoor Scenes	Park et al. (2009) [298]
	Marker Free	Indoor Scenes	Habermann et al. (2019) [145]
Human-Machine Interaction (HMI)	**Arts**	**People**	
	Static Camera	Interaction real/virtual	Levin (2006)[234]
	Games	**People**	
	RGB-D Camera	Interaction real/virtual	Microsoft Kinect
	Ludo-Multimedia		
		Fish	
	Aqu@theque Project	Aquarium of La Rochelle	Penciuc et al. (2006)[304]
	Aqu@theque Project	Aquarium of La Rochelle	Baf et al. (2007)[22]
	Aqu@theque Project	Aquarium of La Rochelle	Baf et al. (2007)[20]
Vision-based Hand Gesture Recognition	**Human-Computer Interface (HCI**	**Hands**	
	Augmented Screen	Indoor Scenes	Park and Hyun (2013) [299]
	Hand Detection	Indoor Scenes	Stergiopoulou et al. (2014) [380]
	Behavior Analysis	**Hands**	
	Hand Detection	Indoor Car Scenes	Perrett et al. (2016) [305]
	Sign Language Interpretation and Learning	**Hands**	
	Hand Gesture Segmentation	Indoor/Outdoor Scenes	Elsayed et al. (2015) [112]
	Robotics	**Hands**	
	Control robot movements	Indoor Scenes	Khaled et al. (2015) [205]
Content based Video Coding	**Video Content**	**Objects**	
	Static Camera	MPEG-4	Chien et al. (2012) [83]
	Static Camera	H.264/AVC	Paul et al. (2010)[300]
	Static Camera	H.264/AVC	Paul et al. (2013)[302]
	Static Camera	H.264/AVC	Paul et al. (2013)[301]
	Static Camera	H.264/AVC	Zhang et al. (2010)[456]
	Static Camera	H.264/AVC	Zhang et al. (2012)[458]
	Static Camera	H.264/AVC	Chen et al. (2012)[80]
	Moving Camera	H.264/AVC	Han et al. (2012)[152]
	Static Camera	H.264/AVC	Zhang et al. (2012)[459]
	Static Camera	H.264/AVC	Geng et al. (2012)[129]
	Static Camera	H.264/AVC	Zhang et al. (2014)[455]
	Static Camera	HEVC	Zhao et al. (2014)[467]
	Static Camera	HEVC	Zhang et al. (2014)[457]
	Static Camera	HEVC	Chakraborty et al. (2014)[70]
	Static Camera	HEVC	Chakraborty et al. (2014)[71]
	Static Camera	HEVC	Chakraborty et al. (2017)[72]
	Static Camera	H.264/AVC	Chen et al. (2012) [79]
	Static Camera	H.264/AVC	Guo et al. (2013) [140]
	Static Camera	HEVC	Zhao et al. (2013) [468]

Table 5 Miscellaneous applications: An Overview (Part V)

– **Foreground Objects:** Foreground objects are all road users which have different appearance in terms of color, shape and behavior. Thus, moving foreground objects of interest are **(1)** any kind of moving vehicles such as cars, trucks, motorcycles (motorbikes,...), **(2)** cyclists on bicycles, and **(3)** pedestrians on a pedestrian crossing.

Practically, all these characteristics generate intrinsic specificities and challenges as developed by Song and Tai [373] and Hao et al. [153], and they can be classified as follows:

1. **Background Values:** The intensity of background scene is generally the most frequently recorded one at its pixel position. So, the background intensity can be determined by analyzing the intensity histogram. However, sensing variation and noise from image acquisition devices may result in erroneous estimation and

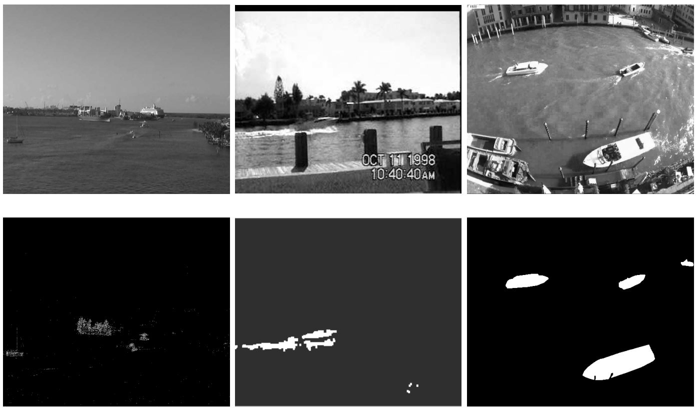

Fig. 3 Maritime surveillance with different types of ships and maritime environments: The first row presents original frames in open sea (Culibrk et al. [97]), fluvial canal in Miami (Socek et al. [372]) and fluvial canal (Bloisi et al. [40]). The second row shows the segmentation results or the ground truth (GT).

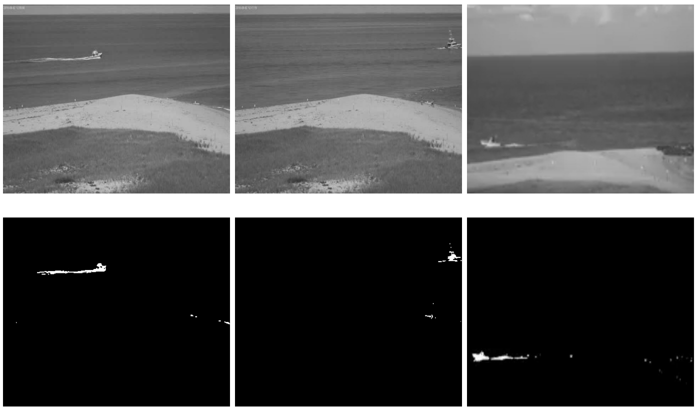

Fig. 4 Coastal surveillance (Cullen et al. [98]): The first row presents original frames. The second row shows the segmentation results.

cause a foreground object to have the maximum intensity frequency in the histogram.

2. **Challenges due the cameras:**
   - In the case of cameras placed on a tall tripod, tripod may moves due the wind [359,358]. In the case of aerial videos, the detection has particular challenges due to high object distance, simultaneous object and camera motion, shadows, or weak contrast [326,391,115,113,114,117,116]. In the case of satellite videos, small size of the objects and the weak contrast are the main challenges [217]. Figure 6 shows different images corresponding to different locations of the camera which induce small or large sizes of the objects of interest.

Fig. 5 Different locations of cameras: Cameras placed on tripod (Images from Alldieck [9]), and cameras placed on the Skybox satellite with an high definition image and its enlarged part (Images from Yang et al. [438]).

- In the case of low quality cameras, the video is low quality, noisy, with compression artifacts, and low frame rate as developed by Toropov et al. [396].

3. **Challenges due the environments:**
   - Traffic surveillance needs to work 24/7 with different weather conditions in day and night scenes. Thus, a background scene dramatically changes over time by the shadow of background objects (e.g., trees) and varying illumination. Figure 7 shows different illumination conditions of the same scene: good illumination conditions, presence of large shadows and presence of large reflections. Figure 8 shows different illumination conditions in different scenes.
   - Shadows may move with the wind in the trees, which may makes the detection result too noisy.

4. **Challenges due the foreground objects:**
   - The moving objects may have similar colors to those of the road and the shadow. Then, the background may be falsely detected as an object or vice-versa.
   - Vehicles may stop occasionally at intersections because of traffic light or control signals. Such kind of transient stops increase the weight of non-background Gaussian and seriously degrade the background estimation quality of a traffic image sequence.
   - In scenarios where vehicles are driving on busy streets, this is even more challenging due to possible merged detections. Figure 9 shows the case of empty traffic, heavy traffic and jam traffic.
   - False detection are caused by vehicle headlights during nighttime as developed by Li et al. [241].

Fig. 6 Various locations: The first column shows orginal images from cameras placed on a tripod at a medium distance in a intersection (Hao et al. [153]) and on a road (Lan et al. [225]). The second column shows images taken from highly mounted cameras on buildings located on the Lankershim Boulevard dataset[3] (Rodriguez and Wohlberg [331]) and located in China (Gao et al. [127]), respectively. The third column shows images taken in aerial location in Germany (Teutsch et al. [391]) and in China (Lin et al. [249]), respectively. The fourth colum shows very high resolution satellite images of Las Vegas (Kopsiaftis and Karantzalos [217]), and Dubai (Yang et al. [438]) obtained from the SkySat-1, respectively.

Fig. 7 Various conditions: (a) Good conditions, (b) Large shadows and (c) Large reflections. Images from Alldieck [9]

5. **Challenges in the implementation:** The computation time needs to be as low as possible because most of the applications require real-time detection.

Table 6 and Table 7 show an overview of the different publications in the field of traffic surveillance with information about the background model, the background maintenance, the foreground detection, the color space and the strategies used by the authors. Authors used uni-modal model or multi-modal model following where the camera is placed. For example, if the camera mainly filmed the road, the most used models are uni-modal models like the median, the histogram and the single Gaussian while if the camera is in a dynamic environment with waving trees, the most models used are multi-modal models like MOG models. For the color space, the authors often used the RGB color space but intensity and YCrCb are also employed to be more robust against illumination changes. For additional strategies, it concerns most of the time shadows detection because it is the most met challenges in this kind of applications.

[3] https://www.fhwa.dot.gov/publications/research/operations/07029/index.cfm

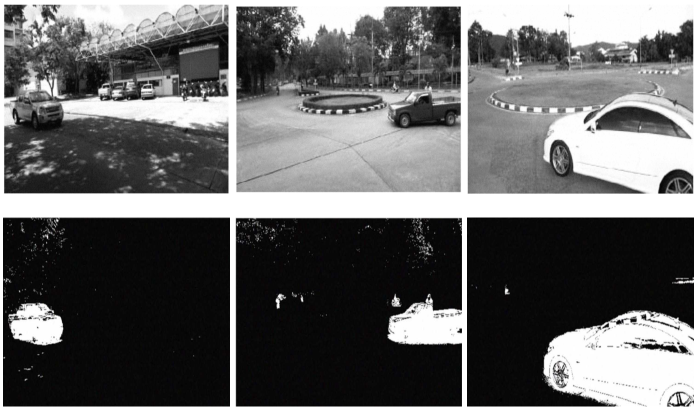

Fig. 8 Different illumination conditions: (a) under shadows, (b) cloudy day and (c) sunny day. The first row shows the original images. The second row shows the foreground detection. Images from Intachak and Kaewapichai [178]

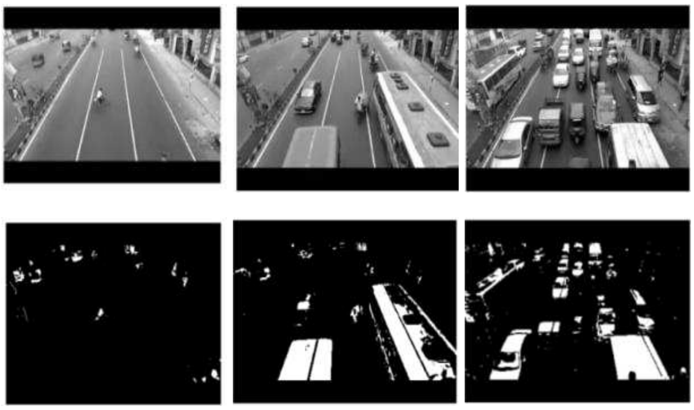

Fig. 9 Different traffic density in urban traffic in India (Muniruzzaman et al. (2016) [285]): empty traffic, heavy traffic and jam traffic. The second row shows the segmentation results.

4.2 Airport surveillance

Airport visual surveillance mainly concerns the area where aircrafts are parked and maintained by specialized ground vehicles such as fueling vehicles and baggage cars as well as tracking of individuals such as workers. The need of visual surveillance is given due to the following reasons: **(1)** an airports apron is a security relevant area, **(2)** it helps to improve transit time, i.e. the time the aircraft is parking on the apron, and **(3)** it helps to minimize costs for the company operating the airport, as personal can be deployed more efficiently, and to minimize latencies for the passengers, as the time needed for accomplishing ground services decreases. Practically, airport surveillance videos present their own characteristics as follows:

1. **Challenges due the environments:** Weather and light changes are very challenging problems. For example, haze is very common and the visibility is then

significantly reduced. In addition, both slow and fast non-uniform illumination changes often occur.

2. **Challenges due the foreground objects:**
   – Ground support vehicles may change their shape in an extended degree during the tracking period, e.g. a baggage car. Strictly rigid motion and object models may not work on tracking those vehicles.
   – Vehicles may build blobs, either with the aircraft or with other vehicles, for a longer period of time, e.g. a fueling vehicle during refueling process.
   – Large airplanes with clear details and small airplanes with blurred outlines can be present on the tarmac.
   – Camouflage in ground airport surveillance often occur because similar gray-white color distributions are shared by airplanes and grounds. The airplanes and airport grounds over the world all have this tone.

Table 8 shows an overview of the different publications in the field of airport surveillance with information about the background model, the background maintenance, the foreground detection, the color space and the strategies used by the authors. We can see that only the median or the single Gaussian are employed for the background model as the tarmac is an uni-modal background. The color space used is the RGB color space for all the works.

4.3 Maritime surveillance

Maritime surveillance can be achieved in visible spectrum [40,453] or IR spectrum [254,476]. The idea is to count, to track and to recognize boats in fluvial canals, in river, or in open sea. For fluvial canals of Miami, Socek et al. [372] proposed a hybrid foreground detection approach which combined a Bayesian background subtraction framework with an image color segmentation technique to improve accuracy. In an other work, Bloisi et al. [40] used the Independent Multimodal Background Subtraction (IMBS[39]) which has been designed for dealing with highly dynamic scenarios characterized by non-regular and high frequency noise, such as water background. IMBS is a per-pixel, non-recursive, and non-predictive model. In the case of open sea environments, Culibrk et al. [97] used the original MOG [379] implemented on General Regression Neural Network (GRNN) to detect cargos. In an other work, Zhang et al. [453] used the median model to detect ships to track them. To detect foreign floating objects, Borghgraef et al. [42] employed the improved MOG model called Zivkovic-Heijden GMM [481]. To detect many kinds of vessels, Szpak and Tapamo used the single Gaussian model [382], and can tracked in their experiments jet-skis, sailboats, rigid-hulled inflatable boats, tankers, ferries and patrol boats. In infrared video, Liu et al. [254] employed a modified histogram model. To detect sailboats, Sobral et al. [369] developed a double constrained RPCA based on saliency detection. In a comparison work, Tran and Le [398] compared the original MOG and ViBe to detect boats, and they concluded that ViBe is a suitable algorithm for detecting different kinds of boats such as cargo ships, fishing boats, cruise ships, and canoes which have very different appearance in terms of size, shape, texture and structure. To automatically detects and tracks ships (intruders) in the case of cage aquaculture, Hu et al. [167] used an approximated median based on AM [271] with a wave ripple removal. In Table 8, we show an overview of the different publications in the field of maritime surveillance with information about the background model, the background maintenance, the foreground detection, the color space and the strategies used by the authors. We can remark that the authors prefer to use multi-modal background models (MOG, Zivkovic-Heijden GMM, ...) in this context because water presents a dynamic aspect. For the color space, RGB is often used in diurnal conditions as well as infrared for night conditions. Most of the time, additional strategies are employed like morphological processing and saliency detection to deal with the false positive detections due to the movement of the water.

4.4 Coastal Surveillance

Cullen et al. [99,98] used the behavior subtraction model developed by Jodoin et al. [195] to detect salient events in coastal environments[4] which can be interesting for many organizations to learn about the wildlife, land erosion, impact of humans on the environment, etc. For example, biologists interested in marine mammal protection wish to know whether humans have come too close to seals on a beach. US Fish and Wildlife Service wish to know how many people and cars have been on the beach each day, and whether they have disturbed the fragile sand dunes. Practically, Cullen et al. [99,98] detected boats, motor vehicles and people appearing close to the shoreline. In Table 8, the reader can find an overview of the different publications in the field of coastal surveillance with information about the background model, the background maintenance, the foreground detection, the color space and the strategies used by the authors.

4.5 Store surveillance

The interest is more and more coming across detection of human in stores because marketing researchers in academia and industry seek for tools to aid their decision making. Unlike other types of sensors, vision presents an ability to observe customer experience without separating it from the environment. By tracking the path traveled by the customer along the store, important pieces of information, such as customer dwell time and product interaction statistics can be collected. One of the most important customer statistics is the information about the shopper groups such as in Leykin and Tuceryan [237,236,235], Avinash et al. [17] and Nogueira et al. [294]. Other applications concern the amount product estimation [159,160] for improving on-shelf availability. The most common method to accurately understand the product amount on the shelves employes Radio Frequency Identification (RFID) tags on each product in the store [479]. In practice, RFID allows detection of products in real-time allowing the shopper to accurately know the number of remaining products on the shelves. But, RFID tags are still expensive and a large amount of time is needed for attaching them on all products in the store. To address these issues, Higa and Iwamoto

[4] http://vip.bu.edu/projects/vsns/coastal-surveillance/

[159, 160] proposed a low-cost solution based on videos using background subtraction applied to videos taken by a fixed camera attached on the ceiling. Table 8 shows an overview of the different publications in the field of store surveillance with information about the background model, the background maintenance, the foreground detection, the color space and the strategies used by the authors. Either a uni-modal model and multi-modal model in RGB color space are used because the videos are generally filmed in indoor scenes.

4.6 Military surveillance

Detection of moving objects in military surveillance is often referred as target detection in literature. Most of the time, it used specific sensors like infrared cameras [25] and Synthetic-aperture radar (SAR) imaging [360]. In practice, the goal is to detect persons and/or vehicles in challenging environments (forest, ...) and challenging conditions (night scenes, ...). In literature, several authors employed background subtraction methods for target detection either in infrared or SAR imaging as follows:

- **Infrared imaging:** For example, El Baf et al. [25, 21] used a median background subtraction algorithm with a Choquet integral approach to classify objects as background or foreground in infrared video whilst El Baf et al. [29] used a Type-2 Fuzzy MOG (T2-FMOG) model. Figure 10 shows results obtained by El Baf et al. [25]on videos from the OCTBVS datasets[5].
- **SAR imaging:** A sub-aperture image can be considered as combination of background image that contains clutter and foreground image that contains moving target. For clutter, its scattered field is varying slowly in limit angular sector. For moving target, its image position is changing along azimuth viewing angle because of circular flight. Thus, target signature is moving in consecutive sub-aperture images. Thus, value of certain pixel would have sudden change when target signature is moving onto and leaving it. Based on this idea, Shen et al. [360] employed a temporal median to obtain the background image. A log-ratio operator is then integrated into the process. The operator can be defined as the logarithm of the ratio of two images. This is equivalent to subtract two logarithm images. Finally, the moving targets are detected by applying Constant False Alarm Rate (CFAR) detector. Figure 11 shows results obtained by Shen et al. [360].

Because in this field videos are very confidential, authors present results on publicly civil datasets for publication.

[5] http://vcipl-okstate.org/pbvs/bench/

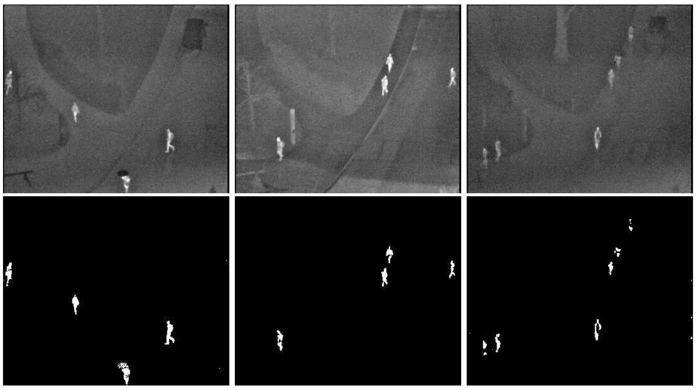

Fig. 10 Infrared imaging: The first row shows original images. The second row shows segmentation results. Images from El Baf et al. [25].

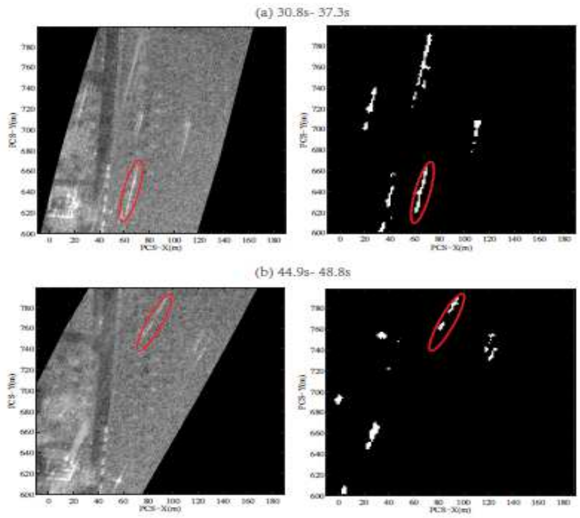

Fig. 11 SAR imaging. The first column shows input subaperture images. The right column are segmentation results in two intervals 30.837.3s and 44.948.8s. Boxes are unimportant detected moving objects. Images from Shen et al. [360].

Human Activities	Type	Background model	Background Maintenance	Foreground Detection	Color Space	Strategies
Traffic Surveillance	**1) Road Surveillance**					
	1.1) Road/Highways Traffic					
	Zheng et al. (2006) [473]	Histogram mode	Blind Maintenance	Difference	RGB	Aggregation for small range
	Batista et al. (2006) [32]	Average	Running Average	Minimum	RGB	Double Backgrounds
	Monteiro et al. (2008) [284]	Median	Sliding Median	Minimum	RGB	Double Backgrounds-Shadow/Highlight Removal
	Monteiro et al. (2008) [283]	Median	Sliding Median	Minimum	RGB	Double Backgrounds-Shadow/Highlight Removal
	Monteiro et al. (2008) [235]	Codebook [210]	Idem Codebook	Idem Codebook	RGB	Shadow/Highlight Detection [165]
	Monteiro (2009) [281]	Sliding Median		Minimum	RGB	Double Backgrounds-Shadow/Highlight Removal
	Hao et al. (2013) [154]	KDE [358]	Idem KDE	Idem KDE	Joint Domain-Range Features [358]	Foreground Model
	Ling et al. (2014) [250]	Dual Layer Approach [250]				
	Sawalakhe and Metkar (2014) [346]	Spatio-Temporal BS/FD (ST-BSFD)	-	AND	RGB	-
	Huang and Chen (2013) [173]	Cerebellar-Model-Articulation-Controller (CMAC) [173]	-	Threshold	YCrCb	Block
	Chen and Huang (2014) [78]	PCA-based RBF Network [78]	Selective Maintenance	Euclidean distance	YCrCb	Block
	Lee et al (2015) [229]	GDSM with Running Average [229]	Selective Maintenance	AND	RGB	Shadow Detection [95]
	Aqel et al. (2015) [12]	SG on Intensity Transition	Idem SG	Idem SG	RGB	-
	Aqel et al. (2016) [11]	SG on Intensity Transition	Idem SG	Idem SG	RGB	Shadow Detection
	Wang et al. (2016) [413]	Median				
	Dey and Praveen (2016) [102]	GDSM with background subtraction [146]	Yes	AND	RGB	Post-processing
	Hadiuzzaman et al. (2017) [148]	Median	Yes	Idem Median	Intensity	Shadow Detection
	1.2) Urban Traffic					
	Conventional camera					
	Hwang et al. (2009) [177]	MOG-ALR [177]	Adaptive Learning Rate	Idem MOG	RGB	-
	Intachak and Kaewapichai (2011) [178]	Mean (Clean images)	Selective Maintenance	Idem Mean	RGB	Illumination Adjustment
	Milla et al. (2013) [278]	$\Sigma-\Delta$ filter	$\Sigma-\Delta$ filter	$\Sigma-\Delta$ filter	Intensity	Short-term/Long-term backgrounds
	Toropov et al. (2015) [396]	MOG [379]	Idem MOG	Idem MOG	Color	Brightness Adjustment
	Zhang et al. (2016) [463]	GMMCM [463]	Idem MOG [379]	Confidence Period	Intensity	Classification of traffic density
	Highly-mounted camera					
	Gao et al. (2014) [127]	SG [427]	Idem SG	Idem SG	YCrCb	Shadow detection [316]
	Quesada and Rodriguez (2016) [322]	incPCP [335]	Idem incPCP	Idem incPCP	Intensity	-
	Headlight Removal					
	Li et al. (2016) [241]	GMM [379]	Idem GMM	Idem GMM	RGB	Headlight/Shadow Removal
	Intersection					
	Ding et al. (2012) [107]	CPS based GMM [107]	Idem GMM	Idem GMM	RGB	Cyber Physical System
	Ding et al. (2012) [107]	CPS based FGD [239]	Idem FGD	Idem FGD	RGB	Cyber Physical System
	Alldieck (2015) [9]	Zivkovic-Heijden GMM [481]	Idem GMM	Idem GMM	RGB/IR	Multimodal cameras
	Mendoca et Oliveira (2015) [274]	Context supported ROad iNformation (CRON) [274]	Selective Maintenance	Idem AM	RGB	
	Li et al. (2016) [238]	incPCP [330]	Idem incPCP	Idem incPCP	RGB	-
	Obstacle Detection					
	Lan et al. (2015) [225]	SUOG [225]	Selective Maintenance	Idem GMM	RGB	Obstacle Detection Model

Table 6 Intelligent Visual Surveillance of Human Activities: An Overview (Part 1). ”-” indicated that the corresponding step used is not indicated in the paper.

Human Activities	Type	Background model	Background Maintenance	Foreground Detection	Color Space	Strategies
	2)Vehicle Counting					
	Unzueta et al. (2012) [402]	Multicue approach [402]				
	Virginas-Tar et al. (2014) [406]	MOG-EM [199]	Idem MOG-EM	Idem MOG-EM		Shadow Detection [165]
	3) Vehicle Detection					
	3.1) Conventional Video					
	Wang and Song (2011) [409]	GMM with Spatial Correlation Method [409]	Idem GMM	Idem GMM	HSV	Spatial Correlation Method
	Hadi et al. (2014) [147]	Histogram mode	Blind Maintenance	Absolute Difference	RGB	Morphological Processing
	Hadi et al. (2017) [146]	Histogram mode	Blind Maintenance	Absolute Difference	RGB	Morphological Processing
	3.2) Aerial Video					
	Lin et al. (2009) [249]	Two CFD	-	-	RGB	Translation
	Reilly (2012) [326]	Median	Idem Median	Idem Median	Intensity	-
	Teutsch et al. (2014) [391]	Independent Motion Detection [391]				
	3.3) Satellite Video					
	Kopsiaftis and Karantzalos (2015) [217]	Mean	Idem Mean	Idem Mean	Intensity	-
	Yang et al. (2016) [438]	Local Saliency based Background Model based on ViBe (LS-ViBe) [438]	Idem ViBe	Idem ViBe	Intensity	-
	4) Illegally Parked Vehicles					
	Lee et al. (2007) [230]	Median	Selective Maintenance	Difference	RGB	Morphological Processing
	Zhao et al. (2013)[470]	Average	Running Average	Difference	HIS	Morphological Processing-Salient Object Detection
	Saker et al. (2015)[339]	GMM-PUC [243]	Idem GMM	Idem GMM	RGB	Detection of Stationary Object
	Chunyang et al. (2015) [87]	MOG [379]	Idem MOG	Idem MOG	RGB	Morphological Processing
	Wahyono et al. (2015) [408]	Running Average	Selective Maintenance	Difference	RGB	Dual background models
	5) Vacant parking area					
	Postigo et al. (2015) [312]	MOG-EM [199]	Idem MOG-EM	Idem MOG-EM	RGB	Transience Map
	Neuhausen (2015) [290]	SG	Selective Maintenance	Choquet Integral [256]	YCrCb-ULBP [445]	Adaptive weight on Illumination Normalization
	6) Motorcycle (Motorbike) Detection					
	Silva et al. (2013) [365]	Zivkovic-Heijden GMM [481]	Idem GMM	Idem GMM	Intensity	-
	Waranusast et al. (2013) [418]	Zivkovic-Heijden GMM [481]	Idem GMM	Idem GMM	RGB	Morphological Operators
	Dahiya et al. (2016) [100]	Zivkovic-Heijden GMM [481]	Idem GMM	Idem GMM	Intensity	Detection Bike-riders

Table 7 Intelligent Visual Surveillance of Human Activities: An Overview (Part 2). ”-” indicated that the corresponding step used is not indicated in the paper.

Human Activities	Type	Background model	Background Maintenance	Foreground Detection	Color Space	Strategies
Airport Surveillance	Blauensteiner and Kampel (2004) [38]	Median	Idem Median	Idem Median	RGB	-
	Aguilera et al. (2005) [5]	Single Gaussian	Idem SG	Idem SG	RGB	-
	Thirde et al. (2006) [392]	Single Gaussian	Idem SG	Idem SG	RGB	-
	Aguilera et al. (2006) [4]	Single Gaussian	Idem SG	Idem SG	RGB	-
Maritime Surveillance	**1) Fluvial canals environment**					
	Socek et al. (2005) [372]	Two CFD	-	Bayesian Decision [236]	RGB	Color Segmentation
	Bloisi et al. (2014) [40]	IMBS [39]				
	2) River environment					
	Zheng et al. (2013) [474]	LBP Histogram [158]	LBP Histogram [158]	LBP Histogram [158]	RGB	-
	Lu et al. (2016) [255]	K-SVD dictionary [255]	Yes	Yes	RGB	-
	Mei et al. (2017) [273]	EAdaDGS [273]	Improved mechanism	Idem AdaDGS [171]	RGB	Multi-resolution
	3) Open sea environment					
	Culibrk et al. (2006) [97]	MOG-GNN [379]	Idem MOG	Idem MOG	Intensity	-
	Zhang et al. (2009) [453]	Median				
	Borghgraef et al. (2010) [42]	Zivkovic-Heijden GMM [481]	Idem Zivkovic-Heijden GMM	Idem Zivkovic-Heijden GMM	IR	-
	Arshad et al. (2010) [13]	-	-	-	RGB	Morphological Processing
	Arshad et al. (2011) [14]	-	-	-	RGB	Morphological Processing
	Szpak and Tapamo (2011) [382]	Single Gaussian [427]	Idem SG	Idem SG	Intensity	Spatial Smoothness
	Hu et al. (2011) [167]	Modified AM [167]	Idem Modified AM	Idem Modified AM	RGB	Fast 4-Connected Component Labeling
	Saghafi et al. (2012) [338]	Modified ViBe [338]	Modified ViBe	Modified ViBe	RGB	Backwash Cancellation Algorithm
	Xie et al. (2012) [429]	Three CFD	Selective Maintenance	AND	Intensity	
	Zhang et al. (2012) [450]	PCA [295]/GMM [379]	Idem PCA [295]/GMM [379]	Idem PCA [295]/GMM [379]	Intensity	-
	Arshad et al. (2014) [15]	-	-	-	RGB	Morphological Processing
	Liu et al. (2014) [254]	Modified Histogram [254]	Idem Histogram	Idem Histogram	IR	Adaptive Row Mean Filter
	Sobral et al. (2015) [369]	Double Constrained RPCA [369]	Idem RPCA	Idem RPCA	RGB	Saliency Detection
	Tran and Le (2016) [398]	ViBe [30]	Idem ViBe	Idem ViBe	RGB	Saliency Detection
	Chen et al. (2018) [82]	MoG [379]	Idem MoG	Idem MoG	RGB	-
Store Surveillance	Leykin and Tuceryan (2007) [237]	Codebook [210]	Idem Codebook	Idem Codebook	RGB	-
	Leykin and Tuceryan (2005) [236]	Codebook [210]	Idem Codebook	Idem Codebook	RGB	-
	Leykin and Tuceryan (2005) [235]	Codebook [210]	Idem Codebook	Idem Codebook	RGB	-
	Avinash et al. (2012) [17]	Single Gaussian [427]	Idem SG	Idem SG	RGB	-
	Higa and Iwamoto (2017) [159]	Zivkovic-Heijden GMM [481]	Idem Zivkovic-Heijden GMM	Idem Zivkovic-Heijden GMM	RGB	-
	Higa and Iwamoto (2018) [160]	Zivkovic-Heijden GMM [481]	Idem Zivkovic-Heijden GMM	Idem Zivkovic-Heijden GMM	RGB	-
Coastal Surveillance	Cullen et al. (2012) [99]	Behavior Subtraction [195]	Idem BS	Idem BS	RGB	-
	Cullen (2012) [98]	Behavior Subtraction [195]	Idem BS	Idem BS	RGB	-
Swimming Pools Surveillance	**1) Online videos**					
	1.1) Top view videos					
	Eng et al. (2003) [118]	Block-based median [118]	Sliding Mean	Block-based Difference	CIELa*b*	Partial Occlusion Handling
	Eng et al. (2004) [119]	Region-based single multivariate Gaussian [119]	Idem SG	Region based SG	CIELa*b*	Handling Specular Reflection
	Chan (2011) [73]	SG/optical flow [73]	Idem SG	Designed Distance	RGB	-
	Chan (2014) [74]	MOG/optical flow [74]	Idem MOG	Idem MOG	RGB	-
	1.2) Underwater videos					
	Fei et al. (2009) [124]	MOG	Idem MOG			Shadow Removal
	Lei and Zhao (2010) [233]	Kalman filter [328]	Idem Kalman filter	Idem Kalman filter	RGB	
	Zhang et al. (2015) [449]	-	-	-	RGB	Inter-frame based denoising
	2) Archived videos					
	Sha et al. (2014) [348]	-	-	-	RGB	-
	3) Private swimming pools					
	Peixoto et al. (2012) [303]	Mean/Two CFD [303]	Selective Maintenance	Mean Distance/Two CFD Distance	HSV	

Table 8 Intelligent Visual Surveillance of Human Activities: An Overview (Part 3). ”-” indicated that the background model used is not indicated in the paper.

5 Intelligent Visual Observation of Animals and Insects

Surveillance with fixed cameras can also concern census and activities of animals in open protected areas (river, ocean, forest, ...), as well as ethology in closed areas (zoo, cages, ...). In these applications, the detection done by background subtraction is followed by a tracking phase or a recognition phase [34,444,424]. In practice, the objects of interest are then animals such as birds [214,215,354,105,106], fish [374, 377], honeybees [213,63,211,19], hinds [206,131], squirrels [91,90], mice [327,16] or pigs [399]. In practice, animals live and evolve in different environments that can be classified in three main categories: 1) Natural environments like forest, river and ocean, 2) study's environment like tank for fish and cages for mice, 3) farm environments for surveillance of pigs [271] and livestock [466,293,10,141]. Thus, videos for visual observation of animals and insects present their own characteristics due the intrinsic appearance and behavior of the detected animals or insects, and the environments in which they are filmed. Figure 12 and Figure 13 show sample of detection on **(1)** birds, honeybees and fish and **(2)** hinds and spiders. Figure 14 show detection of hinds, tapirs and dolphins with RPCA models. Therefore, there are generally no a priori on the shape and the color of the objects when background subtraction is applied. In addition of these different intrinsic appearances in terms of size, shape, color and texture, and behavior in terms of velocity, videos which contains animals or insects present challenging characteristics due the background in natural environments which are developed in Section 6. However, it is important to notice that only the fact that biological foreground objects occur in dynamic outdoor scenes is very challenging. Finding and differentiating a robotic object in such scenes would be equally quasi difficult. The additional difficulty inherent to biological foreground objects is that animals/insects present sometimes characteristics that allow them to use a camouflage strategy adding more difficulties to detect them. Events such as VAIB (Visual observation and analysis of animal and insect behavior) in conjunction with ICPR addressed the problem of the detection of animals and insects employing real datasets.

5.1 Visual Observation of Animals

Animals live in different kind of environments (i.e air, water and ground) and their activities can be monitored to better know them. The most monitored species are the following ones:

1. **Birds Surveillance:** Detection of birds is a crucial problem for multiple applications such as aviation safety, avian protection, and ecological science of migrant bird species. There are three kinds of bird observations: **(1)** observations at human made feeder stations [214,215], **(2)** observation at natural nesting stations [132, 105,106], and **(3)** observation in the air with camera looking at the roof-top of a building or recorded footages on lakes [354]. In the first case, as developed by Ko et al. [214], birds at a feeder station present a larger per-pixel variance due to changes in the background generated by the presence of a bird. Rapid background

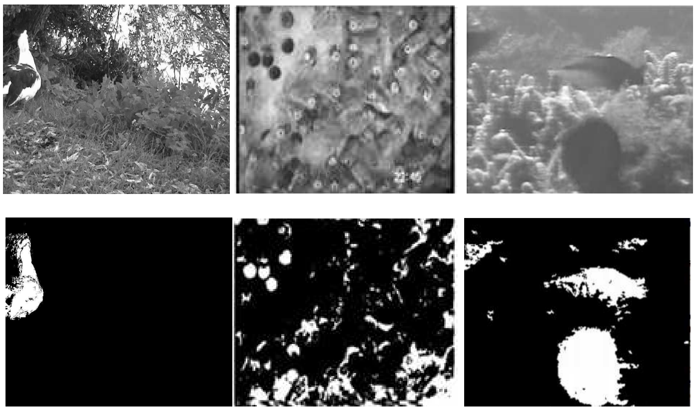

Fig. 12 The first row presents original frames in animals surveillance: Seabirds (Dickinson et al. [105]), Honeybees (Knauer et al. [213]) and Fish (Spampinato et al. [374]). The second row shows the segmentation results.

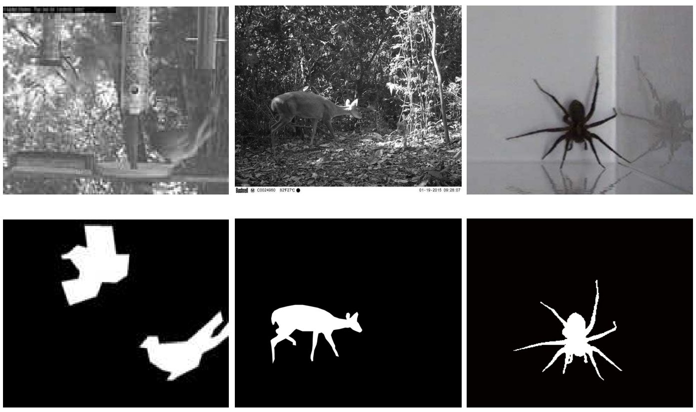

Fig. 13 The first row presents original frames in animals surveillance: Birds (Ko et al. [214]), Hinds (Giraldo-Zuluaga et al. [131]), and Spiders (Iwatani et al. [179]). The second row shows the corresponding ground truth (GT)

adaptation fails because birds, when present, are often moving less than the background and often end up being incorporated into it. To address this challenge, Ko et al. [214,215] designed a background subtraction based on distributions. In the second case, Goehner et al. [132] compared three background models (MOG [313], ViBe [30], PBAS [161]) to detect events of interest within uncontrolled outdoor avian nesting video for the Wildlife@Home[6] project. The video are taken in environments which require background models that can handle quick correction of camera lighting problems while still being sensitive enough to detect the motion of a small to medium sized animal with cryptic coloration. To address these problems, Goehner et al. [132] added modifications to both the ViBe and PBAS algorithms by making these algorithms second-frame-ready and by adding a morphological opening and closing filter to quickly adjust to the noise present

Fig. 14 The first row presents original frames in animals surveillance: Hinds (Khorrami et al. [206]), Tapir (Khorrami et al. [206]) and Dolphin (Karnowski et al. [201]). The second row shows the corresponding the low rank (background image). The third row the sparse component (foreground) obtained by RPCA [65].

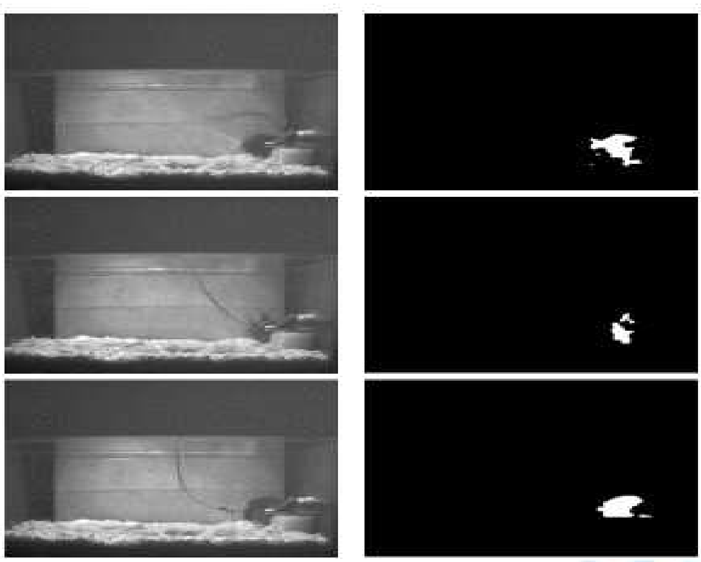

Fig. 15 The first column presents original frames. The second column shows the segmentation results.

in the videos. Moreover, Goehner et al. [132] also added a convex hull around the connected foreground regions to help counter cryptic coloration.

2. **Fish Surveillance:** Automated video analysis of fish is required in several types of applications: **(1)** detection of fish for recognition of species in a tank as in El Baf et al. [22,304] and in open sea as in Wang et al. [410], **(2)** detection and tracking of fish for counting in a tank in the case of industrial aquaculture as in Abe et al. [2], **(3)** detection and tracking of fish in open sea to study their behaviors in different weather conditions as in Spampinato et al. [374,377,375, 376], and **(4)** in fish catch tracking and size measurement as in Huang et al. [174]. In all these applications, the appearances of fish are variable as they are non-rigid and deformable objects, and therefore it makes their identification very complex. In 2007, El Baf et al. [22,304] studied different models (SG [427], MOG [379] and KDE [110]) for the Aqu@theque project, and concluded that SG and MOG offer good performance in time consuming and memory requirement, and that KDE is too slow for the application and requires too much memory. Because MOG gives better results than SG, MOG is revealed to be the most suitable for fish detection in a tank. In the EcoGrid project in 2008, Sampinato et al. [374, 375] used both a sliding Average and the Zivkovic-Heijden GMM [481] for fish detection in submarine video. In 2014, Sampinato et al. [377] developed a specific model called textons based KDE model to address the problem that the movement of fish is fairly erratic with frequent direction changes. For videos taken by Remotely Operated Vehicles (ROVs) or Autonomous Underwater Vehicles (AUVs), Liu et al. [252] proposed to combine with a logical "AND" the foreground detection obtained by the running average and the three CFD. In an other work, Huang et al. [174] proposed a live tracking of rail-based fish catching by combining background subtraction and motion trajectories techniques in highly noisy sea surface environment. First, the foreground masks are obtained using SuBSENSE [378]. Then, the fish are tracked and separated from noise based on their trajectories. In 2017, Huang et al. [175] detected moving deep-sea fishes with the MOG model and tracked them with an algorithm combining Camshift with Kalman filtering. By analyzing the directions and trails of targets, both distance and velocity are determined.

3. **Marine Mammals Surveillance:** Detecting and tracking social marine mammals such as bottle nose dolphins, allow researchers such as ethologists and ecologists to explain their social dynamics, predict their behavior, and measure the impact of human interference. Practically, multi-camera systems give an opportunity to record the behavior of captive dolphins with a massive dataset from which long term statistics can be extracted. In this context, Karnowski et al. [201] used a background subtraction method to detect dolphins over time, and to visualize the paths by which dolphins regularly traverse their environment.

[6] http://csgrid.org/csg/wildlife/

4. **Lizards Surveillance:** To census of endangered lizard species, it is important to automatically identify them from videos. In this context, Nguwi et al. [291] used a background subtraction method to detect lizards in video, followed by experiments comparing thresholding values and methods.

5. **Mice Surveillance:** Mice surveillance concern mice engaging in social behavior [61, 162]. Most of the time experts reviewed frame-by-frame the behavior but it is time consuming. In practice, the main challenge in this kind of surveillance is the big amount of videos and thus it requires fast algorithms to detect and recognize behaviors. In this context, Rezaei and Ostadabbas [327, 16] provided a fast Robust Matrix Completion (fRMC) algorithm for in-cage mice detection using the Caltech resident intruder mice dataset [61].

6. **Pigs Surveillance:** Behavior analysis of livestock animals such as pigs under farm conditions is an important task to allow better management and climate regulation to improve the life of the animals. There are three main problems in farrowing pens as developed by Tu et al. [399]:
   (a) **Dynamic background objects:** The nesting materials in the farrowing pen are often moved around because of movements of the sow and piglets. The nesting materials can be detected as moving backgrounds.
   (b) **Light changes:** Lights are often switched on and off in the pig house. In the worst case, the whole segmented image often appears as foreground in most statistical models when the strong global illumination change suddenly occurs.
   (c) **Motionless foreground objects:** Pigs and sows often sleep over a long period. In this case, a foreground object that becomes motionless can be incorporated in the background.

   First, Guo et al. [142] used a Prediction mechanism-Mixture of Gaussians algorithm called PM-MOG for detection of group-housed pigs in overhead views. In an other work, Tu et al. [399] employed a combination of modified MOG model and the Dyadic Wavelet Transform (DWT) in gray scale videos. This algorithm accurately extracts the shapes of a sow under complex environments. In a further work, Tu et al. [400] proposed an illumination and reflectance estimation by using an Homomorphic Wavelet Filter (HWF) and a Wavelet Quotient Image (WQI) model based on DWT. Based on this illumination and reflectance estimation, Tu et al. [400] used the CFD algorithm of Li and Leung [240] which combined intensity and texture differences to detect sows in gray scale video.

7. **Hinds Surveillance:** There are three main problems to detect hinds as developed by Khorrami et al. [206]:
   (a) **Camouflage:** Hinds may blend with the forest background by necessity
   (b) **Motionless foreground objects:** Hinds may sleep over a long period. In this case, hinds can be incorporated in the background.
   (c) **Rapid Motion:** Hinds can quickly move to escape a predator.

In camera-trap sequences, Giraldo-Zuluaga et al. [131, 130] used a multi-layer RPCA to detect hinds in forest in Colombia. Experimental results [131] against other RPCA models show the robustness of the multi-layer RPCA model in presence of challenges such as illumination changes.

5.2 Visual Observation of Insects

All insects have significant influence on the environment and the most monitored species are the following ones:

1. **Honeybees Surveillance:** Honeybees are generally filmed at the entrance of a hive to track and count different goal as follows: **(1)** detection of external honeybee parasites as in Knauer et al. [213], **(2)** monitoring arrivals and departures at the hive entrance as in Campbell et al. [63], **(3)** study of their sociability as in Kimura et al. [211], and **(4)** remote pollination monitoring as in Babic et al. [19]. There are several reasons why honeybees detection is a difficult computer vision problem as developed by Campbell et al. [63] and Babic et al. [19].
   (a) Honeybees are small. In a typical image acquired from a hive-mounted camera a single bee occupies only a very small portion of the image (approximately 6×14 pixels). Honeybee detection can be easier with higher-resolution cameras or with multiple cameras placed closer to the hive entrance, but only at a substantial increase in cost as well as physical and computational complexity, limiting utility in practical setting.
   (b) Honeybees are fast targets. Indeed, honeybees cover a significant distance between frames. This movement complicates frame-to-frame matching as worker bees from a hive are virtually identical in appearance.
   (c) Honeybees present motion which appears to be chaotic. Indeed, honeybees transition quickly between loitering, crawling, and flying modes of movement and change directions unpredictably; this makes it impossible to track them using one unimodal motion model.
   (d) Bee hives are in outdoor scenes where lighting conditions vary significantly with the time of day, season and weather. Moreover, shadows are cast by the camera enclosure, moving bees, and moving foliage overhead. Even if it is possible to have clear lighting in the hive entry area, it demands onerous hive placement constraints vis-a-vis trees, buildings, and compass points. Artificial lighting is difficult to place in the field and could affect honeybee behavior.
   (e) The scene in front of a hive is often cluttered with honeybees grouping, occluding and/or overlapping each other. Thus, the moving objects detection aware of the fact that the detected moving object can sometimes contain more than one honeybee.
   (f) In the case of the pollen assessment [19], it is needed to at least obtain additional information that is whether the group of honeybees has a pollen load or not, when it is not possible to segment individual honeybees.

   In 2016, Pilipovic et al. [307] studied different background models (frame differencing, median model [271], MOG model [379] and Kalman filter [200]) in this field, and concluded that MOG is best suited for detection honeybees in hive entrance video. In 2017, Yang [437] confirmed the adequacy of MOG by using a modified version [414].

2. **Spiders Surveillance:** In 2016, Iwatani et al. [179] proposed to design a hunting robot that mimics the spiders hunting locomotion. To estimate the two-dimensional position and direction of a wolf spider in an observation box from video imaging, a simple background subtraction method is used because the environment is controlled. Practically, a gray scale image without the spider is taken in advance. Then, a pixel in each captured image is selected as a spider component candidate, if the difference between the background image and the captured image in grayscale is larger than a given threshold.

3. **Miscellaneous Insects Surveillance:** In 2019, Geus et al. [101] employed a running average method to detect insect in videos recorded within a container with bulk maize. For the foreground detection, Geus et al. [101] used local and global thresholds.

6 Intelligent Visual Observation of Natural Environments

The aim is to detect foreign objects in natural environments such as forest, ocean and river to protect the biodiversity in terms of fauna and flora. For example, foreign objects in river and ocean can be floating bottles [475], floating wood [6] [7] or mines [42]. There is a general used tool to detect motion in ecological environments called MotionMeerkat [421]. MotionMeerkat[7] alleviates the process of video stream data analysis by extracting frames with motion from a video. MotionMeerkat can either use Running Gaussian Average or MOG as background model. MotionMeerkat is successful in many ecological environments but is still subject to problems such as rapid lighting changes, and camouflage. In a further work, Weinstein [423] proposed to employ a background modeling based on convolutional neural networks and developed the software DeepMeerkat[8] for biodiversity detection. For marine environment, there is a open source framework called Video and Image Analytics for a Marine Environment (VIAME) but it does currently no contain background subtraction algorithms. Thus, advanced and designed background models are needed in natural environments. Practically, natural environments such as the forest canopy, river and ocean present an extreme challenge because the foreground objects may blend with the background by necessity. Furthermore, the background itself mainly changes following its characteristics as described in the following sections.

[7] http://benweinstein.weebly.com/motionmeerkat.html

[8] http://benweinstein.weebly.com/deepmeerkat.html

Animal and Insect Behaviors	Type	Background model	Background Maintenance	Foreground Detection	Color Space	Strategies
Birds Surveillance	**Birds**					
	1) Feeder stations					
	Ko et al. (2008) [214]	KDE [359]	Blind Maintenance	Bhattacharyya Distance	RGB	Temporal Consistency
	Ko et al. (2010) [215]	Set of Warping Layer [215]	Blind Maintenance	Bhattacharyya distance	UYV	-
	2) Birds in air					
	Shakeri and Zhang (2012) [354]	Zivkovic-Heijden GMM [481]	Idem GMM	Idem GMM	RGB	Correspondence Component based on Point-Tracking
	Nazir et al. (2017) [289]	OpenCV Background Subtraction (WiseEye)	-	-	-	-
	3) Avian nesting					
	Goehner et al. (2015) [132]	MOG [313], ViBe [30], PBAS [161]	Idem MOG/PBAS/ViBe	Idem MOG/PBAS/ViBe	RGB	Morphological Processing
	Dickinson et al. (2008) [105]	MOG [105]	Idem MOG	Idem MOG	RGB	Spatially Coherent Segmentation [104]
	Dickinson et al. (2010) [106]	MOG [105]	Idem MOG	Idem MOG	RGB	Spatially Coherent Segmentation [104]
Fish Surveillance	**1) Tank environment**					
	1.1) Species Identification					
	Penciuc et al. (2006) [20]	MOG [379]	Idem MOG	Idem MOG	RGB	-
	El Baf et al. (2007) [22]	MOG [379]	Idem MOG	Idem MOG	RGB	-
	El Baf et al. (2007) [304]	MOG [379]	Idem MOG	Idem MOG	RGB	-
	1.2) Industrial Aquaculture					
	Pinkiewicz (2012) [308]	Average/Median	Idem Average/Idem Median	Idem Average/Idem Median	RGB	-
	Abe et al. (2017) [2]	Average	Idem average	Idem average	RGB	-
	Zhou et al. (2017) [477]	Average	Idem average	Idem average	Near Infrared	-
	2) Open sea environment					
	Spampinato et al. (2008) [374]	Sliding Average/Zivkovic-Heijden GMM [481]	-	-	-	AND
	Spampinato et al. (2010) [375]	Sliding Average/Zivkovic-Heijden GMM [481]	-	-	-	AND
	Spampinato et al. (2014) [376]	GMM [379], APMM [122], IM [309], Wave-Back [311]	-	-	-	Fish Detector
	Spampinato et al. (2014) [377]	Textons based KDE [377]				
	Liu et al. (2016) [252]	Running Average-Three CFD [252]	Idem RA-TCFD	Idem RA-TCFD	RGB	AND
	Huang et al. (2016) [174]	SuBSENSE [378]	Idem SuBSENSE	Idem SuBSENSE	RGB-LBSP [37]	Trajectory Feedback
	Seese et al. (2006) [347]	MOG [379]/Kalman filter [328]	Idem MOG/Kalman filter	Idem MOG/Kalman filter	Intensity	Intersection
	Wang et al. (2006) [410]	GMM [379]	Idem GMM	Idem GMM	RGB	Double Local Thresholding
	Huang et al. (2017) [175]	GMM [379]	Idem GMM	Idem GMM	RGB	-
	Salman et al. (2019) [341]	Improved GMM [379]	Idem GMM	Idem GMM	RGB	Pixel-wise Posteriors [36]
	Salman et al. (2019) [342]	R-CNN	-	-	RGB	Three inputs
Dolphins Surveillance	Karnowski et al. (2015) [201]	RPCA [65]	-	-	Intensity	-
Honeybees Surveillance	Knauer et al. (2005) [213]	K-Clusters [62]	Idem K-Clusters	Idem K-Clusters	intensity	-
	Campbell et al. (2008) [63]	MOG [379]	Idem MOG	Idem MOG	RGB	-
	Kimura et al. (2012) [211]	-	-	-	-	-
	Babic et al. (2016) [19]	MOG [379]	Idem MOG	Idem MOG	RGB	-
	Pilipovic et al. (2016) [307]	MOG [379]	Idem MOG	Idem MOG	RGB	-
Spiders Surveillance	Iwatani et al. (2016) [179]	Image without foreground objects	-	Threshold	Intensity	-
Insects Surveillance	De Geus et al. (2019) [101]	Running average	Running average	Two Thresholds (Global/Local	RGB	-
Lizards Surveillance	Nguwi et al. (2016) [291]	-	-	-	-	-
Pigs Surveillance	McFarlane and Schofield (1995) [271]	Approximated Median	YEs	Idem Median	Intensity	-
	Guo et al. (2015) [142]	PM-MOG [142]	Idem MOG	Idem MOG	RGB	Prediction Mechanism
	Tu et al. (2014) [399]	MOG-DWT [399]	Idem MOG	Idem MOG	Intensity/Texture	OR
	Tu et al. (2015) [400]	CFD [240]	-	Difference	RGB	Illumination and Reflectance Estimation
Hinds Surveillance	Khorrami et al. (2012) [206]	RPCA [65]	-	-	Intensity	-
	Giraldo-Zuluaga et al. (2017) [131]	Multi-Layer RPCA [131]	-	-	RGB	-
	Giraldo-Zuluaga et al. (2017) [130]	Multi-Layer RPCA [131]	-	-	RGB	-

Table 9 Background models used for intelligent visual observation of animals and insects: An Overview. ”-” indicated that the background model used is not indicated in the paper.

6.1 Forest Environments

The aim is to detect humans or animals but the motion of the foliage generate rapid transitions between light and shadow. Furthermore, humans or animals can be partially occluded by branches. First, Boult et al. [43] addressed these problems to detect humans into the woods with omnidirectional cameras. In 2017, Shakeri and Zhang [355] employed a robust PCA method to detect animals in clearing zones. A camera trap is used to capture the videos. Experiments show that the proposed method outperforms most of the previous RPCA methods on the illumination change dataset (ICD). In 2017, Yousif et al. [442] designed a joint background modeling and deep learning classification to detect and distinguish human and animals. The block-wise background modeling employ three features (intensity, histogram, Local Binary Pattern [53], and Histogram of Oriented Gradient (HOG) [53]) to be robust against waving trees. In 2018, Yousif et al. [443] presented Animal Scanner which is a software for classifying humans, animals, and empty frames in camera trap images taken in forest environments. First, each frame from a given camera trap sequence are rescaled into a specific width and height and then divided the rescaled image into 736 (or 32×23) regular blocks. Then different features (i.e LBP, HOG, Gray Level Cooccurrence Matrix) are extracted from each block. To determine which block contain moving animals, Yousif et al. [443] used the minimum feature distance (MFD). Second, a given frame is subtracted from the background frame abd a threshold value determines whether this block belongs to background or foreground. The foreground blocks are then connected to represent the foreground regions which are the region candidates to be verified as human, animal, or background in order to label them with tagged bounding boxes. In 2017, Janzen et al. [181] detected movement in a region via background subtraction, image thresholding, and fragment histogram analysis. This system reduced the number of images for human consideration to one third of the input set with a success rate of proper identification of ungulate crossing between 60% and 92%, which is suitable in larger study context.

6.2 River Environments

The idea is to detect foreign objects in the river (bottles, floating woods, ...) for **(1)** the preservation of the river or for **(2)** the preservation of the civil infrastructures such as bridges and dams on the rivers [6,7]. In the first case, foreign objects pollute the environment and then animals are affected [475]. In the second case, foreign objects such as fallen trees, bushes, branches of fallen trees and other small pieces of wood can damage bridges and dams on the rivers. The risk of damage by trees is directly proportional to their size. Larger fallen trees are more dangerous than the smaller parts of fallen trees. These trees often remain wedged against the pillars of bridges and help in the accumulation of small branches, bushes and debris around. In these river environments, both background water waves and floating objects of interest are in motion. Moreover, the flow of river water varies from the normal flow during floods, and thus it causes more motion. In addition, small waves and water surface illumination, cloud shadows, and similar phenomena add non-periodical background

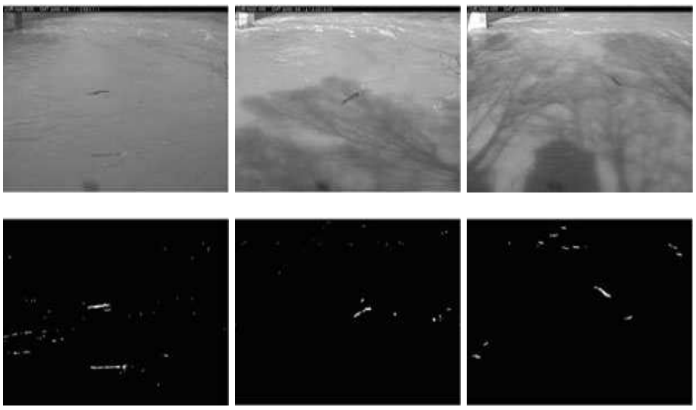

Fig. 16 The first row presents original frames in a river environment with floating woods under different illumination conditions. The second row shows the corresponding foreground detection masks. Images from Ali et al. [7].

changes. In 2003, Zhong et al. [475] used Kalman filter to detect bottles in waving river for an application of type (1). In 2012, Ali et al. [6] used a space-time spectral mode for an application of type (2). In a further work, Ali et al. [7] added to the MOG model a rigid motion model to detect floating woods. Figure 16 shows images of a river environment in which floating woods need to be detected.

6.3 Ocean Environments

The aim is to detect **(1)** ships for the optimization of the traffic which is the case in most of the applications, **(2)** foreign objects to avoid the collision with foreign objects [42], and **(3)** foreign people (intruders) because ships are in danger of pirate attacks both in open waters and in a harbor environment [382]. These scenes are more challenging than calm water scenes because of the waves breaking near the shore. Moreover boat wakes and weather issues contribute to generate a highly dynamic background. Practically, the motion of the waves generate false positive detections in the foreground detection [3]. In a valuable study, Prasad et al. [314] [315] provided a list of the challenges met in videos acquired in maritime environments, and applied on the Singapore-Marine dataset the 34 algorithms that participated in the ChangeDetection.net competition. Experimental results [314] show that all these methods are ineffective, and produced false positives in the water region or false negatives while suppressing water background. Practically, the challenges can be classified as follows as developed by Prasad et al. [314]:

- Weather and illumination conditions such as bright sunlight, twilight conditions, night, haze, rain, fog, ...
- The solar angles induce different speckle and glint conditions in the water.
- Tides also influence the dynamicity of water.
- Situations that affect the visibility influence the contrast, statistical distribution of sea and water, and visibility of far located objects.

- Effects such as speckle and glint create non-uniform background statistics which need extremely complicated modeling such that foreground is not detected as the background and vice versa.
- Color gamuts for illumination conditions such as night (dominantly dark), sunset (dominantly yellow and red), and bright daylight (dominantly blue), and hazy conditions (dominantly gray) vary significantly.

Figure 3 and Figure 4 show images of ocean environments in which ships, cars and persons need to be detected in maritime and coastal surveillance.

6.4 Submarine Environments

There are three kinds of aquatic underwater environments (also called underwater environments): **(1)** swimming pools, **(2)** tank environments for fish, and **(3)** open sea environments.

6.4.1 Swimming pools

In swimming pools, there are water ripples, splashes and specular reflections. First, Eng et al. [118] designed a block-based median background model and used the CIELa*b* color space for outdoor pool environments to detect swimmers under amid reflections, ripples, splashes and rapid lighting changes. Partial occlusions are resolved using a Markov Random Field framework that enhances the tracking capability of the system. In a further work, Eng et al. [119] employed a region-based single multivariate Gaussian model to detect swimmers, and used the CIELa*b* color space as in Eng et al. [118]. Then, a spatio-temporal filtering scheme enhanced the detection because swimmers are often partially hidden by specular reflections of artificial nighttime lighting. In an other work, Lei and Zhao [233] employed a Kalman filter [328] to deal with light spot and water ripple. In a further work, Chan et al. [74] detected swimmers by computing dense optical flow and the MOG model on video sequences captured at daytime, and nighttime, and of different swimming styles (breaststroke, freestyle, backstroke). For private swimming, Peixoto et al. [303] combined the mean model and the two CFD, and used the HSV invariant color model space. Figure 17 and Figure 18 show example of segmentation of swimmers.

6.4.2 Tank environments

There are three kinds of tanks: **(1)** tanks in aquarium which reproduce the maritime environment, **(2)** tanks for industrial aquaculture, and **(3)** tanks for studies of fish's behaviors. The challenges met in tank environments can be classified as follows

- **Challenges related to the environments:** Illumination changes are owed to the ambient light, the spotlights which light the tank from the inside and from the outside, the movement of the water due to fish and the continuous renewal of the water. In addition for tank in aquarium, moving algae generate false detections as developed by El Baf et al. [22,304].

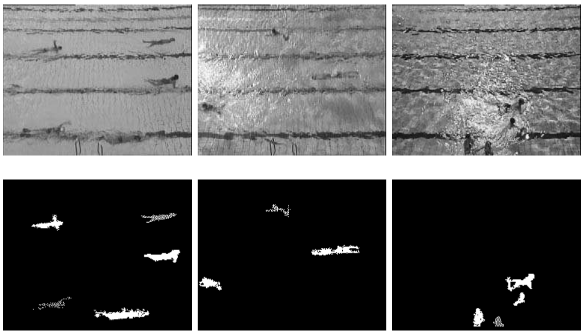

Fig. 17 The first row presents original frames in swimming pools at different time intervals from 9am to 8pm. The second row shows the segmentation of swimmers obtained by Eng et al. [119].

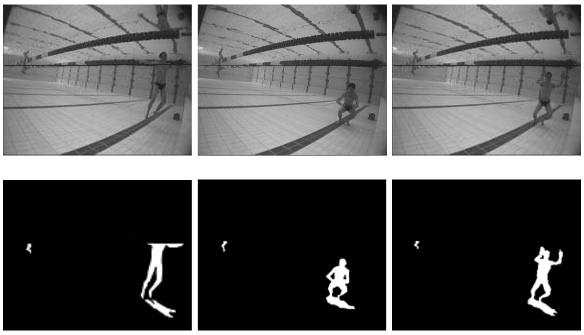

Fig. 18 The first row presents original frames in swimming pools. The second row shows the segmentation of swimmers obtained by Fei et al. [124]. Remark: Only the underwater part is segmented. It is the reason why the head is missing in the first segmented image.

– **Challenges related to fish:** The movement of fish is different due to their species and the kind of tank. Furthermore, their number is different. In tank for aquarium, there are different species as in El Baf et al. [22] where there are ten species of tropical fish. However, fish of the same species tend to have the same behavior. But, there are several species which swim at different depths. Furthermore, they can be occluded by algae or other fish. In an other way in industrial aquaculture, the number of fish is bigger than aquarium, and all the fish are from the same species and thus they have the same behavior. For example in Abe et al. [2], there were 250 Japanese rice fish, of which 170-190 were detected by naked eye observation in the visible area during the recording period. Furthermore, these fish tend to swim at various depths in the tank.

6.4.3 Open sea environments

In underwater open sea environments, the degree of luminosity and water flow vary depending upon the weather and the time of the day. The water may also have varying degrees of clearness and cleanness as developed by Spampinato et al. [374]. In

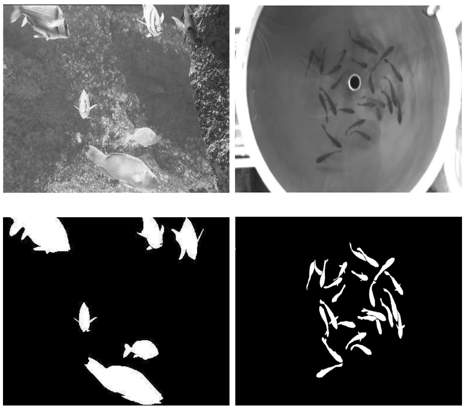

Fig. 19 The first row presents original frames in tank environments: Aquarium (El Baf et al. [22]), Aquaculture (Zhou et al. [477]). The second row shows the foreground detection mask or the corresponding ground truth (GT).

addition in subtropical waters, algae grow rapidly and appear on camera lens. Consequently, different degrees of greenish and bluish videos are produced. In order to reduce the presence of the algae, lens is frequently and manually cleaned. Practically, the challenges in open sea environments can classified as developed by Kavasidis and Palazzo [202] and Spampinato et al. [376]:

– **Light changes:** Every possible lighting conditions need to be taken into account because the video feeds are captured during the whole day and the background subtraction algorithms should consider the light transition.
– **Physical phenomena:** Image contrast is influenced by various physical phenomena such as typhoons, storms or sea currents which can easily compromise the contrast and the clearness of the acquired videos.
– **Murky water:** It is important to consider that the clarity of the water during the day could change due to the drift and the presence of plankton in order to investigate the movements of fish in their natural habitat. Under these conditions, targets that are not fish might be detected as false positives.
– **Grades of freedom:** In underwater videos the moving objects can move in all three dimensions whilst videos containing traffic or pedestrians videos are virtually confined in two dimensions.
– **Algae formation on camera lens:** The contact of sea water with the cameras lens facilitates the quick formation of algae on top of camera.
– **Periodic and multi-modal background:** Arbitrarily moving objects such as stones and periodically moving ojects such as plants subject to flood-tide and drift can generated false positive detections.

In a comparative evaluation done in 2012 on the Fish4Knowledge dataset [203], Kavasidis and Palazzo [202] evaluated the performance of six state-of-the-art background subtraction algorithms (GMM [379], APMM [122], IM [309], Wave-Back [311], Codebook [210] and ViBe [30]) in the task of fish detection in unconstrained and underwater video. Kavasidis and Palazzo [202] concluded that:

1. At the blob level, the performance is generally good in videos which presented scenes under normal weather and lighting conditions and more or less static backgrounds, except from the Wave-Back algorithm. On the other hand, the ViBe algorithm excelled in nearly all the videos. GMM and APMM performed somewhere in the middle, with the GMM algorithm resulting slightly better than the PMM algorithm. The codebook algorithm gave the best results in the high resolution videos.
2. At the pixel level, all the algorithms show a good pixel detection rate, i.e. they are able to correctly identify pixels belonging to an object, with values in the range between 83.2% for the APMM algorithm and 83.4% for ViBe. But, they provide a relatively high pixel false alarm rate, especially the Intrinsic Model and Wave-back algorithms when the contrast of the video was low, a condition encountered during low light scenes and when violent weather phenomena were present (typhoons and storms).

In a further work in 2017, Radolko et al. [323] identified the following five main difficulties:

- **Blur:** It is due to the forward scattering in water and makes it impossible to get a sharp image.
- **Haze:** Small particles in the water cause back scatter. The effect is similar to a sheer veil in front of the scene.
- **Color Attenuation:** Water absorbs light stronger than air. Also, the absorption effect depends on the wavelength of the light and this leads to underwater images with strongly distorted and mitigated colors.
- **Caustics:** Light reflections on the ground caused by ripples on the water surface. They are similar to strong, fast moving shadows which makes them very hard to differentiate from dark objects.
- **Marine Snow:** Small floating particles which strongly reflect light. Mostly they are small enough that they are filtered out during the segmentation process, however, they still corrupt the image and complicate for example the modeling of the static background.

Experimental results [323] done on the dataset UnderwaterChangeDetection.eu [323] show that GSM [324] gives better performance than MOG-EM [199], Zivkovic-Heijden GMM (also called ZHGMM) [481] and EA-KDE (also called KNN) [480]. In an other work, Rout et al. [337] designed a spatio-Contextual GMM (SC-GMM) that outperforms 18 background subtraction algorithms such as **1)** classical algorithms like the original MOG, the original KDE and the original PCA, and **2)** advanced algorithms like DPGMM (VarDMM) [149], PBAS [161], PBAS-PID [394], SuBSENSE [378] and SOBS [260] both on the Fish4Knowledge dataset [203] and the dataset UnderwaterChangeDetection.eu [323].

Thus, natural environments involve multi-modal backgrounds, and changes of the background structure need to be captured from the background model to avoid a big amount of false detection rate. Practically, events such as CVAUI (Computer Vision for Analysis of Underwater Imagery) 2015 and 2016 in conjunction with ICPR ad-

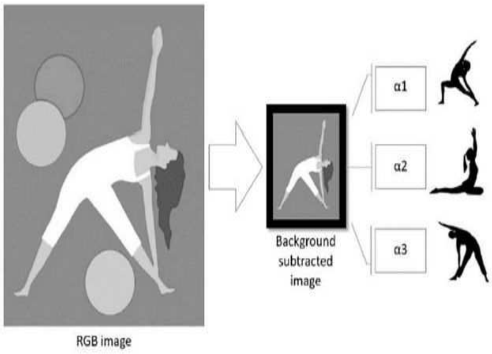

Fig. 20 Aerobic routine detection. Images from John et al. [196].

dressed the problem of the detection in ocean surface and underwater environments.

7 Automated Visual Analysis of Human Activities

Background subtraction is also used for visual analysis of human activities like in sport **(1)** when important decisions need to be made quickly, **(2)** for precise analysis of athletic performance, and **(3)** for surveillance in dangerous activities. For example, John et al. [196] provide a system that allow a coach to obtain real-time feedbacks to ensure that the routine is performed in a correct manner (See Figure 20). During the initialization stage, the stationary camera captures the background image without the user, and then each current image is subtracted to obtain the silhouette. This method is the simplest way to obtain the silhouette and is useful in this context as it is indoor scene with control on the light. In practice, the method was implemented inside an Augmented Reality (AR) desktop app that employs a single RGB camera. The detected body pose image is compared against the exemplar body pose image at specific intervals. The pose matching function is reported to have an accuracy of 93.67%. However, the current trend in pose estimation is not to use background subtraction, but to work on individual images. In an other work, Tamas [385] designed a system for estimating the pose of athletes exercising on indoor rowing machines. Practically, Zivkovic-Heijden GMM [481] and Zivkovic-KDE [480] available in OpenCV were tested with success but with the main drawbacks that these methods mark the shadows projecting to the background as foreground. Then, Tamas [385] developed a fast and accurate background subtraction method which allows to extract the head, shoulder, elbow, hand, hip, knee, ankle and back positions of a rowing athlete (See Figure 21). For surveillance (See Figure 22), Bastos [31] employed the original MOG to detect surfers on waves in Ribeira dIlhas beach (Portugal). The proposed system obtains a false positive rate of 1.77 for a detection rate of 90% but the amount of memory and computational time required to process a video sequence is the main drawback of this system.

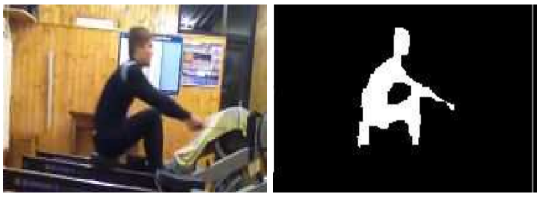

Fig. 21 Rowing Activities: The first image shows the original image. The second image shows the detected silhouette. Images from Tamas [385].

8 Optical Motion Capture

The aim is to obtain a 3D model of an actor filmed by a system with multi-cameras wich can require markers [136] or not [66, 163, 276, 277, 86, 208, 145]. Because it is impossible to have a rigourous 3D reconstruction of a human model, a 3D voxel approximation [276, 277]obtained by shape-from-silhouette (also called visual hull) is computed with the silhouettes obtained from each camera. Then, the movememts are tracked and reproduced on human body model called avatar. Optical motion capture is used in computer game, virtual clothing application and virtual reality. In all optical motion capture systems, it requires in its first step to obtain a full and precise capture of human sihouette and movements from the different point of view provided by the multiple cameras. One common technique for obtaining silhouettes also used in television weather forecasts and for cinematic special effects for background substitution is chromakeying (also called bluescreen matting) which is based on the fact that the actual scene background is a single uniform color that is unlikely to appear in foreground objects. Foreground objects can then be segmented from the background by using color comparisons but chromakey techniques do not admit arbitrary backgrounds, which is a severe limitation as developed by Buehler et al. [60]. Thus, background subtraction is more suitable to obtain the silhouette. Practically, silhouettes are then extracted in each view by background subtraction, and thus this step is also called silhouette detection or silhouette extraction. Because the acquisition is made in indoor scenes, the background model required can be uni-modal, and shadows and highlights are the main challenges in this application. In this context, Wren et al. [426] used a single Gaussian model in YUV color space whilst Horprasert et al. [164, 166] used a statistical model with shadow and highlight suppression. In 2019, Habermann et al. [145] presented a system called LiveCap which allows real-time human motion capture from monocular video. By assuming that the background is static, that its color is sufficiently different from the foreground, and a few frames of the empty scene are recorded before performance capture begins, Habermann et al. [145] employed the Zivkovic-Heijden GMM [481] algorithm. Figure 23 shows a representative illustration of the steps of optical motion capture system until the visual hull. Furthermore, Table 10 shows an overview of the different publications in the field with information about the background model, the background maintenance, the foreground detection, the color space and the strategies used by the authors.

Applications	Type	Background model	Background Maintenance	Foreground Detection	Color Space	Strategies
Visual Hull Computing						
1) Image-based Modeling						
Matusik et al. (2000) [427]	Marker Free	MOG [126]	Idem MOG	Idem MOG	RGB	-
2) Optical Motion Capture						
Wren et al. (1997) [427]	Marker Free (Pfinder [427])	SG [427]	Idem SG	Idem SG	YUV	-
Horprasert et al. (1998) [163]	Marker Free	W4 [157]	Idem W4	Idem W4	Intensity	-
Horprasert et al. (1999) [164]	Marker Free	Codebook [164]	Idem Codebook [164]	Idem Codebook [164]	RGB	Shadow Detection
Horprasert et al. (2000) [166]	Marker Free	Codebook [164]	Idem Codebook [164]	Idem Codebook [164]	RGB	Shadow Detection
Mikic et al. (2002) [276]	Marker Free	Codebook [164]	Idem Codebook [164]	Idem Codebook [164]	RGB	Shadow Detection
Mikic et al. (2003) [277]	Marker Free	Codebook [164]	Idem Codebook [164]	Idem Codebook [164]	RGB	Shadow Detection
Chu et al. (2003) [86]	Marker Free	SG [272]	Idem SG	Idem SG	HSV	-
Carranza et al. (2003) [66]	Marker Free	SG [427]	Idem SG	Idem SG	YUV	Shadow Detecion
Guerra-Filho (2005) [136]	Marker Detection	Median	Idem Median	Idem Median	RGB	-
Kim et al. (2007) [208]	Marker Free	SGG (GGF) [209]	Idem SGG	Idem SGG	RGB	Small regions suppression
Park et al. (2009) [298]	Photorealistic Avatars	Codebook with online MoG	Idem Codebook	Idem Codebook	RGB	Markov Random Field (MRF)
Habermann et al. (2019) [145]	Photorealistic Avatars	Zivkovic-Heijden GMM [481]	Idem Zivkovic-Heijden GMM	Idem Zivkovic-Heijden GMM	RGB	-
Human-Machine Interaction (HMI)						
1) Arts						
Levin (2006) [234]	Art	-	-	-	-	-
2) Games						
3) Ludo-Multimedia Applications						
Penciuc et al. (2006) [304]	Fish Detection	MOG	Idem MOG	Idem MOG	RGB	-
El Baf et al.(2007) [22]	Fish Detection	MOG	Idem MOG	Idem MOG	RGB	-
El Baf et al.(2007) [20]	Fish Detection	MOG	Idem MOG	Idem MOG	RGB	-
Vision-based Hand Gesture Recognition						
1) Human Computer Interface (HCI)						
Park and Hyun (2013) [299]	Hand Detection	Average	Selective Maintenance	Idem Running Average	Intensity	-
Stergiopoulou et al. (2014) [380]	Hand Detection	Three CFD/BGS [92]	Idem CFD/BGS[92]	Idem CFD/BGS[92]	RGB	-
2) Behavior Analysis						
Perrett et al. (2016) [305]	Hand Detection	PBAS [161]	Idem PBAS [161]	PBAS [161]	Intensity	Post-processing (Median filter)
3) Sign Language Interpretation and Learning						
Elsayed et al. (2015) [112]	Hand Detection	First Frame without Foreground Objects	Running Average	Idem Running Average	YCrCb	-
4) Robotics						
Khaled et al. (2015) [205]	Hand Detection	Average	Running Average	Idem Running Average	Intensity/Color	Contour Extraction Algorithm
Video Coding						
Chien et al. (2002) [83]	MPEG-4 (QCIF Format)	Progressive Generation (CFD [1])	Progressive Maintenance	CD [1]	Intensity	Post Processing/Shadow Detection
Paul et al. (2010) [300]	H.264/AVC (CIF Format)	MOG on decoded pixel intensities	Selective Maintenance	Block-based Difference	Intensities	-
Paul et al. (2013) [302]	H.264/AVC (CIF Format)	MOG on decoded pixel intensities	Selective Maintenance	Block-based Difference	Intensities	-
Paul et al. (2013) [301]	H.264/AVC	MOG [155]	-	-	Color	-
Zhang et al. (2010) [456]	H.264/AVC	Non-Parametric Background Generation [456]	Idem BG [456]	Idem BG [456]	Color	-
Zhang et al. (2012) [458]	H.264/AVC	SWRA [458]	Selective Maintenance	-	Color	-
Chen et al. (2012) [80]	H.264/AVC	Average	Selective Maintenance	-	Color	-
Han et al. (2012)[152]	H.264/AVC	-	-	-	Color	Panorama Background/Motion Compensation
Zhang et al. (2012) [459]	H.264/AVC	MSBDC [459]	Selective Maintenance	-	Color	-
Geng et al. (2012)[129]	H.264/AVC	SWRA [458]	Selective Maintenance	-	Color	-
Zhang et al. (2014) [455]		BMAP [455]	Selective Maintenance	-	Color	-
Zhao et al. (2014) [467]	HEVC	BFDS [467]	-	-	Color	-
Zhang et al. (2014) [457]	HEVC	Running Average	-	-	Color	-
Chakraborty et al. (2014) [70]	HEVC	KDE/Median	Selective Maintenance	-	Intensity	-
Chakraborty et al. (2014) [71]	HEVC	KDE/Median	Selective Maintenance	-	Intensity	-
Chakraborty et al. (2017) [72]	HEVC	KDE/Median	Selective Maintenance	-	Intensity	Scene Adaptive Non-Parametric Technique
Chen et al. (2012) [79]	H.264/AVC	RPCA (LRSD [79])	-	-	Intensity	-
Guo et al. (2013) [140]	H.264/AVC	RPCA (Dictionary Learning [140])	-	-	Color	-
Zhao et al. (2013) [468]	HEVC	RPCA (Adaptive Lagrange Multiplier [468])	-	-	Color	-

Table 10 Background models used for optical motion capture and video coding: An Overview. ”-” indicated that the background model used is not indicated in the paper.

Fig. 22 Detection of surfers in Ribeira dIlhas beach (Portugal): The first image shows the original image. The second image shows detected bounding boxes. Images from Bastos [31].

Fig. 23 Optical motion capture system: The first image shows original images taken by different cameras. The second image shows the corresponding foreground mask images. The third image shows two images of the corresponding visual hull. Images from Mikic et al. [277].

9 Visual-based Interaction Applications

9.1 Human-Machine Interaction

Several applications need interactions between human and machine through a video acquired in real-time by fixed cameras such as games (Microsoft's Kinect) and ludo-applications such as Aqu@theque [20, 22, 304].

- **Arts and Games:** First, the person's body pixels are located with background subtraction, and then this information is used as the basis for graphical responses in interactive systems as developed by Levin et al. [234] website[9]). In 2003, Warren [419] presented a vocabulary of various essential interaction techniques which can use this kind of body-pixel data. These schema are useful in "mirror-like" contexts, such as Myron Krueger's Videoplace [10]), or video games like the PlayStation Eye-Toy, in which the participant can observe his own image or silhouette composited into a virtual scene.
- **Ludo-Multimedia Applications:** In this type of applications, the user can select a moving object of interest on a screen, and then information are provided. A representative example is the Aqu@theque project which allows a visitor of an aquarium to select on an interactive interface fishes that are filmed on line by a remote video camera. This interface is a touch screen divided into two parts. The first one shows the list of fishes present in the tank and is useful all the time. The filmed scene is visualized in the remaining part of the screen. The computer can supply information about fishes selected by the user with his finger. A fish is then

[9] http://www.flong.com/texts/essays/essaycvad/
[10] http://www.medienkunstnetz.de/works/videoplace/

automatically identified and some educational information about it is put on the screen. The user can also select each identified fish whose virtual representation is shown on another screen. This second screen is a virtual tank reproducing the natural environment where the fish lives in presence of it preys and predators. The behavior of every fish in the virtual tank is modeled. The project is based on two original elements: the automatic detection and recognition of fish species in a remote tank of an aquarium and the behavioral modeling of virtual fishes by multi-agents method.

9.2 Vision-based Hand Gesture Recognition

In vision-based hand gesture recognition, it is needed to detect, track and recognize hand gesture for several applications such as human-computer interface, behavior studies, sign language interpretation and learning, teleconferencing, distance learning, robotics, games selection and object manipulation in virtual environments. Figure 24 shows sample of hands detection in three applications in which we can see different challenges in hand detection: **(1)** hands can be near or far the cameras and **(1)** fine details need be detected as fingers in the application of sign language. We have classified them as follows:

- **Human-Computer Interface:** Common HCI techniques still rely on simple devices such as keyboard, mice, and joysticks, which are not enough to convoy the latest technology. Hand gesture has become one of the most important attractive alternatives to existing traditional HCI techniques. Practically, hand gesture detection for HCI is achieved using real-time video streaming by removing the background using a background algorithm. Then, every hand gesture can be used for augmented screen as in Park and Hyun [299] or for computer interface in vision-based hand gesture recognition as in Stergiopoulou et al. [380].
- **Behavior Analysis:** Perrett et al. [305] analyzed which vehicle occupant is interacting with a control on the center console when it is activated, enabling the full use of dual-view touch screens and the removal of duplicate controls. The proposed method is first based on hands detection made by a background subtraction algorithm incorporating information from a superpixel segmentation stage. Practically, Perrett et al. [305] chose PBAS [161] as background subtraction algorithm because it allows small foreground objects to decay into the background model quickly whilst larger objects persist, and superpixel Simple Linear Iterative Clustering (SLIC) algorithm as the super-pixel segmentation method. Experimental results [305] on the centers panel of a car show that the hands can be effectively detected both in day and night conditions.
- **Sign Language Interpretation and Learning:** Elsayed et al. [112] proposed to detect moving hand area precisely in a real time video sequence using a threshold based on skin color values to improve the segmentation process. The initial background is the first frame without foreground objects. Then, the foreground detection is obtained with a threshold on the difference between the background and the current frame in YCrCb color space. Experimental results [112] on indoor and outdoor scenes show that this method can efficiently detect the hands.

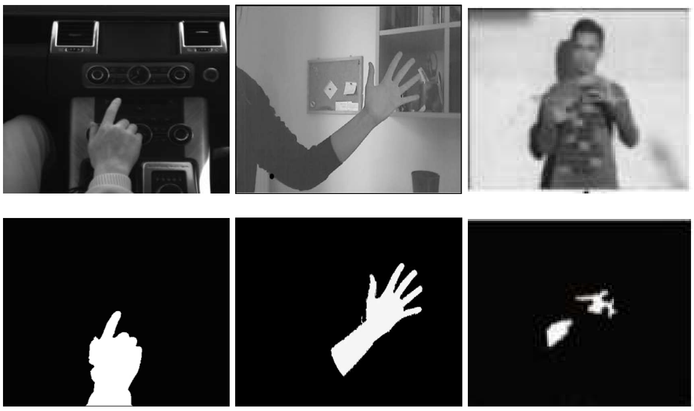

Fig. 24 Vision-based hand detection systems: The first row shows the original images in behavior analyis (Perrett et al. [305]), Human-Computer Interface (Stergiopoulou et al. [380]) and Sign Language Interpretation and Learning (Elsayed et al. [112]), respectively. The second row the corresponding foreground detection masks.

– **Robotics:** Khaled et al. [205] used a running average to detect the hands, and the 1$ algorithm [205] for hands template matching. Then, five hand gestures are detected and translated into commands that can be used to control robot movements.

10 Content-based Video Coding

To generate video contents, videos have to be segmented into video objects and tracked as they transverse across the video frames. The registered background and the video objects are then encoded separately to allow the transmission of video objects only in the case when the background does not change over time as in video surveillance scenes taken by a fixed camera. So, video coding needs an effective method to separate moving objects from static and dynamic environments [83,446].

For H.264 video coding, Paul et al. [302,300] proposed a video coding method using a reference frame which is the most common frame in scene generated by dynamic background modeling based on the MOG model with decoded pixel intensities instead of the original pixel intensities. Thus, the proposed model focuses on rate-distortion optimization whereas the original MOG primarily focuses on moving object detection. In a further work, Paul et al. [301] presented an arbitrary shaped pattern-based video coding (ASPVC) for dynamic background modeling based on MD-MOG. Even if these dynamic background frame based video coding methods based on MoG based background modeling achieve better rate distortion performance compared to the H.264 standard, they need high computation time, present low coding efficiency for dynamic videos, and prior knowledge requirement of video content. To address these limitations, Chakraborty et al. [70,71] presented an Adaptive Weighted non-Parametric (WNP) background modeling technique, and further embedded it into HEVC video coding standard for better rate-distortion performance.

Being non-parametric, WNP outperforms in dynamic background scenarios compared to MoG-based techniques without a priori knowledge of video data distribution. In a further work, Chakraborty et al. [72] improved WNP by using a scene adaptive non-parametric (SANP) technique developed to handle video sequences with high dynamic background.

To address the limitations of the H.264/AVC video coding, Zhang et al. [456] presented a coding scheme for surveillance videos captured by fixed cameras. This scheme used a nonparametric background generation proposed by Liu et al. [253]. In a further work, Zhang et al. [458] proposed a Segment-and-Weight based Running Average (SWRA) method for surveillance video coding. In a similar approach, Chen et al. [80] used a timely and bit saving background maintenance model. In a further work, Zhang et al. [459] used a macro-block-level selective background difference coding method (MSBDC). In an other work, Zhang et al. [455] presented a Background Modeling based Adaptive Background Prediction (BMAP) method. In an other approach, Zhao et al. [467] proposed a background-foreground division based search algorithm (BFDS) to address the limitations of the HECV coding whilst Zhang et al. [457] used a running average. For moving cameras, Han et al. [152] proposed to compute a panorama background with motion compensation.

Because H.264/AVC is not sufficiently efficient for encoding surveillance videos since it does not exploit the strong background temporal redundancy, Chen et al. [79] used for the compression the RPCA decomposition model [65] which decomposed a surveillance video into the low-rank component (background), and the sparse component (moving objects). Then, Chen et al. [79] developed different coding methods for the two different components by representing the frames of the background by very few independent frames based on their linear dependency. Experimental results [79] show that the proposed RPCA method called Low-Rank and Sparse Decomposition (LRSD) outperforms H.264/AVC, up to 3 dB PSNR gain, especially at relatively low bit rate. In an other work, Guo et al. [140] trained a background dictionary based on a small number of observed frames, and then separated every frame into the background and motion (foreground) by using the RPCA decomposition model [65]. In a further step, Guo et al. [140] stored the compressed motion with the reconstruction coefficient of the background corresponding to the background dictionary. Experimental results [140] show that this RPCA method significantly reduces the size of videos while gains much higher PSNR compared to the state-of-the-art codecs. In these RPCA video coding standards, the selection of Lagrange multiplier is crucial to achieve trade-off between the choices of low-distortion and low-bitrate prediction modes, and the rate-distortion analysis shows that a larger Lagrange multiplier should be used if the background in a coding unit took a larger proportion. To take into account this fact, Zhao et al. [79] proposed a modified Lagrange multiplier for rate-distortion optimization by performing an in-depth analysis on the relationship between the optimal Lagrange multiplier and the background proportion, and then Zhao et al. [79] presented a Lagrange multiplier selection model to obtain the optimal coding performance. Experimental results show that this Adaptive Lagrange Multiplier Coding Method (ALMCM) [79] achieves 18.07% bit-rate saving on CIF sequences and 11.88% on SD sequences against the background-irrelevant Lagrange multiplier selection method.

Fig. 25 Background substitution: The first image shows the original image. The second image shows the image with background substitution. Images from Huang et al. [169].

Table 10 shows an overview of the different publications in the field with information about the background model, the background maintenance, the foreground detection, the color space and the strategies used by the authors.

11 Miscellaneous Applications

1. **Background Substitution:** This task is also called background cut and video matting. It consist in extracting the foreground from the input video and then combine it with a new background. Thus, background subtraction can be used in the first step as in Huang et al. [169].
2. **Carried Baggage Detection:** Tzanidou [401] proposed a carried baggage detection based on background modeling which used multi-directional gradient and phase congruency. Then, the detected foreground contours are refined by classifying the edge segments as either belonging to the foreground or background. Finally, a contour completion technique by anisotropic diffusion is employed. The proposed method targets cast shadow removal, gradual illumination change invariance, and closed contour extraction.
3. **Fire detection:** Several fire detection systems [395, 329, 150, 135] use in their first step background subtraction to detect moving pixels. Second, colors of moving pixels are checked to evaluate if they match to pre-specified fire-colors, then wavelet analysis in both temporal and spatial domains is carried out to determine high-frequency activity as developed by Toreyin et al. [395].
4. **OLED defect detection:** Organic Light Emitting Diode (OLED) is a light-emitting diode which is popular in the display industry due to its advantages such as colorfulness, light weight, large viewing angle, and power efficiency as developed by Jeon and Yoo [192]. But, the complex manufacturing process also produces defects. which may consistently affect the quality and life of the display. In this context, an automated inspection system based on computer vision is needed. Practically, OLED presents a feature of gray scale and repeating patterns, but significant intensity variations are also presented. Thus, background subtraction can be used for the inspection. For example, KDE based background model [192] can be built by using multiple repetitive images around the target area. Then, the model is used to compute likelihood at each pixel of the target image.

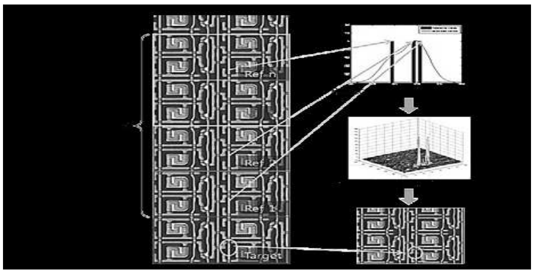

Fig. 26 OLED defect detection: The first image shows the original image. The second image shows the image with background substitution. Images from Jeon and Yoo [192].

12 Discussion

12.1 Solved and Unsolved Challenges

In this section, we have grouped in Table 11 the solved and unsolved challenges by application to give an overview in which applications investigations need to be made. We can see that the main difficult challenges are camera jitter, illumination changes and dynamic backgrounds that occur in outdoor scenes. Future research may concern these challenges.

12.2 Prospective Solutions

Prospective solutions to handle the unsolved challenges can be the use of recent background subtraction methods based on fuzzy models [27,24,26,84], RPCA models [185,186,191,187], deep learning models [56,246,244,279,280,340], and semantic models [226,345,57,247,447] . Among these recent models, several algorithms are potential usable algorithms for real applications:

12.2.1 Statistical models

First, it is interesting to consider improvements of the current used models such as MOG [379], codebook [210], ViBe [30], and PBAS [161]. Instead of the original versions of MOG, codebook and ViBe algorithms employed in most of the reviewed works in this paper, several improvements of MOG [269,212,123,452,337,259] as well as codebook [471,356,222,464], ViBe [172,151,451,478] and PBAS [190] algorithms are potential usable methods for these real applications.

For example, Goyal and Singhai [133] evaluated six improvements of MOG on the CDnet 2012 dataset showing that Shah et al.'s MOG [351] and Chen et Ellis'MOG [81] both published in 2014 achieve significantly better detection while being usable in real applications than previously published MOG algorithms, that are MOG in 1999, Adaptive GMM P1C2-MOG-92 in 2003, Zivkovic-Heijden GMM [481] in 2004, and Effective GMM [228] in 2005.

Applications	Scenes	Challenges	Solved-unsolved
1) Intelligent Visual Surveillance of Human Activities			
1.1) Traffic Surveillance	Outdoor scenes	Multi-modal backgrounds	Partially solved
		Illumination changes	Partially solved
		Camera jitter	Partially solved
1.2 Airport Surveillance	Outdoor scenes	Illumination changes	Partially solved
		Camera jitter	Partially solved
		Illumination changes	Partially solved
1.3) Maritime Surveillance	Outdoor scenes	Multimodal backgrounds	Partially solved
		Illumination changes	Partially solved
		Camera jitter	Partially solved
1.4) Store Surveillance	Indoor scenes	Multimodal backgrounds	Splved
1.5) Coastal Surveillance	Outdoor scenes	Multimodal backgrounds	Partially solved
1.6) Swimming Pools Surveillance	Outdoor scenes	Multimodal backgrounds	Solved
2) Intelligent Visual Observation of Animal and Insect Behaviors			
2.1) Birds Surveillance	Outdoor scenes	Multimodal backgrounds	Partially solved
		Illumination changes	Partially solved
		Camera jitter	Partially solved
2.2) Fish Surveillance	Aquatic scenes	Multimodal backgrounds	Partially solved
		Illumination changes	Partially solved
2.3) Dolphins Surveillance	Aquatic scenes	Multimodal backgrounds	Partially solved
		Illumination changes	Partially solved
2.4) Honeybees Surveillance	Outdoor scenes	Small objects	Partially solved
2.5) Spiders Surveillance	Outdoor scenes		Partially solved
2.6) Lizards Surveillance	Outdoor scenes	Multimodal backgrounds	Partially solved
2.7) Pigs Surveillance	Indoor scenes	Illumination changes	Partially solved
2.7) Hinds Surveillance	Outdoor scenes	Multimodal backgrounds	Partially solved
		Low-frame rate	Partially solved
3) Intelligent Visual Observation of Natural Environments			
3.1) Forest	Outdoor scenes	Multimodal backgrounds	Partially solved
		Illumination changes	Partially solved
		Low-frame rate	Partially solved
3.2) River	Aquatic scenes	Multimodal backgrounds	Partially solved
		Illumination changes	Partially solved
3.3) Ocean	Aquatic scenes	Multimodal backgrounds	Partially solved
		Illumination changes	Partially solved
3.4) Submarine	Aquatic scenes	Multimodal backgrounds	Partially solved
		Illumination changes	Partially solved
4) Intelligent Analysis of Human Activities			
4.1) Soccer	Outdoor scenes	Small objects	Solved
		Illumination changes	Solved
4.2) Rowing	Indoor scenes		Solved
4.3) Surf	Aquatic scenes	Dynamic backgrounds	Partially solved
		Illumination changes	Partially solved
5) Visual Hull Computing			
Image-based Modeling	Indoor scenes	Shadows/highlights	Solved
Optical Motion Capture	Indoor scenes	Shadows/highlights	Solved (SG)
6) Human-Machine Interaction (HMI)			
Arts	Indoor scenes		
Games	Indoor scenes		
Ludo-Multimedia	Indoor scenes/Outdoor scenes		
7) Vision-based Hand Gesture Recognition			
Human Computer Interface (HCI)	Indoor scenes		
Behavior Analysis	Indoor scenes/Outdoor scenes		
Sign Language Interpretation and Learning	Indoor scenes/Outdoor scenes		
Robotics	Indoor scenes		
7) Content based Video Coding		Indoor scenes/Outdoor scenes	

Table 11 Solved and unsolved issues : An Overview

Furthermore, there also exist real-time implementation of MOG [306, 245, 344, 343, 384], codebook [383], ViBe [218, 221] and PBAS [219, 220, 128]. In addition, robust background initialization methods [321, 387] as well as robust deep learned features [349, 350] with the MOG model could also be considered for very challenging environments like maritime and submarine environments.

12.2.2 Fuzzy concepts

Critical situations met in video surveillance generate imprecision and uncertainties in the whole process of background subtraction. Therefore, some authors have introduced fuzzy concepts in the different steps of background subtraction. The reader can refer to the survey of Bouwmans [46] for more details on fuzzy concepts used in background subtraction. The interest of fuzzy background subtraction methods is that they are no time consuming. In 2008, Sigari et al. [363,362] designed a Fuzzy Running Average (FRA) method for background subtraction. In practice, a fuzzy membership value is obtained by a fuzzy classification based on saturating linear function and then is used to compute the learning rate for the maintenance of the background average model. Experimental results on road surveillance show the pertinence of this approach in the case of camouflage. In 2008, Shakeri et al. [352,353] adapted this method in cellular automata for urban traffic applications. Each frame sequence is modeled by a cellular automata, and specific cellular automata rules are applied on pixels. Computation is done independently in all cells. Experimental results show better performance for the fuzzy cellular running average against FRA. In 2011, Yeo et al. [441,440] proposed an extension of FRA for moving vehicle detection using infrared videos. Even if these models are often surveyed as fuzzy background models, they should be more considered as fuzzy background maintenance and fuzzy foreground detection models as the fuzzy memberships only intervene in these two steps. In 2012, the foreground detection based on Choquet integral was tested with success for moving vehicles detection by Lu et al. [258,257].

12.2.3 Robust subspace models

Subspace learning method (either linear or non linear, local or global) such as Principal Component Analysis (PCA) [295,109,108,204], Independent Component Analysis (ICA) [434], Non-negative Matrix Factorization (NMF) [59,58] offer a suitable framework for background subtraction, particularly in the presence of illumination changes. However, these conventional methods are very sensitive to outliers in the data making them very sensitive to the challenges met in video-surveillance such as noise, camera jitter and dynamic backgrounds. After an empty period, subspace learning have generated renewed interest in this area in 2009 owing to the theoretical advances of robust PCA, created by Candès et al. [65]. Even if these methods are more robust than conventional subspace learning [143,187–189,184], they require batch algorithms making them not applicable in real-time applications [50]. To address this limitation, several authors proposed dynamic RPCA algorithms reviewed by Vaswani et al. [405,404] in terms of detection and algorithms's properties. Among online algorithms, incPCP[11] algorithm [331] and its corresponding improvements [332,330,335,333,364,334,336,77] as well as the ReProCS[12] algorithm [318] and its numerous variants [319,137,138,288] present both advantages in terms of detection, real-time and memory requirements. In traffic surveillance, incPCP was tested with success for vehicle counting [322,390] whereas an online RPCA algorithm was employed for vehicle and person detection [433]. In animals surveillance, Rezaei and Ostadabbas [327,16] provided a fast Robust Matrix Completion (fRMC) algorithm for in-cage mice detection using the Caltech resident intruder mice dataset.

12.2.4 Deep neural networks concepts

DNNs have been successfully applied to background generation [139,321,430,432, 431], background subtraction [18,33,56,89,246], foreground detection enhancement [448], ground-truth generation [416], and the learning of deep spatial features [231, 292, 349, 350, 461]. In 2014, Restricted Boltzman Machines (RBMs) were first employed by Guo and Qi [139] and Xu et al. [430] for background generation to further achieve moving object detection through background subtraction. In a similar manner, Xu et al. [432, 431] used deep auto-encoder networks to achieve the same task whereas Qu et al. [321] used context-encoder for background initialization. Furthermore, Convolutional Neural Networks (CNNs) has also been employed to background subtraction by Braham and Droogenbroeck [56], Bautista et al. [33] and Cinelli [89]. Other authors have employed improved CNNs such as cascaded CNNs [416], deep CNNs [18], structured CNNs [246] and two stage CNNs [469]. Through another study, Zhang et al. [461] used a Stacked Denoising Auto-Encoder (SDAE) to learn robust spatial features and modeled the background with density analysis, whereas Shafiee et al. [349] employed Neural Reponse Mixture (NeREM) to learn deep features used in the Mixture of Gaussians (MOG) model [379]. In another study, Chan [75] proposed a deep learning-based scene-awareness approach for change detection in video sequences thus applying the suitable background subtraction algorithm for the corresponding type of challenges.

All these approaches were developed by academic researchers and were not tested for real applications. Only Bautista et al. [33] tested the convolutional neural network for vehicle detection in low resolution traffic videos. But, even if the recent deep learning methods can be naturally considered owing to their robustness in presence of the concerned unsolved challenges, most of the time they are still too time and memory consuming to be currently employed in real application cases. In addition, these methods required hand-labeled data for the training and are often scene-specific. Indeed, DNNs based background subtraction can only process a certain type of scenery, and needs to be retrained for other video scenes [18]. This fact is usually not a problem because the camera is fixed when filming similar scenes. However, this may not be the case for certain applications, as pointed out by Hu et al. [168]. Thus, deep learning based methods seems for the moment to be only interesting in a theoretical point of you because there are not usable in practice. Only progress related to unsupervised and online deep learning methods can alleviate this current incompatibility [51].

12.2.5 Semantic concepts

In 2017, Braham et al. [57] proposed to leverage object-level semantics to address the variety of challenging scenarios for background subtraction. The information of a semantic segmentation algorithm is combied with the output of any background subtraction algorithm to reduce false positive detections produced by illumination changes, dynamic backgrounds, shadows, and ghosts. In addition, Braham et al. [57] maintained a fully semantic background model to improve the detection of camouflaged foreground objects. In 2019, Zeng et al. [447] designed a background subtraction algorithm with real-time semantic segmentation making it usable for real applications. This method achieves state-of-the-art performance among unsupervised background subtraction methods while operating at real-time, and even performs better than some supervised deep learning algorithms. Moreover, semantic concepts have been also used for background initialization [226,345] allowing there use in applications such as video impainting, privacy protection and computational photography.

12.3 Datasets for Evaluation

Suitable experiments set up are mostly based on a large-scale dataset sufficiently large to ensure a fair evaluation. Most of the time, the authors either used public datasets or their own datasets. In this part, we quickly survey public available datasets to evaluate algorithms in similar conditions than the real ones. For visual surveillance of human activities, there are several available dataset. Fist, Toyama et al. [397] provided in 1999 the Wallflower dataset but it was limited to person detection with one Ground-Truth (GT) image by video. In 2004,Li et al. [239] developed a more larger dataset called I2R dataset with videos with both persons and cars in indoor and outdoor scenes. But, this dataset did not cover a very large spectrum of challenges and the GTs are also limited to 20 by video. In 2012, a breakthrough was done by the BMC 2012 dataset and especially by the CDnet 2012 dataset [134] that are very realistic large scale datasets with a big amount of videos and corresponding GTs. In 2014, the CDnet 2012 dataset was extended with additional camera-captured videos ($70,000$ new frames) spanning 5 new categories, and became the CDnet 2014 dataset [415]. In addition, there also dataset for RGB-D videos (SBM-RGBD dataset [64]), infrared videos (OTCBVS 2006), and multi-spectral videos (FluxData FD-1665 [35]). For visual surveillance of animals and insects, there are very less datasets. At the best of our knowledge, there are the following main datasets that are 1) Aqu@theque [23] for fish in tank, Fish4knowledge [203] for fish in open sea, 2) the Caltech resident intruder mice dataset [61] for social behavior recognition of mice, 3) the Caltech Camera Traps (CCT[13]) dataset [34] which contains sequences of images taken at approximately one frame per second for census and recognition of species and 4) the eMammal dataset which also provides camera trap sequences. All the link to these datasets are available on the Background Subtraction Website[14]. Practically, we can note the absence of a large-scale dataset for visual surveillance of animals and insects, and for visual surveillance of natural environments.

12.4 Libraries for Evaluation

Most of the time, authors as for example Wei et al. [420] in 2018 employed one of the three background subtraction algorithms based on MOG that are available in OpenCV[15], or provided an evaluation of these three algorithms in the context of traffic surveillance like in Marcomini and Cunha [266] in 2018. But, these algorithms are less efficient that more recent algorithms available in the BGSLibrary[16] and LRS Library[17]. Indeed, BGSLibrary [366,368] provides a C++ framework to perform background subtraction with currently more than 43 background subtraction algorithms.

[11] https://sites.google.com/a/istec.net/prodrig/Home/en/pubs/incpcp

[12] http://www.ece.iastate.edu/ hanguo/PracReProCS.html

[13] https://beerys.github.io/CaltechCameraTraps/

[14] https://sites.google.com/site/backgroundsubtraction/test-sequences

In addition, Sobral [371] provided a study of five algorithms from BGSLibrary in the context of highway traffic congestion classification showing that these recent algorithms are more efficient than the three background subtraction algorithms available in OpenCV. For LRSLibrary, it is implemented in MATLAB and focus on decomposition in low-rank and sparse components. The LRSLibrary this provides a collection of state-of-the-art matrix-based and tensor-based factorization algorithms. Several algorithms can be implemented in C to reach real-time requirements.

13 Conclusion

In this review, we have firstly presented all the applications where background subtraction is used to detect static or moving objects of interest. Then, we reviewed the challenges related to the different environments and the different moving objects. Thus, the following conclusions can be made:

- All these applications show the importance of the moving object detection in video as it is the first step that is followed by tracking, recognition or behavior analysis. So, the foreground mask needs to be the most precise as possible and quickly obtained as possible for real-time applications. A valuable study of the influence of background subtraction on the further steps can be found in Varona et al. [403].
- These different applications present several specificities and need to deal with specific critical situations due to **(1)** the location of the cameras which can generate small or large moving objects to detect in respect of the size of the images, **(2)** the type of the environments, and **(3)** the type of the moving objects.
- Because the environments are very different, the background model needs to handle different challenges following the application. Furthermore, the moving objects of interest present very different intrinsic properties in terms of appearance. So, it is required to develop specific background models for a specific application or to find a universal background model which can be used in all the applications. To have a universal background model, the best way may to develop a dedicated background model for particular challenges, and to pick up the suitable background model when the corresponding challenges are detected.
- Basic models are sufficiently robust for applications which are in controlled environments such as optical motion capture in indoor scenes. For traffic surveillance, statistical models offer a suitable framework but challenges such as illumination changes and sleeping/beginning foreground objects need to add specific developments. For natural environments and in particular maritime and aquatic environments, more robust background methods than the top methods of ChangeDetection.net competition are required for maritime and submarine environment as developed by Prasad et al. [314] in 2017. Thus, recent RPCA and deep learning models have to be considered for this kind of environments.

[15] https://opencv.org/
[16] https://github.com/andrewssobral/bgslibrary
[17] https://github.com/andrewssobral/lrslibrary

The main conclusion is that BGS methods currently used in real application systems are most of time old methods slightly improved. It is due to two phenomena. First, engineers often limit their research to old well-known methods with public available implementations because they are stressed by the development time. Second, recent methods are sometimes not usable for real applications owing to their memory and computation requirements. Nevertheless, some of them has real time abilities such as incPCP [331] and ReProCS [318] for robust subspace learning.

Acknowledgements The authors acknowledge CONACYT (Consejo Nacional de Ciencia y Tecnologia) of México and Instituto Politëcnico Nacional which have provided a grant to Belmar Garcia-Garcia for one year research to the Lab. MIA at Univ. La Rochelle.

References

1. T. Aach, A. Kaup, and R. Mester. Statistical model-based change detection in moving video. *Signal Processing,*, pages 165–180, 1993.
2. S. Abe, T. Takagi, K. Takehara, N. Kimura, T. Hiraishi, K. Komeyama, S. Torisawa, and S. Asaumi. How many fish in a tank? Constructing an automated fish counting system by using PTV analysis. *International Congress on High-Speed Imaging and Photonics*, 2016.
3. A. Adam, I. Shimshoni, and E. Rivlin. Aggregated dynamic background modeling. *Pattern Analysis and Applications*, July 2013.
4. J. Aguilera, D. Thirde, M. Kampel, M. Borg, G. Fernandez, and J. Ferryman. Visual surveillance for airport monitoring applications. *Computer Vision Winter Workshop, CVW 2006*, 2006.
5. J. Aguilera, H. Wildenauer, M. Kampel, M. Borg, D. Thirde, and J. Ferryman. Evaluation of motion segmentation quality for aircraft activity surveillance. *IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pages 293–300, 2005.
6. I. Ali, J. Mille, and L. Tougne. Space-time spectral model for object detection in dynamic textured background. *International Conference on Advanced Video and Signal Based Surveillance, AVSS 2010*, 33(13):1710–1716, October 2012.
7. I. Ali, J. Mille, and L. Tougne. Adding a rigid motion model to foreground detection: Application to moving object detection in rivers. *Pattern Analysis and Applications*, July 2013.
8. S. Ali, K. Goyal, and J. Singhai. Moving object detection using self adaptive Gaussian Mixture Model for real time applications. *International Conference on Recent Innovations in Signal processing and Embedded Systems, RISE 2017*, pages 153–156, 2017.
9. T. Alldieck. Information based multimodal background subtraction for traffic monitoring applications. *Master Thesis, Aalborg University, Danemark*, 2015.
10. M. Aotani, R. Nishide, Y. Takaki, C. Ohta, and K. Oyama. Refined Cattle Detection using Composite Background Subtraction and Brightness Intensity from Birds Eye Images. *International Symposium on Information and Communication Technology, SoICT 2018*, December 2018.
11. S. Aqel, A. Aarab, and M. Sabri. Traffic video surveillance: Background modeling and shadow elimination. *International Conference on Information Technology for Organizations Development, IT4OD 2016*, pages 1–6, 2016.
12. S. Aqel, M. Sabri, and A. Aarab. Background modeling algorithm based on transitions intensities. *International Review on Computers and Software, IRECOS*, 10(4), 2015.
13. N. Arshad, K. Moon, and J.Kim. Multiple ship detection and tracking using background registration and morphological operations. *International conference on Signal Processing and Multimedia Communication*, 123:121–126, 2010.
14. N. Arshad, K. Moon, and J.Kim. An adaptive moving ship detection and tracking based on edge information and morphological operations. *SPIE International Conference on Graphic and Image Processing, ICGIP*, 2011.
15. N. Arshad, K. Moon, and J.Kim. Adaptive real-time ship detection and tracking using morphological operations. *Journal of Information Communication Convergence Engineering*, 3(12):168–172, September 2014.

16. M. Ashraphijuo, V. Aggarwal, and X. Wang. On Deterministic Sampling Patterns for Robust Low-Rank Matrix Completion. *IEEE Signal Processing Letters*, 25(3):343–347, March 2018.
17. N. Avinash, M. Shashi Kumar, and S. Sagar. Automated video surveillance for retail store statistics generation. *International Conference on Signal and Image Processing, ICSIP*, pages 585–596, 2012.
18. M. Babaee, D. Dinh, and G. Rigoll. A deep convolutional neural network for background subtraction. *Preprint*, 2017.
19. Z. Babic, R. Pilipovic, V. Risojevic, and G. Mirjanic. Pollen bearing honey bee detection in hive entrance video recorded by remote embedded system for pollination monitoring. *ISPRS Congress*, July 2016.
20. F. El Baf and T. Bouwmans. Comparison of background subtraction methods for a multimedia learning space. *International Conference on Signal Processing and Multimedia, SIGMAP 2007*, July 2007.
21. F. El Baf and T. Bouwmans. Robust fuzzy statistical modeling of dynamic backgrounds in ir videos. *SPIE News room*, January 2009.
22. F. El Baf, T. Bouwmans, and B. Vachon. Comparison of background subtraction methods for a multimedia application. *International Conference on systems, Signals and Image Processing, IWSSIP 2007*, pages 385–388, June 2007.
23. F. El Baf, T. Bouwmans, and B. Vachon. Comparison of background subtraction methods for a multimedia application. *International Conference on systems, Signals and Image Processing, IWSSIP 2007*, pages 385–388, June 2007.
24. F. El Baf, T. Bouwmans, and B. Vachon. Foreground detection using the Choquet integral. *International Workshop on Image Analysis for Multimedia Interactive Integral, WIAMIS 2008*, pages 187–190, May 2008.
25. F. El Baf, T. Bouwmans, and B. Vachon. Fuzzy foreground detection for infrared videos. *IEEE International Conference on Computer Vision and Pattern Recognition, CVPR-Workshop OTCBVS 2008*, pages 1–6, June 2008.
26. F. El Baf, T. Bouwmans, and B. Vachon. Fuzzy integral for moving object detection. *IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2008*, pages 1729–1736, June 2008.
27. F. El Baf, T. Bouwmans, and B. Vachon. Type-2 fuzzy mixture of Gaussians model: Application to background modeling. *International Symposium on Visual Computing, ISVC 2008*, pages 772–781, December 2008.
28. F. El Baf, T. Bouwmans, and B. Vachon. Fuzzy statistical modeling of dynamic backgrounds for moving object detection in infrared videos. *IEEE International Conference on Computer Vision and Pattern Recognition, CVPR-Workshop OTCBVS 2009*, pages 60–65, June 2009.
29. F. El Baf, T. Bouwmans, and B. Vachon. Fuzzy statistical modeling of dynamic backgrounds for moving object detection in infrared videos. *IEEE International Conference on Computer Vision and Pattern Recognition, CVPR-Workshop OTCBVS 2009*, pages 60–65, June 2009.
30. O. Barnich and M. Van Droogenbroeck. ViBe: a powerful random technique to estimate the background in video sequences. *International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2009*, pages 945–948, April 2009.
31. R. Bastos. Monitoring of human sport activities at sea. *Master Thesis, Portugal*, 2015.
32. J. Batista, P. Peixoto, C. Fernandes, and M. Ribeiro. A dual-stage robust vehicle detection and tracking for real-time traffic monitoring. *IEEE Intelligent Transportation Systems Conference, ITSC 2006*, 2006.
33. C. Bautista, C. Dy, M. Manalac, and R. Orbe andM. Cordel. Convolutional neural network for vehicle detection in low resolution traffic videos. *TENCON 2016*, 2016.
34. S. Beery, G. Horn, and P. Peron. Recognition in Terra Incognita. *ECCV 2018*, 2018.
35. Y. Benezeth, D. Sidibe, and J. Thomas. Background subtraction with multispectral video sequences. *IEEE International Conference on Robotics and Automation, ICRA 2014*, 2014.
36. C. Bibby and I. Reid. Robust real-time visual tracking using pixel-wise posteriors. *European Conference on Computer Vision, ECCV 2008*, page 831844, 2008.
37. G. Bilodeau, J. Jodoin, and N. Saunier. Change detection in feature space using local binary similarity patterns. *Conference on Computer and Robot Vision, CRV 2013*, pages 106–112, May 2013.
38. P. Blauensteiner and M. Kampel. Visual surveillance of an airport's apron - an overview of the AVITRACK project. *Workshop of the Austrian Association for Pattern Recognition, AAPR 2004*, pages 213–220, 2004.
39. D. Bloisi and L. Iocchi. Independent multimodal background subtraction. *International Conference on Computational Modeling of Objects Presented in Images: Fundamentals, Methods and Applications*, pages 39–44, 2012.

40. D. Bloisi, A. Pennisi, and L. Iocchi. Background modeling in the maritime domain. *Machine Vision and Applications*, 25(5):1257–1269, July 2014.
41. M. Boninsegna and A. Bozzoli. A tunable algorithm to update a reference image. *Signal Processing: Image Communication*, 16(4):1353–365, 2000.
42. A. Borghgraef, O. Barnich, F. Lapierre, M. Droogenbroeck, W. Philips, and M. Acheroy. An evaluation of pixel-based methods for the detection of floating objects on the sea surface. *EURASIP Journal on Advances in Signal Processing*, 2010.
43. T. Boult, R. Micheals, X. Gao, and M. Eckmann. Into the woods: visual surveillance of noncooperative and camouflaged targets in complex outdoor settings. *Proceedings of the IEEE*, 89(10):1382–1402, October 2001.
44. T. Bouwmans. Subspace Learning for Background Modeling: A Survey. *Recent Patents on Computer Science, RPCS 2009*, 2(3):223–234, November 2009.
45. T. Bouwmans. Recent Advanced Statistical Background Modeling for Foreground Detection: A Systematic Survey. *Recent Patents on Computer Science, RPCS 2011*, 4(3):147–176, September 2011.
46. T. Bouwmans. Background Subtraction For Visual Surveillance: A Fuzzy Approach. *Chapter 5, Handbook on Soft Computing for Video Surveillance, Taylor and Francis Group, S.K. Pal, A. Petrosino, L. Maddalena*, pages 103–139, March 2012.
47. T. Bouwmans. Traditional and recent approaches in background modeling for foreground detection: An overview. *Computer Science Review*, 11(31-66), May 2014.
48. T. Bouwmans, F. El Baf, and B. Vachon. Background Modeling using Mixture of Gaussians for Foreground Detection - A Survey. *Recent Patents on Computer Science, RPCS 2008*, 1(3):219–237, November 2008.
49. T. Bouwmans, F. El Baf, and B. Vachon. Statistical Background Modeling for Foreground Detection: A Survey. *Part 2, Chapter 3, Handbook of Pattern Recognition and Computer Vision, World Scientific Publishing, Prof C.H. Chen*, 4:181–199, January 2010.
50. T. Bouwmans, S. Javed, H. Zhang, Z. Lin, and R. Otazo. ”on the applications of robust pca in image and video processing”. *Proceedings of IEEE*, July 2018.
51. T. Bouwmans, Z. Javed, M. Sultana, and S. Jung. Deep neural network concepts for background subtraction: A systematic review and comparative evaluation. *Neural Networks*, 2019.
52. T. Bouwmans, L. Maddalena, and A. Petrosino. Scene background initialization: a taxonomy. *Pattern Recognition Letters*, January 2017.
53. T. Bouwmans, C. Silva, C. Marghes, M. Zitouni, H. Bhaskar, and C. Frelicot. On the role and the importance of features for background modeling and foreground detection. *Computer Science Review*, 28(26-91), November 2018.
54. T. Bouwmans, A. Sobral, S. Javed, S. Jung, and E. Zahzah. Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset. *Computer Science Review*, November 2016.
55. T. Bouwmans and E. Zahzah. Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance. *Special Isssue on Background Models Challenge, Computer Vision and Image Understanding, CVIU 2014*, 122:22–34, May 2014.
56. M. Braham and M. Van Droogenbroeck. Deep background subtraction with scene-specific convolutional neural networks. *International Conference on Systems, Signals and Image Processing, IWSSIP 2016*, pages 1–4, May 2016.
57. M. Braham, S. Pierard, and M. Van Droogenbroeck. Semantic background subtraction. *IEEE International Conference on Image Processing, ICIP 2017*, 2017.
58. S. Bucak and B. Gunsel. Incremental subspace learning and generating sparse representations via non-negative matrix factorization. *Pattern Recognition*, 2008.
59. S. Bucak, B. Gunsel, and O. Gursoy. Incremental non-negative matrix factorization for dynamic background modeling. *International Workshop on Pattern Recognition in Information Systems*, June 2007.
60. C. Buehler, W. Matusik, L. McMillan, and S. Gortler. Creating and rendering image-based visual hulls. *MIT Technical Report 780*, March 1999.
61. X. Burgos-Artizzu, P. Dollar, D. Lin, D. Anderson, and P. Perona. Social behavior recognition in continuous videos. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2012*, 2012.
62. D. Butler, S. Shridharan, and V. Bove. Real-time adaptive background segmentation. *International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2003*, 2003.

63. J. Campbell, L. Mummert, and R. Sukthankar. Video monitoring of honey bee colonies at the hive entrance. *ICPR Workshop on Visual Observation and Analysis of Animal and Insect Behavior, VAIB 2008*, December 2008.
64. M. Camplani, L. Maddalena, G. Moya Alcover, A. Petrosino, and L. Salgado. http://rgbd2017.na.icar.cnr.it/SBM-RGBDchallenge.html. *Workshop on Background learning for detection and tracking from RGBD videos, RGBD2017 in conjunction with ICIAP 2017*, 2017.
65. E. Candes, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *International Journal of ACM*, 58(3), May 2011.
66. J. Carranza, C. Theobalt, M. Magnor, and H. Seidel. Free-viewpoint video of human actors. *ACM Transactions on Graphics*, 22(3):569–577, 2003.
67. M. Chacon-Muguia and S. Gonzalez-Duarte. An adaptive neural-fuzzy approach for object detection in dynamic backgrounds for surveillance systems. *IEEE Transactions on Industrial Electronics*, 2011.
68. M. Chacon-Muguia, S. Gonzalez-Duarte, and P. Vega. Simplified SOM-neural model for video segmentation of moving objects. *International Joint Conference on Neural Networks, IJCNN 2009*, pages 474–480, 2009.
69. M. Chacon-Murguia, G. Ramirez-Alonso, and S. Gonzalez-Duarte. Improvement of a neural-fuzzy motion detection vision model for complex scenario conditions. *International Joint Conference on Neural Networks, IJCNN 2013*, August 2013.
70. S. Chakraborty, M. Paul, M. Murshed, and M. Ali. An efficient video coding technique using a novel non-parametric background model. *IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2014*, pages 1–6, July 2014.
71. S. Chakraborty, M. Paul, M. Murshed, and M. Ali. A novel video coding scheme using a scene adaptive non-parametric background model. *IEEE International Conference on Multimedia Signal Processing*, pages 1–6, September 2014.
72. S. Chakraborty, M. Paul, M. Murshed, and M. Ali. Adaptive weighted non-parametric background model for efficient video coding. *Neurocomputing*, 226:35–45, 2017.
73. K. Chan. Detection of swimmer based on joint utilization of motion and intensity information. *IAPR Conference on Machine Vision Applications*, June 2011.
74. K. Chan. Detection of swimmer using dense optical flow motion map and intensity information. *Machine Vision and Applications*, 24(1):75–101, January 2013.
75. Y. Chan. Deep learning-based scene-awareness approach for intelligent change detection in videos. *Journal of Electronic Imaging*, 28(1):013038, February 2019.
76. T. Chang, T. Ghandi, and M. Trivedi. Vision modules for a multi sensory bridge monitoring approach. *International Conference on Intelligent Transportation Systems, ITSC 2004*, pages 971–976, October 2004.
77. G. Chau and P. Rodriguez. Panning and jitter invariant incremental principal component pursuit for video background modeling. *International Workshop RSL-CV 2017 in conjunction with ICCV 2017*, October 2017.
78. B. Chen and S. Huang. An advanced moving object detection algorithm for automatic traffic monitoring in real-world limited bandwidth networks. *IEEE Transactions on Multimedia*, 16(3):837–847, April 2014.
79. C. Chen, J. Cai, W. Lin, and G. Shi. Surveillance video coding via low-rank and sparse decomposition. *ACM international conference on Multimedia*, pages 713–716, 2012.
80. W. Chen, X. Zhang, Y. Tian, and T. Huang. An efficient surveillance coding method based on a timely and bit-saving background updating model. *IEEE International Conference on Visual Communication and Image Processing, VCIP 2012*, November 2012.
81. Z. Chen and T. Ellis. A self-adaptive Gaussian mixture model. *Computer Vision and Image Computing, CVIU 2014*, 122:3546, May 2014.
82. Z. Chen, J. Yang, and Z. Kang. Moving ship detection algorithm based on Gaussian mixture model. *International Conference on Modelling, Simulation and Applied Mathematics*, pages 197–201, 2018.
83. S. Chien, S. Ma, and L. Chen. Efficient moving object segmentation algorithm using background registration technique. *IEEE Transaction on Circuits and Systems for Video Technology*, 12(7):577–586, July 2002.
84. P. Chiranjeevi and S. Sengupta. Neighborhood supported model level fuzzy aggregation for moving object segmentation. *IEEE Transactions on Image Processing*, 23(2):645–657, February 2014.
85. J. Cho, J. Park, U. Baek, D. Hwang, S. Choi, S. Kim, and K. Kim. Automatic parking system using background subtraction with CCTV environment. *International Conference on Control, Automation and Systems, ICCAS 2016*, pages 1649–1652, 2016.

86. C. Chu, O. Jenkins, and M. Mataric. Markerless kinematic model and motion capture from volume sequences. *IEEE Computer Vision and Pattern Recognition, CVPR 2003*, 2003.
87. M. Chunyang, M. Xing, Z. Panpan, M. Chunyang andM. Xing, and Z. Panpan. Smart detection of vehicle in illegal parking area by fusing of multi-features. *International Conference on Next Generation Mobile Applications, Services and Technologies*, pages 388–392, 2015.
88. G. Cinar and J. Principe. Adaptive background estimation using an information theoretic cost for hidden state estimation. *International Joint Conference on Neural Networks, IJCNN 2011*, August 2011.
89. L. Pinheiro Cinelli. Anomaly detection in surveillance videos using deep residual networks. *Master Thesis, Universidade de Rio de Janeiro*, February 2017.
90. C. Cohen, D. Haanpaa, S. Rowe, and J. Zott. Vision algorithms for automated census of animals. *IEEE Applied Imagery Pattern Recognition Workshop, AIPR 2011*, pages 1–5, 2011.
91. C. Cohen, D. Haanpaa, and J. Zott. Machine vision algorithms for robust animal species identification. *IEEE Applied Imagery Pattern Recognition Workshop, AIPR 2015*, pages 1–7, 2015.
92. R. Collins, A. Lipton, T.Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt, and L. Wixson. A system for video surveillance and monitoring. *Technical Report, Carnegie Mellon University:The Robotics Institute*, 2000.
93. R. Collins, X. Zhou, and S. Teh. An open source tracking testbed and evaluation web site. *IEEE International Workshop on PETS 2005*, 2005.
94. A. Colombari, A. Fusiello, and V. Murino. Exemplar based background model initialization. *VSSN 2005*, November 2005.
95. R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirottia. Improving shadow suppression in moving object detection with HSV color information. *IEEE Intelligent Transportation Systems*, pages 334–339, 2001.
96. C. Cuevas, R. Martinez, and N. Garcia. Detection of stationary foreground objects: A survey. *Computer Vision and Image Understanding*, 2016.
97. D. Culibrk, O. Marques, D. Socek, H. Kalva, and B. Furht. A neural network approach to Bayesian background modeling for video object segmentation. *International Conference on Computer Vision Theory and Applications, VISAPP 2006*, February 2006.
98. D. Cullen. Detecting and summarizing salient events in coastal videos. *TR 2012-06, Boston University*, May 2012.
99. D. Cullen, J. Konrad, and T. Little. Detection and summarization of salient events in coastal environments. *IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2012*, 2012.
100. K. Dahiya, D. Singh, and C. Mohan. Automatic detection of bike-riders without helmet using surveillance videos in real-time. *International Joint Conference on Neural Networks*, pages 3046–3051, 2016.
101. A. de Geus, M. Batista, M. Rabelo, C. Barcelos, and S. da Silva. Maize Insects Classification Through Endoscopic Video Analysis. *Canadian Conference on Artificial Intelligence, AI 2019*, pages 251–262, 2019.
102. J. Dey and N. Praveen. Moving object detection using genetic algorithm for traffic surveillance. *International Conference on Electrical, Electronics, and Optimization Techniques, ICEEOT 2016*, pages 2289–2293, 2016.
103. R. Diaz, S. Hallman, and C. Fowlkes. Detecting dynamic objects with multi-view background subtraction. *International Conference on Computer Vision, ICCV 2013*, 2013.
104. P. Dickinson. Monitoring the vulnerable using automated visual surveillance. *PhD thesis, University of Lincoln, UK*, 2008.
105. P. Dickinson, R. Freeman, S. Patrick, and S. Lawson. Autonomous monitoring of cliff nesting seabirds using computer vision. *International Workshop on Distributed Sensing and Collective Intelligence in Biodiversity Monitoring*, 2008.
106. P. Dickinson, C. Qing, S. Lawson, and R. Freeman. Automated visual monitoring of nesting seabirds. *ICPR Workshop on Visual Observation and Analysis of Animal and Insect Behavior, VAIB 2010*, 2010.
107. R. Ding, X. Liu, W. Cui, and Y. Wang. Intersection foreground detection based on the cyber-physical systems. *IET International Conference on Information Science and Control Engineering, ICISCE 2012*, pages 1–7, 2012.
108. Y. Dong and G. DeSouza. Adaptive learning of multi-subspace for foreground detection under illumination changes. *Computer Vision and Image Understanding*, 2011.

109. Y. Dong, T. Han, and G. DeSouza. Illumination invariant foreground detection using multi-subspace learningg. *Journal International of Knowledge-Based and Intelligent Engineering Systems,*, 14(1):31–41, 2010.
110. A. Elgammal, D. Harwood, and L. Davis. Non-parametric model for background subtraction. *Frame Rate Workshop, IEEE International Conference on Computer Vision, ICCV 1999*, September 1999.
111. A. Elqursh and A. Elgammal. Online moving camera background subtraction. *European Conference on Computer Vision, ECCV 2012*, 2012.
112. R. Elsayed, M. Sayed, and M. Abdalla. Skin-based adaptive background subtraction for hand gesture segmentation. *IEEE International Conference on Electronics, Circuits, and Systems, ICECS 2015*, 35:33–36, 2015.
113. A. ElTantawy and M. Shehata. Moving object detection from moving platforms using Lagrange multiplier. *IEEE International Conference on Image Processing, ICIP 2015*, 2015.
114. A. ElTantawy and M. Shehata. UT-MARO: unscented transformation and matrix rank optimization for moving objects detection in aerial imagery. *International Symposium on Visual Computing, ISVC 2015*, pages 275–284, 2015.
115. A. ElTantawy and M. Shehata. A novel method for segmenting moving objects in aerial imagery using matrix recovery and physical spring model. *International Conference on Pattern Recognition, ICPR 2016*, pages 3898–3903, 2016.
116. A. ElTantawy and M. Shehata. KRMARO: aerial detection of small-size ground moving objects using kinematic regularization and matrix rank optimization. *IEEE Transactions on Circuits and Systems for Video Technolog*, June 2018.
117. A. ElTantawy and M. Shehata. MARO: matrix rank optimization for the detection of small-size moving objects from aerial camera platforms. *Signal, Image and Video Processing*, 12(4):641–649, May 2018.
118. H. Eng, K. Toh, A. Kam, J. Wang, and W. Yau. An automatic drowning detection surveillance system for challenging outdoorpool environments. *IEEE International Conference on Computer Vision, ICCV 2003*, pages 532–539, 2003.
119. H. Eng, K. Toh, A. Kam, J. Wang, and W. Yau. Novel region-based modeling for human detection within highly dynamic aquatic environment. *IEEE International Conference onComputer Vision and Pattern Recognition, CVPR 2004*, 2:390, 2004.
120. N. Ershadi, J. Menendez, and D. Jimenez. Robust vehicle detection in different weather conditions: Using mipm. *PLOS 2018*, 2018.
121. D. Farcas, C. Marghes, and T. Bouwmans. Background subtraction via incremental maximum margin criterion: A discriminative approach. *Machine Vision and Applications*, 23(6):1093–1101, October 2012.
122. A. Faro, D. Giordano, and C. Spampinato. Adaptive background modeling integrated with luminosity sensors and occlusion processing for reliable vehicle detection. *IEEE Transactions on Intelligent Transportation Systems*, 12(4):1398–1412, December 2011.
123. B. Farou, M. Kouahla, H. Seridi, and H. Akdag. Efficient local monitoring approach for the task of background subtraction. *Engineering Applications of Artificial Intelligence*, 64:1–12, September 2017.
124. L. Fei, W. Xueli, and C. Dongsheng. Drowning detection based on background subtraction. *International Conference on Embedded Software and Systems, ICESS 2009*, pages 341–343, 2009.
125. V. Fernandez-Arguedas, G. Pallotta, and M. Vespe. Maritime Traffic Networks: From Historical Positioning Data to Unsupervised Maritime Traffic Monitoring. *IEEE Transactions on Intelligent Transportation Systems*, 19(3):722–732, March 2018.
126. N. Friedman and S. Russell. Image segmentation in video sequences: A probabilistic approach. *Conference on Uncertainty in Artificial Intelligence, UAI 1997*, pages 175–181, 1997.
127. Z. Gao, W. Wang, H. Xiong, W. Yu, X. Liu, W. Liu, and L. Pei. Traffic surveillance using highly-mounted video camera. *IEEE International Conference on Progress in Informatics and Computing*, pages 336–340, 2014.
128. D. Garcia-Lesta, V. Brea, P. Lopez, and D. Cabello. Impact of analog memories non-idealities on the performance of foreground detection algorithms. *IEEE International Symposium on Circuits and Systems, ISCAS 2018*, pages 1–5, 2018.
129. M. Geng, X. Zhang, Y. Tian, L. Liang, and T. Huang. A fast and performance-maintained transcoding method based on background modeling for surveillance video. *IEEE International Conference on Multimedia and Expo, ICME 2012*, pages 61–67, July 2012.

130. J. Giraldo-Zuluaga, A. Gomez, A. Salazar, and A. Diaz-Pulido. Automatic recognition of mammal genera on camera-trap images using multi-layer robust principal component analysis and mixture neural networks. *Preprint*, 2017.
131. J. Giraldo-Zuluaga, A. Gomez, A. Salazar, and A. Diaz-Pulido. Camera-trap images segmentation using multi-layer robust principal component analysis. *Preprint*, 2017.
132. K. Goehner, T. Desell, and al. A comparison of background subtraction algorithms for detecting avian nesting events in uncontrolled outdoor video. *IEEE International Conference on e-Science*, pages 187–195, 2015.
133. K. Goyal and J. Singhai. Review of background subtraction methods using Gaussian mixture model for video surveillance systems. *Artificial Intelligence Review*, 50(2):241–259, 2018.
134. N. Goyette, P. Jodoin, F. Porikli, J. Konrad, and P. Ishwar. Changedetection.net: A new change detection benchmark dataset. *IEEE Workshop on Change Detection, CDW 2012 in conjunction with CVPR 2012*, June 2012.
135. C. Gu, D. Tian, Y. Cong, Y. Zhang, and S. Wang. Flame detection based on spatio-temporal covariance matrix. *International Conference on Advanced Robotics and Mechatronics, ICARM 2016*, 2016.
136. G. Guerra-Filho. Optical motion capture: Theory and implementation. *Journal of Theoretical and Applied Informatics*, 2(12):61–89, 2005.
137. H. Guo, C. Qiu, and N. Vaswani. Practical ReProCS for separating sparse and low-dimensional signal sequences from their sum - part 1. *International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2014*, May 2014.
138. H. Guo, C. Qiu, and N. Vaswani. Practical ReProCS for separating sparse and low-dimensional signal sequences from their sum - part 2. *GlobalSIP 2014*, 2014.
139. R. Guo and H. Qi. Partially-sparse restricted Boltzmann machine for background modeling and subtraction. *International Conference on Machine Learning and Applications, ICMLA 2013*, pages 209–214, December 2013.
140. X. Guo, S. Li, and X. Cao. Motion matters: A novel framework for compressing surveillance videos. *ACM International Conference on Multimedia*, October 2013.
141. Y. Guo, Z. Zhang, D. He, J. Niu, and Y Tan. Detection of cow mounting behavior using region geometry and optical flow characteristics. *Computers and Electronics in Agriculture*, 163, June 2019.
142. Y. Guo, W. Zhu, P. Jiao, and J. Chen. Foreground detection of group-housed pigs based on the combination of mixture of gaussians using prediction mechanism and threshold segmentation. *Biosystems Engineering*, 125:98–104, September 2014.
143. C. Guyon, T. Bouwmans, and E. Zahzah. Foreground detection by robust PCA solved via a linearized alternating direction method. *International Conference on Image Analysis and Recognition, ICIAR 2012*, June 2012.
144. C. Guyon, T. Bouwmans, and E. Zahzah. Robust Principal Component Analysis for Background Subtraction: Systematic Evaluation and Comparative Analysis. *Book 1, Chapter 12, Principal Component Analysis, INTECH*, pages 223–238, March 2012.
145. M. Habermann, W. Xu, M. Zollhofer, G. Pons-Moll, and C. Theobalt. LiveCap: Real-time Human Performance Capture from Monocular Video. *Preprint*, January 2019.
146. R. Hadi, L. George, and M. Mohammed. A computationally economic novel approach for real-time moving multi-vehicle detection and tracking toward efficient traffic surveillance. *Arabian Journal for Science and Engineering*, 42(2):817–831, February 2017.
147. R. Hadi, G. Sulong, and L. George. An innovative vehicle detection approach based on background subtraction and morphological binary methods. *Life Science Journal*, pages 230–238, 2014.
148. M. Hadiuzzaman, N. Haque, F. Rahman, S. Hossain, M. Rayeedul, K. Siam, and T. Qiu. Pixel-based heterogeneous traffic measurement considering shadow and illumination variation. *Signal, Image and Video Processing*, pages 1–8, April 2017.
149. T. Haines and T. Xiang. Background subtraction with Dirichlet processes. *European Conference on Computer Vision, ECCV 2012*, October 2012.
150. S. Ham, B. Ko, and J. Nam. Fire-flame detection based on fuzzy finite automation. *International Conference on Pattern Recognition, ICPR 2010*, 2010.
151. G. Han, J. Wang, and X. Ca. Improved visual background extractor using an adaptive distance threshold. *Journal of Electronic Imaging*, 23(6), December 2014.
152. S. Han, X. Zhang, Y. Tian, and T. Huang. An efficient background reconstruction based coding method for surveillance videos captured by moving camera. *IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2012*, pages 160–165, September 2012.

153. J. Hao, C. Li, Z. Kim, and Z. Xiong. Spatio-temporal traffic scene modeling for object motion detection. *IEEE Transactions on Intelligent Transportation Systems*, 1(14):295–302, March 2013.
154. J. Hao, C. Li, Z. Kim, and Z. Xiong. Spatio-temporal traffic scene modeling for object motion detection. *IEEE Transactions on Intelligent Transportation Systems*, 1(14):295–302, March 2013.
155. M. Haque, M. Murshed, and M. Paul. On stable dynamic background generation technique using Gaussian mixture models for robust object detection. *IEEE International Conference On Advanced Video and Signal Based Surveillance, AVSS 2008*, pages 41–48, September 2008.
156. S. Hargude and S. Idate. i-surveillance: Intelligent surveillance system using background subtraction technique. *International Conference on Computing Communication Control and automation, ICCUBEA 2016*, pages 1–5, 2016.
157. I. Haritaoglu, D. Harwood, and L. Davis. W4: Who? When? Where? What? A Real Time System for Detecting and Tracking People. *IEEE International Conference on Automatic Face and GestureRecognition*, April 1998.
158. M. Heikkila and M. Pietikainen. A texture-based method for modeling the background and detecting moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI 2006*, 28(4):657–62, 2006.
159. K. Higa and K. Iwamoto. Estimation of Product Amount on Store Shelves using Image Change Classification based on Background Subtraction and CNN. *Forum on Information Technology, FIT 2017*, September 2017.
160. K. Higa and K. Iwamoto. Robust estimation of product amount on store shelves from a surveillance camera for improving on-shelf availability. *IEEE International Conference on Imaging Systems and Techniques, ICIST 2018*, October 2018.
161. M. Hofmann, P. Tiefenbacher, and G. Rigoll. Background segmentation with feedback: The pixel-based adaptive segmenter. *IEEE Workshop on Change Detection, CVPR 2012*, June 2012.
162. W. Hong, A. Kennedy, X. Burgos-Artizzu, M. Zelikowsky, S. Navonne, P. Perona, and D. Anderson. Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning. *PNAS*, 2015.
163. T. Horprasert, I. Haritaoglu, C. Wren, D. Harwood, L. Davis, and A. Pentland. Real-time 3D motion capture. *Workshop on Perceptual User Interfaces, PUI 1998*, pages 87–90, 1998.
164. T. Horprasert, D. Harwood, and L. Davis. A statistical approach for real-time robust background subtraction and shadow detection. *International Conference on Computer Vision, FRAME-RATE Workshop, ICCV 1999*, September 1999.
165. T. Horprasert, D. Harwood, and L. Davis. A statistical approach for real-time robust background subtraction and shadow detection. *IEEE International Conference on Computer Vision, FRAME-RATE Workshop*, September 1999.
166. T. Horprasert, D. Harwood, and L. Davis. A robust background subtraction and shadow detection. *Asian Conference on Computer Vision, ACCV 2009*, page January, 2000.
167. W. Hu, C. Yang, and D. Huang. Robust real-time ship detection and tracking for visual surveillance of cage aquaculture. *Journal of Visual Communication and Image Representation*, 22(6):543–556, August 2011.
168. Z. Hu, T. Turki, N. Phan, and J. Wang. 3d atrous convolutional long short-term memory network for background subtraction. *IEEE Access*, 2018.
169. H. Huang, X. Fang, Y. Ye, S. Zhang, and P. Rosin. Practical automatic background substitution for live video. *Computational Visual Media*, 2017.
170. I. Huang, J. Hwang, and C. Rose. Chute based automated fish length measurement and water drop detection. *IEEE International Conference on Acoustics,Speech and Signal Processing, ICASSP 2016*, pages 1906–1910, 2016.
171. J. Huang, X. Huang, and D. Metaxas. Learning with dynamic group sparsity. *International Conference on Computer Vision, ICCV 2009*, October 2009.
172. L. Huang, Q. Chen, J. Lin, and H. Lin. Block background subtraction method based on ViBe. *Applied Mechanics and Materials*, May 2014.
173. S. Huang and B. Chen. Highly accurate moving object detection in variable bit rate video-based traffic monitoring systems. *IEEE Transactions on Neural Networks and Learning Systems*, 12(24):1920–1231, December 2013.
174. T. Huang, J. Hwang, S. Romain, and F. Wallace. Live tracking of rail-based fish catching on wild sea surface. *Workshop on Computer Vision for Analysis of Underwater Imagery in conjuction with ICPR 2016, CVAUI 2016*, 2016.

175. W. Huang, Q. Zeng, and M. Chen. Motion characteristics estimation of animals in video surveillance. *IEEE Advanced Information Technology, Electronic and Automation Control Conference, IAEAC 2017*, pages 1098–1102, 2017.
176. M. Hung, J. Pan, and C. Hsieh. Speed up temporal median filter for background subtraction. *International Conference on Pervasive Computing, Signal Processing and Applications*, pages 297–300, 2010.
177. P. Hwang, K. Eom, J. Jung, and M. Kim. A statistical approach to robust background subtraction for urban traffic video. *International Workshop on Computer Science and Engineering*, pages 177–181, 2009.
178. T. Intachak and W. Kaewapichai. Real-time illumination feedback system for adaptive background subtraction working in traffic video monitoring. *International Symposium on Intelligent Signal Processing and Communication Systems, ISPACS 2011*, 2011.
179. Y. Iwatani, K. Tsurui, and A. Honma. Position and direction estimation of wolf spiders, pardosa astrigera, from video images. *International Conference on Robotics and Biomimetics, ICRB 2016*, December 2016.
180. N. Jain, R. Saini, and P. Mittal. A Review on Traffic Monitoring System Techniques. *Soft Computing: Theories and Applications*, pages 569–577, 2018.
181. M. Janzen, K. Visser, D. Visscher, I. Mac Leod, D. Vujnovic, and K. Vujnovic. Semi-automated camera trap image processing for the detection of ungulate fence crossing events. *Environmental Monitoring and Assessment*, 189(10):527, 2017.
182. S. Javed, T. Bouwmans, and S. Jung. Combining ARF and OR-PCA background subtraction of noisy videos. *International Conference in Image Analysis and Applications, ICIAP 2015*, September 2015.
183. S. Javed, T. Bouwmans, and S. Jung. Stochastic decomposition into low rank and sparse tensor for robust background subtraction. *ICDP 2015*, July 2015.
184. S. Javed, T. Bouwmans, and S. Jung. Improving or-pca via smoothed spatially-consistent low-rank modeling for background subtraction. *ACM Symposium on Applied Computing, SAC 2017*, 2017.
185. S. Javed, A. Mahmood, T. Bouwmans, and S. Jung. Motion-Aware Graph Regularized RPCA for Background Modeling of Complex Scenes. *Scene Background Modeling Contest, International Conference on Pattern Recognition, ICPR 2016*, December 2016.
186. S. Javed, A. Mahmood, T. Bouwmans, and S. Jung. Spatiotemporal Low-rank Modeling for Complex Scene Background Initialization. *IEEE Transactions on Circuits and Systems for Video Technology*, December 2016.
187. S. Javed, A. Mahmood, T. Bouwmans, and S. Jung. Background-foreground modeling based on spatiotemporal sparse subspace clustering. *IEEE Transactions on Image Processing*, September 2017.
188. S. Javed, A. Mahmood, T. Bouwmans, and S. Jung. Superpixels based manifold structured sparse RPCA for moving object detection. *International Workshop on Activity Monitoring by Multiple Distributed Sensing, BMVC 2017*, September 2017.
189. S. Javed, A. Mahmood, T. Bouwmans S. Al-Maadeed, and S. Jung. Moving object detection in complex scene using spatiotemporal structured-sparse RPCA. *IEEE Transactions on Image Processing*, 28(2):1007–1022, 2019.
190. S. Javed, S. Oh, and S. Jung. Ipbas: Improved pixel based adaptive background segmenter for background subtractio. *Conference: Human Computer Interaction*, January 2014.
191. S. Javed, S. Oh, A. Sobral, T. Bouwmans, and S. Jung. Background subtraction via superpixel-based online matrix decomposition with structured foreground constraints. *Workshop on Robust Subspace Learning and Computer Vision, ICCV 2015*, December 2015.
192. I. Jeon and S. Yoo. Spatial kernel bandwith estimation in background modeling. *SPIE International Conference on Machine Vision, ICMV 2016*, 2016.
193. Q. Jiang, G. Li, J. Yu, and X. Li. A model based method for pedestrian abnormal behavior detection in traffic scene. *IEEE International Smart Cities Conference, ISC2 2015*, pages 1–6, 2015.
194. P. Jodoin, L. Maddalena, and A. Petrosino. Scene background initialization: a taxonomy. *IEEE Transactions on Image Processing*, 2017.
195. P. Jodoin, V. Saligrama, and J. Konrad. Behavior subtraction. *IEEE Transactions on Image Processing*, 2011.
196. F. John, I. Hipiny, H. Ujir, and M. Sunar. Assessing performance of aerobic routines using background subtraction and intersected image region. *Preprint*, 2018.
197. S. Ju and X. Chen andG. Xu. An improved mixture Gaussian models to detect moving object under real-time complex background. *International Conference on Cyberworlds, ICC 2008*, pages 730–734, 2008.

198. I. Junejo, A. Bhutta, and H. Foroosh. Dynamic scene modeling for object detection using single-class SVM. *International Conference on Image Processing, ICIP 2010*, pages 1541–1544, September 2010.
199. P. KaewTraKulPong and R. Bowden. An improved adaptive background mixture model for real-time tracking with shadow detection. *AVBS 2001*, September 2001.
200. K. Karmann and A. Von Brand. Moving object recognition using an adaptive background memory. *Time-Varying Image Processing and Moving Object Recognition, Elsevier*, 1990.
201. J. Karnowski, E. Hutchins, and C. Johnson. Dolphin detection and tracking. *IEEE Winter Applications and Computer Vision Workshops, WACV 2015*, pages 51–56, 2015.
202. I. Kavasidis and S. Palazzo. Quantitative performance analysis of object detection algorithms on underwater video footage. *ACM international workshop on Multimedia Analysis for Ecological Data, MAED 2012*, pages 57–60, 2012.
203. I. Kavasidis, S. Palazzo, and C. Spampinato. An innovative web-based collaborative platform for video annotation. *Multimedia Tools and Applications*, pages 1–20, 2013.
204. Y. Kawanishi, I. Mitsugami, M. Mukunoki, and M. Minoh. Background image generation keeping lighting condition of outdoor scenes. *International Conference on Security Camera Network, Privacy Protection and Community Safety, SPC2009*, October 2009.
205. H. Khaled, S. Sayed, E. Saad, and H. Ali. Hand gesture recognition using modified 1$ and background subtraction algorithms. *Mathematical Problems in Engineering*, 35:33–36, 2015.
206. P. Khorrami, J. Wang, and T. Huang. Multiple animal species detection using robust principal component analysis and large displacement optical flow. *Workshop on Visual Observation and Analysis of Animal and Insect Behavior, VAIB 2012*, 2012.
207. M. Kilger. A shadow handler in a video-based real-time traffic monitoring system. *IEEE Workshop on Applications of Computer Vision*, pages 1062–1066, 1992.
208. H. Kim, R. Sakamoto, I. Kitahara, N. Orman, T. Toriyama, and K. Kogure. Compensated visual hull for defective segmentation and occlusion. *International Conference on Artificial Reality and Telexistence, ICAT 2007*, 2007.
209. H. Kim, R. Sakamoto, I. Kitahara, T. Toriyama, and K. Kogure. Robust foreground extraction technique using Gaussian family model and multiple thresholds. *Asian Conference on Computer Vision, ACCV 2007*, November 2007.
210. K. Kim, T. Chalidabhongse, D. Harwood, and L. Davis. Background modeling and subtraction by codebook construction. *IEEE International Conference on Image Processing, ICIP 2004*, 2004.
211. T. Kimura, M. Ohashi, K. Crailsheim, T. Schmickl, R. Odaka, and H. Ikeno. Tracking of multiple honey bees on a flat surface. *International Conference on Emerging Trends in Engineering and Technology, ICETET 2012*, pages 36–39, November 2012.
212. B. Kiran and S. Yogamani. Real-Time Background Subtraction Using Adaptive Sampling and Cascade of Gaussians. *Preprint*, 2017.
213. U. Knauer, M. Himmelsbach, F. Winkler, F. Zautke, K. Bienefeld, and B. Meffert. Application of an adaptive background model for monitoring honeybees. *VIIP 2005*, 2005.
214. T. Ko, S. Soatto, and D. Estrin. Background subtraction on distributions. *European Conference on Computer Vision, ECCV 2008*, pages 222–230, October 2008.
215. T. Ko, S. Soatto, and D. Estrin. Warping background subtraction. *IEEE International Conference on Computer Vision and Pattern Recognition, CVPR 2010*, June 2010.
216. D. Koller, J. Weber, T. Huang, J. Malik, G. Ogasawara, B. Rao, and S. Russel. Towards robust automatic traffic scene analysis in real-time. *International Conference on Pattern Recognition, ICPR 1994*, pages 126–131, October 1994.
217. G. Kopsiaftis and K. Karantzalos. Vehicle detection and traffic density monitoring from very high resolution satellite video data. *IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2015*, pages 1881–1884, 2015.
218. T. Kryjak and M. Gorgon. Real-time implementation of the ViBe foreground object segmentation algorithm. *Federated Conference on Computer Science and Information Systems, FedCSIS 2013*, pages 591–596, 2013.
219. T. Kryjak, M. Komorkiewicz, and M. Gorgon. Hardware implementation of the pbas foreground detection method in fpga. *International Conference Mixed Design of Integrated Circuits and Systems , MIXDES 2013*, pages 479–484, 2013.
220. T. Kryjak, M. Komorkiewicz, and M. Gorgon. Real-time foreground object detection combining the pbas background modelling algorithm and feedback from scene analysis module. *International Journal of Electronics and Telecommunications, IJET 2014*, 2014.

221. T. Kryjak, M. Komorkiewicz, and M. Gorgon. Real-time implementation of foreground object detection from a moving camera using the ViBe algorithm. *Computer Science and Information Systems*, 2014.
222. W. Kusakunniran and R. Krungkaew. Dynamic codebook for foreground segmentation in a video. *ECTI Transactions on Computer and Information Technology, ECTI-CIT 2017*, 10(2), 2017.
223. A. Kushwaha and R. Srivastava. A Framework of Moving Object Segmentation in Maritime Surveillance Inside a Dynamic Backgrounds. *Transactions on Computational Science*, pages 35–54, April 2015.
224. A. Lai, H. Yoon, and G. Lee. Robust background extraction scheme using histogram-wise for real-time tracking in urban traffic video. *International Conference on Computer and Information Technology, CIT 2008*, 1944:845–850, July 2008.
225. J. Lan, Y. Jiang, and D. Yu. A new automatic obstacle detection method based on selective updating of Gaussian mixture model. *International Conference on Transportation Information and Safety, ICTIS 2015*, pages 21–25, 2015.
226. B. Laugraud, S. Pierard, and M. Van Droogenbroeck. Labgen-p-semantic: A first step for leveraging semantic segmentation in background generation. *MDPI Journal of Imaging*, 2018.
227. A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2), February 1999.
228. D. Lee. Effective Gaussian mixture learning for video background subtraction. *IEEE Transactions on Pattern Recognition and Machine Intelligence, PAMI 2005*, 27:827–832, 2005.
229. G. Lee, R. Mallipeddi, G. Jang, and M. Lee. A genetic algorithm-based moving object detection for real-time traffic surveillance. *IEEE Signal Processing Letters*, 10(22):1619–1622, 2015.
230. J. Lee, M. Ryoo, M. Riley, and J. Aggarwal. Real-time Detection of Illegally Parked Vehicles using 1-D Transformation. *IEEE Advanced Video and Signal-based Surveillance, AVSS 2007*, 2007.
231. S. Lee and D. Kim. Background subtraction using the factored 3-way restricted boltzmann machines. *Preprint*, 2018.
232. S. Lee, N. Kim, I. Paek, M. Hayes, and J. Paik. Moving object detection using unstable camera for consumer surveillance systems. *International Conference on Consumer Electronics, ICCE 2013*, pages 145–146, January 2013.
233. F. Lei and X. Zhao. Adaptive background estimation of underwater using Kalman-filtering. *International Congress on Image and Signal Processing, CISP 2010*, October 2010.
234. G. Levin. Computer vision for artists and designers: Pedagogic tools and techniques for novice programmers. *Journal of Artificial Intelligence and Society*, 20(4), 2006.
235. A. Leykin and M. Tuceryan. Tracking and activity analysis in retail environments. *TR 620 Computer Science Department, Indiana University*, 2005.
236. A. Leykin and M. Tuceryan. A vision system for automated customer tracking for marketing analysis: Low level feature extraction. *HAREM 2005*, 2005.
237. A. Leykin and M. Tuceryan. Detecting shopper groups in video sequences. *International Conference on Advanced Video and Signal-Based Surveillance, AVSBS 2007*, 2007.
238. C. Li, A. Chiang, G. Dobler, Y. Wang, K. Xie, K. Ozbay, M. Ghandehari, J. Zhou, and D. Wang. Robust vehicle tracking for urban traffic videos at intersections. *IEEE Advanced Video and Signal Based Surveillance, AVSS 2016,*, 2016.
239. L. Li and W. Huang. Statistical modeling of complex background for foreground object detection. *IEEE Transaction on Image Processing*, 13(11):1459–1472, November 2004.
240. L. Li and M. Leung. Integrating intensity and texture differences for robust change detection. *IEEE Transactions on Image Processing*, 11:105–112, February 2002.
241. Q. Li, E. Bernal, M. Shreve, and R. Loce. Scene-independent feature- and classifier-based vehicle headlight and shadow removal in video sequences. *IEEE Winter Applications of Computer Vision Workshops, WACVW 2016*, pages 1–8, 2016.
242. X. Li, M. Ng, and X. Yuan. Median filtering-based methods for static background extraction from surveillance video. *Numerical Linear Algebra with Applications, Wiley Online Library*, March 2015.
243. X. Li and Y. Wu. Image object detection algorithm based on improved Gaussian mixture model. *Journal of Multimedia*, 9(1):152–158, 2014.
244. X. Li, M. Ye, Y. Liu, and C. Zhu. Adaptive deep convolutional neural networks for scene-specific object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, September 2017.
245. Y. Li, G. Wang, and X. Lin. Three-level GPU accelerated gaussian mixture model for background subtraction. *Proceedings of the SPIE*, 2012.

246. K. Lim, W. Jang, and C. Kim. Background subtraction using encoder-decoder structured convolutional neural network. *IEEE International Conference on Advanced Video and Signal based Surveillance, AVSS 2017*, 2017.
247. C. Lin, B. Yan, and W. Tan. Foreground detection in surveillance video with fully convolutional semantic network. *IEEE International Conference on Image Processing, ICIP 2018*, pages 4118–4122, 2018.
248. H. Lin, T. Liu, and J. Chuang. A probabilistic SVM approach for background scene initialization. *International Conference on Image Processing, ICIP 2002*, 3:893–896, September 2002.
249. R. Lin, X. Cao, Y. Xu, C. Wu, and H. Qiao. Airborne moving vehicle detection for video surveillance of urban traffic. *IEEE Intelligent Vehicles Symposium, IVS 2009*, pages 203–208, 2009.
250. Q. Ling, J. Yan, F. Li, and Y. Zhang. A background modeling and foreground segmentation approach based on the feedback of moving objects in traffic surveillance systems. *Neurocomputing*, 2014.
251. I. Lipovac, T. Hrkac, K. Brkic, Z. Kalafatic, and S. Segv. Experimental evaluation of vehicle detection based on background modelling in daytime and night-time video. *Croatian Computer Vision Workshop, CCVW 2014*, 2014.
252. H. Liu, J. Dai, R. Wang, H. Zheng, and B. Zheng. Combining background subtraction and three-frame difference to detect moving object from underwater video. *OCEANS 2016*, pages 1–5, 2016.
253. Y. Liu, H. Yao, W. Gao, X. Chen, and D. Zhao. Nonparametric background generations. *Journal of Visual Communication and Image Representation*, 18(3):253–263, 2007.
254. Z. Liu, F. Zhou, X. Chen, X. Bai, and C. Sun. Iterative infrared ship target segmentation based on multiple features. *Pattern Recognition*, 47(9):28392852, 2014.
255. P. Lu, Q. Liu, F. Teng, L. Me, and J. Li. Inland moving ships detection via compressive sensing and saliency detection. *CSIC 2016*, 2016.
256. X. Lu, T. Izumi, T. Takahashi, and L. Wang. Moving vehicle detection based on fuzzy background subtraction. *IEEE International Conference on Fuzzy Systems*, pages 529–532, 2014.
257. X. Lu, T. Izumi, T. Takahashi, and L. Wang. Moving vehicle detection based on fuzzy background subtraction. *IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2014*, pages 529–532, July 2014.
258. X. Lu, T. Izumi, L. Teng, T. Hori, and L. Wang. A novel background subtraction method for moving vehicle detection. *IEEJ Transactions on Fundamentals and Materials*, 132(10):857–863, 2012.
259. X. Lu, C. Xu, L. Wang, and L. Teng. Improved Background Subtraction Method for Detecting Moving Objects based on GMM. *IEEJ Transactions on Electrical and Electronics Engineering*, June 2018.
260. L. Maddalena and A. Petrosino. A self organizing approach to background subtraction for visual surveillance applications. *IEEE Transactions on Image Processing*, 17(7):1168–1177, July 2008.
261. L. Maddalena and A. Petrosino. The 3dSOBS+ algorithm for moving object detection. *Computer Vision and Image Understanding, CVIU 2014*, 122:65–73, May 2014.
262. L. Maddalena and A. Petrosino. Background model initialization for static cameras. *Handbook on Background Modeling and Foreground Detection for Video Surveillance, CRC Press, Taylor and Francis Group*, 3, July 2014.
263. L. Maddalena and A. Petrosino. Towards benchmarking scene background initialization. *Workshop on Scene Background Modeling and Initialization in conjunction with ICIAP 2015*, 1:469–476, September 2015.
264. L. Maddalena and A. Petrosino. Self-organizing background subtraction using color and depth data. *Multimedia Tools and Applications*, October 2018.
265. S. Maity, A. Chakrabarti, and D. Bhattacharjee. Block-based quantized histogram (bbqh) for efficient background modeling and foreground extraction in video. *International Conference on Data Management, Analytics and Innovation, ICDMAI 2017*, pages 224–229, 2017.
266. L. Marcomini and A. Cunha. A Comparison between Background Modelling Methods for Vehicle Segmentation in Highway Traffic Videos. *Preprint*, 2018.
267. C. Marghes and T. Bouwmans. Background modeling via incremental maximum margin criterion. *International Workshop on Subspace Methods, ACCV 2010 Workshop Subspace 2010*, November 2010.
268. C. Marghes, T. Bouwmans, and R. Vasiu. Background modeling and foreground detection via a reconstructive and discriminative subspace learning approach. *International Conference on Image Processing, Computer Vision, and Pattern Recognition, IPCV 2012*, July 2012.
269. I. Martins, P. Carvalho, L. Corte-Real, and J. Alba-Castro. BMOG: Boosted Gaussian Mixture Model with Controlled Complexity. *IbPRIA 2017*, 2017.

270. W. Matusik, C. Buehler, R. Raskar, L. McMillan, and S. Gortler. Image-based visual hulls. *ACM Transactions on Graphic, SIGGRAPH 2000*, 2000.
271. N. McFarlane and C. Schofield. Segmentation and tracking of piglets in images. *British Machine Vision and Applications, BMVA 1995*, pages 187–193, 1995.
272. G. Medioni. Adaptive color background modeling for real-time segmentation of video streams. *International Conference on Imaging Science, Systems, and Technology*, pages 227–232, June 1999.
273. L. Mei, J. Guo, P. Lu, Q. Liu, and F. Teng. Inland ship detection based on dynamic group sparsity. *International Conference on Advanced Computational Intelligence, ICACI 2017*, pages 1–6, 2017.
274. M. Mendonca and L. Oliveira. Context-supported road information for background modeling. *SIBGRAPI Conference on Graphics, Patterns and Images*, 2015.
275. S. Messelodi, C. Modena, N. Segata, and M. Zanin. A Kalman filter based background updating algorithm robust to sharp illumination changes. *International Conference on Image Analysis and Processing, ICIAP 2005*, 3617:163–170, September 2005.
276. I. Mikic, M. Trivedi, E. Hunter, and P. Cosman. Human body model acquisition and motion capture using voxel data. *AMDO 2002*, pages 104–118, 2002.
277. I. Mikic, M. Trivedi, E. Hunter, and P. Cosman. Human body model acquisition and tracking using voxel data. *International Journal of Computer Vision*, 3(53):199–223, 2003.
278. J. Milla, S. Toral, M. Vargas, and F. Barrero. Dual-rate background subtraction approach for estimating traffic queue parameters in urban scenes. *IET Intelligent Transport Systems*, 2013.
279. T. Minematsu, A. Shimada, and R. Taniguchi. Analytics of deep neural network in change detection. *IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2017*, September 2017.
280. T. Minematsu, H. Uchiyama, A. Shimada, and R. Taniguchi. Analytics of deep neural network-based background subtraction. *MDPI Journal of Imaging*, 2018.
281. G. Monteiro. Traffic video surveillance for automatic incident detection on highways. *Master Thesis*, 2009.
282. G. Monteiro, J. Marcos, and J. Batista. Stopped vehicle detection system for outdoor traffic surveillance. *RECPAD 2008*, 2008.
283. G. Monteiro, J. Marcos, M. Ribeiro, and J. Batista. Robust segmentation process to detect incidents on highways. *International Conference on Image Analysis and Recognition, ICIAR 2008*, 2008.
284. G. Monteiro, M. Ribeiro, J. Marcos, and J. Batista. Robust segmentation for outdoor traffic surveillance. *IEEE International Conference on Image Processing, ICIP 2008*, 2008.
285. S. Muniruzzaman, N. Haque, F. Rahman, M. Siam, R. Musabbir, M. Hadiuzzaman, and S. Hossain. Deterministic algorithm for traffic detection in free-flow and congestion using video sensor. *Journal of Built Environment, Technology and Engineering*, September 2016.
286. O. Munteanu, T. Bouwmans, E. Zahzah, and R. Vasiu. The detection of moving objects in video by background subtraction using Dempster-Shafer theory. *Scientific Bulletin of the University of Timisoara*, 2015.
287. Y. Nakashima, N. Babaguchi, and J. Fan. Automatic generation of privacy protected using background estimation. *IEEE International Conference on Multimedia and Expo, ICME 2011*, pages 1–6, 2011.
288. P. Narayanamurthy and N. Vaswani. A Fast and Memory-efficient Algorithm for Robust PCA (MEROP). *IEEE International Conference on Acoustics, Speech, and Signal, ICASSP 2018*, April 2018.
289. S. Nazir, S. Newey, R. Irvine, F. Verdicchio, P. Davidson, G. Fairhurst, and R. van der Wal. WiseEye: Next Generation Expandable and Programmable Camera Trap Platform for Wildlife Research. *PLOS*, 2017.
290. M. Neuhausen. Video-based parking space detection: Localisation of vehicles and development of an infrastructure for a routeing system. *Institut fr Neuroinformatik*, 2015.
291. Y. Nguwi, A. Kouzani, J. Kumar, and D. Driscoll. Automatic detection of lizards. *International Conference on Advanced Mechatronic Systems, ICAMechS 2006*, pages 300–305, 2016.
292. T. Nguyen, C. Pham, S. Ha, and J. Jeon. Change detection by training a triplet network for motion feature extraction. *IEEE Transactions on Circuits and Systems for Video Technology*, January 2018.
293. R. Nishide, Y. Hosomi, T. Ohkawa, K. Oyama, and C. Ohta. Detecting and Tracking Breeding Cows from Birds Eye Video of Pasture. *IIAE International Conference on Intelligent Systems and Image Processing*, pages 239–246, 2017.
294. V. Nogueira, H. Oliveira, J. Silva, T. Vieira, and K. Oliveira. RetailNet: A deep learning approach for people counting and hot spots detection in retail stores. *SIBGRAPI 2019*, 2019.

295. N. Oliver, B. Rosario, and A. Pentland. A Bayesian computer vision system for modeling human interactions. *International Conference on Vision Systems, ICVS 1999*, January 1999.
296. D. Panda and S. Meher. A new Wronskian change detection model based codebook background subtraction for visual surveillance applications. *Journal of Visual Communication and Image Representation*, 2018.
297. S. Pang, J. Zhao, B. Hartill, and A. Sarrafzadeh. Modelling land water composition scene for maritime traffic surveillance. *Inderscience*, 2016.
298. A. Park, K. Jung, and T. Kurata. Codebook-based background subtraction to generate photorealistic avatars in a walkthrough simulator. *International Symposium on Visual Computing, ISVC 2009*, December 2009.
299. H. Park and K. Hyun. Real-time hand gesture recognition for augmented screen using average background and CAMshift. *Korea-Japan Joint Workshop on Frontiers of Computer Vision, FCV 2013*, February 2013.
300. M. Paul, W. Lin, C. Lau, and B. Lee. Video coding using the most common frame in scene. *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2010*, 2010.
301. M. Paul, W. Lin, C. Lau, and B. Lee. Pattern-based video coding with dynamic background modeling. *EURASIP Journal on Advances in Signal Processing*, page 138, 2013.
302. M. Paul, W. Lin, C. Lau, and B. Lee. Video coding with dynamic backgrounds. *EURASIP Journal on Advances in Signal Processing*, 2013.
303. N. Peixoto, N. Cardoso, P. Goncalves, P. Cardoso, J. Cabral, A. Tavares, and J. Mendes. Motion segmentation object detection in complex aquatic scenes and its surroundings. *International Conference on Industrial Informatics, INDIN 2012*, pages 162–166, 2012.
304. D. Penciuc, F. El Baf, and T. Bouwmans. Comparison of background subtraction methods for an interactive learning space. *NETTIES 2006*, September 2006.
305. T. Perrett, M. Mirmehdi, and E. Dias. Visual monitoring of driver and passenger control panel interactions. *IEEE Transactions on Intelligent Transportation Systems*, 2016.
306. V. Pham, P. Vo, H. Vu Thanh, and B. Le Hoai. GPU implementation of extended gaussian mixture model for background subtraction. *IEEE International Conference on Computing and Telecommunication Technologies, RIVF 2010*, November 2010.
307. R. Pilipovic, V. Risojevic, Z. Babic, and G. Mirjanic. Background subtraction for honey bee detection in hive entrance video. *INFOTEH-JAHORINA*, 15, March 2016.
308. T. Pinkiewicz. Computational Techniques for Automated Tracking and Analysis of Fish Movement in Controlled Aquatic Environments. *PhD Thesis, University of Tasmania, Australia*, 2012.
309. F. Porikli. Multiplicative background-foreground estimation under uncontrolled illumination using intrinsic images. *IEEE Motion Multi-Workshop*, 2005.
310. F. Porikli, Y. Ivanov, and T. Haga. Robust abandoned object detection using dual foregrounds. *EURASIP Journal on Advances in Signal Processing*, page 30, 2008.
311. F. Porikli and C. Wren. Change detection by frequency decomposition: Wave-Back. *International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2005*, April 2005.
312. C. Postigo, J. Torres, and J. Menendez. Vacant parking area estimation through background subtraction and transience map analysis. *IET Intelligent Transport Systems*, 9(9):835–841, 2015.
313. P. Power and J. Schoonees. Understanding background mixture models for foreground segmentation. *Imaging and Vision Computing New Zealand*, November 2002.
314. D. Prasad, C. Prasath, D. Rajan, L. Rachmawati, E. Rajabally, and C. Quek. Challenges in video based object detection in maritime scenario using computer vision. *WASET International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 11(1), January 2017.
315. D. Prasad, D. Rajan, L. Rachmawati, E. Rajabally, and C. Quek. Video processing from electro-optical sensors for object detection and tracking in maritime environment: A survey. *Preprint*, November 2016.
316. A. Prati, I. Mikic, M. Trivedi, and R. Cucchiara. Detecting moving shadows: Algorithms and evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(4):918–923, July 2003.
317. J. Pulgarin-Giraldo, A. Alvarez-Meza, D. Insuasti-Ceballos, T. Bouwmans, and G. Castellanos-Dominguez. GMM background modeling using divergence-based weight updating. *Conference Ibero American Congress on Pattern Recognition, CIARP 2016*, 2016.
318. C. Qiu and N. Vaswani. Real-time robust principal components pursuit. *International Conference on Communication Control and Computing*, 2010.
319. C. Qiu and N. Vaswani. Support predicted modified-CS for recursive robust principal components' pursuit. *IEEE International Symposium on Information Theory, ISIT 2011*, 2011.

320. Z. Qu, M. Yu, and J. Liu. Real-time traffic vehicle tracking based on improved MoG background extraction and motion segmentation. *International Symposium on Systems and Control in Aeronautics and Astronautics, ISSCAA 2010*, pages 676–680, June 2010.
321. Z. Qu, S. Yu, and M. Fu. Motion background modeling based on context-encoder. *IEEE International Conference on Artificial Intelligence and Pattern Recognition, ICAIPR 2016*, September 2016.
322. J. Quesada and P. Rodriguez. Automatic vehicle counting method based on principal component pursuit background modeling. *IEEE International Conference on Image Processing, ICIP 2016*, 2016.
323. M. Radolko, F. Farhadifard, and U. von Lukas. Dataset on Underwater Change Detection. *IEEE Monterey OCEANS 2016*, 2016.
324. M. Radolko and E. Gutzeit. Video Segmentation via a Gaussian Switch Background Model and Higher Order Markov Random Fields. *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISAPP 2015*, 2015.
325. M. Razif, M. Mokji, and M. Zabidi. Low complexity maritime surveillance video using background subtraction on H.26. *International Symposium on Technology Management and Emerging Technologies, ISTMET 2015*, pages 364–368, 2015.
326. V. Reilly. Detecting, tracking, and recognizing activities in aerial video. *PhD thesis, University of Central Florida, USA*, 2012.
327. B. Rezaei and S. Ostadabbas. Moving Object Detection through Robust Matrix Completion Augmented with Objectness . *IEEE Journal of Selected Topics in Signal Processing*, December 2018.
328. C. Ridder, O. Munkelt, and H. Kirchner. Adaptive background estimation and foreground detection using kalman-filtering. *International Conference on recent Advances in Mechatronics, ICRAM 1995*, pages 193–195, 1995.
329. S. Rinsurongkawong, M. Ekpanyapong, and M. Dailey. Fire detection for early fire alarm based on optical flow video processing. *IEEE International Conference on Electrical Engineering/Electronic, ECTICon 2012*, pages 1–4, 2012.
330. P. Rodriguez. Real-time incremental principal component pursuit for video background modeling on the TK1. *GPU Technical Conference, GTC 2015*, March 2015.
331. P. Rodriguez and B. Wohlberg. Fast principal component pursuit via alternating minimization. *IEEE International Conference on Image Processing, ICIP 2013*, September 2013.
332. P. Rodriguez and B. Wohlberg. A Matlab implementation of a fast incremental principal component pursuit algorithm for video background modeling. *IEEE International Conference on Image Processing, ICIP 2014*, October 2014.
333. P. Rodriguez and B. Wohlberg. Translational and rotational jitter invariant incremental principal-component pursuit for video background modeling. *IEEE International Conference on Image Processing, ICIP 2015*, 2015.
334. P. Rodriguez and B. Wohlberg. Ghosting suppression for incremental principal component pursuit algorithms. *IEEE Global Conference on Signal and Information Processing, GlobalSIP 2016*, December 2016.
335. P. Rodriguez and B. Wohlberg. Incremental principal component pursuit for video background modeling. *Journal of Mathematical Imaging and Vision*, 55(1):1–18, 2016.
336. P. Rodriguez and B. Wohlberg. An incremental principal component pursuit algorithm via projections onto the l_1 ball. *IEEE International Conference on Electronics, Electrical Engineering and Computing, INTERCON 2017*, pages 1–4, 2017.
337. D. Rout, B. Subudhi, T. Veerakumar, and S. Chaudhury. Spatio-contextual Gaussian Mixture Model for Local Change Detection in Underwater Video. *Expert Systems With Applications*, 2017.
338. M. Saghafi, S. Javadein, S. Noorhossein, and H. Khalili. Robust ship detection and tracking using modified ViBe and backwash cancellation algorithm. *International Conference on Computational Intelligence and Information Technology, CIIT 2012*, 2012.
339. M. Saker, C. Weihua, and M. Song. Detection and recognition of illegally parked vehicle based on adaptive Gaussian mixture model and seed fill algorithm. *Journal of Information and Communication Convergence Engineering*, pages 197–204, 2015.
340. D. Sakkos, H. Liu, J. Han, and L. Shao. End-to-end video background subtraction with 3D convolutional neural networks. *Multimedia Tools and Applications*, pages 1–19, December 2017.
341. A. Salman, S. Maqbool, A. Khan, A. Jalal, and F. Shafait. Real-time fish detection in complex backgrounds using probabilistic background modelling. *Ecological Informatics*, 51(44-51), May 2019.

342. A. Salman, S. Siddiqui, F. Shafait, A. Mian, M. Shortis, K. Khurshid, A. Ulges, and U. Schwanecke. Automatic fish detection in underwater videos by a deep neural network-based hybrid motion learning system. *ICES Journal of Marine Science*, 2019.
343. C. Salvadori, D. Makris, M. Petracca, J. Rincon, and S. Velastin. Gaussian mixture background modelling optimisation for micro-controllers. *International Symposium on Visual Computing, ISVC 2012*, pages 241–251, 2012.
344. C. Salvadori, M. Petracca, J. Rincon, S. Velastin, and D. Makris. An optimisation of gaussian mixture models for integer processing units. *Real-Time Image Processing*, February 2014.
345. A. Savakis and A. Shringarpure. Semantic background estimation in video sequences. *IEEE International Conference on Signal Processing and Integrated Networks, SPIN 2018*, pages 597–601, 2018.
346. S. Sawalakhe and S. Metkar. Foreground background traffic scene modeling for object motion detection. *IEEE India Conference, INDICON 2014*, pages 1–6, 2014.
347. N. Seese, A. Myers, K. Smith, and A. Smith. Adaptive foreground extraction for deep fish classification. *Workshop on Computer Vision for Analysis of Underwater Imagery in conjuction with ICPR 2016, CVAUI 2016*, pages 19–24, 2016.
348. L. Sha, P. Lucey, S. Sridharan, S. Morgan, and D. Pease. Understanding and analyzing a large collection of archived swimming videos. *IEEE Winter Conference on Applications of Computer Vision, WACV 2014*, pages 674–681, March 2014.
349. M. Shafiee, P. Siva, P. Fieguth, and A. Wong. Embedded motion detection via neural response mixture background modeling. *International Conference on Computer Vision and Pattern Recognition, CVPR 2016*, June 2016.
350. M. Shafiee, P. Siva, P. Fieguth, and A. Wong. Real-time embedded motion detection via neural response mixture modeling. *Journal of Signal Processing Systems*, June 2017.
351. M. Shah, J. Deng, and B. Woodford. Video background modeling: recent approaches, issues and our proposed techniques. *Machine Vision and Applications*, 5(25):1105–1119, 2014.
352. M. Shakeri and H. Deldari. Fuzzy-cellular background subtraction technique for urban traffic applications. *World Applied Sciences Journal*, 5(1), 2008.
353. M. Shakeri, H. Deldari, H. Foroughi, A. Saberi, and A. Naseri. A novel fuzzy background subtraction method based on cellular automata for urban traffic applications. *International Conference on Signal Processing, ICSP 2008*, pages 899–902, October 2008.
354. M. Shakeri and H. Zhang. Real-time bird detection based on background subtraction. *World Congress on Intelligent Control and Automation, WCICA 2012*, pages 4507–4510, July 2012.
355. M. Shakeri and H. Zhang. Moving object detection in time-lapse or motion trigger image sequences using low-rank and invariant sparse decomposition. *IEEE ICCV 2017*, October 2017.
356. V. Sharma, N. Nain, and T. Badal. A redefined codebook model for dynamic backgrounds. *International Conference on Computer Vision and Image Processing, CVIP 2016*, 2016.
357. Y. Sheikh, O. Javed, and T. Kanade. Background subtraction for freely moving cameras. *IEEE International Conference on Computer Vision, ICCV 2009*, pages 1219–1225, October 2009.
358. Y. Sheikh and M. Shah. Bayesian modeling of dynamic scenes for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(11):1778–1792, November 2005.
359. Y. Sheikh and M. Shah. Bayesian object detection in dynamic scenes. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2005*, June 2005.
360. W. Shen, Y. Lin, L. Yu, F. Xue, and W. Hong. Single Channel Circular SAR Moving Target Detection Based on Logarithm Background Subtraction Algorithm. *MPDI Remote Sensing*, 2018.
361. P. Shi, E. Jones, and Q. Zhu. Median model for background subtraction in intelligent transportation systems. *SPIE Symposium on Electronic Imaging, Image Processing: Algorithms and Systems III*, 1, January 2004.
362. M. Sigari. Fuzzy background modeling/subtraction and its application in vehicle detection. *World Congress on Engineering and Computer Science, WCECS 2008*, October 2008.
363. M. Sigari, N. Mozayani, and H. Pourreza. Fuzzy running average and fuzzy background subtraction: Concepts and application. *International Journal of Computer Science and Network Security*, 8(2):138–143, 2008.
364. G. Silva and P. Rodriguez. Jitter invariant incremental principal component pursuit for video background modeling on the TK1. *Asilomar Conference on Signals, Systems, and Computers, ACSSC 2015*, November 2015.
365. R. Silva, K. Aires, T. Santos, K. Abdala, R. Veras, and A. Soares. Automatic detection of motorcyclists without helmet. *Latin American Computing Conference, CLEI 2013*, pages 1–7, 2013.

366. A. Sobral. BGSLibrary: An opencv c++ background subtraction library. *Workshop de Viso Computacional, WVC 2013*, June 2013.
367. A. Sobral, C. Baker, T. Bouwmans, and E. Zahzah. Incremental and multi-feature tensor subspace learning applied for background modeling and subtraction. *International Conference on Image Analysis and Recognition, ICIAR 2014*, October 2014.
368. A. Sobral and T. Bouwmans. Bgs library: A library framework for algorithms evaluation in foreground/background segmentation. *Background Modeling and Foreground Detection for Video Surveillance*, 2014.
369. A. Sobral, T. Bouwmans, and E. Zahzah. Double-constrained RPCA based on saliency maps for foreground detection in automated maritime surveillance. *ISBC 2015 Workshop conjunction with AVSS 2015*, 2015.
370. A. Sobral, S. Javed, S. Jung, T. Bouwmans, and E. Zahzah. Online stochastic tensor decomposition for background subtraction in multispectral video sequences. *Workshop on Robust Subspace Learning and Computer Vision, ICCV 2015*, 2015.
371. A. Sobral, L. Oliveira, L. Schnitman, and F. Souza. Highway traffic congestion classification using holistic properties. *IASTED International Conference on Signal Processing, Pattern Recognition and Applications, SPPRA 2013*, February 2013.
372. D. Socek, D. Culibrk, O. Marques, H. Kalva, and B. Furht. A hybrid color-based foreground object detection method for automated marine surveillance. *Advanced Concepts for Intelligent Vision Systems Conference, ACVIS 2005*, 2005.
373. K. Song and J. Tai. Real-time background estimation of traffic imagery using group-based histogram. *Journal of Information Science and Engineering*, 24:411–423, 2008.
374. C. Spampinato, Y. Burger, G. Nadarajan, and R. Fisher. Detecting, tracking and counting fish in low quality unconstrained underwater videos. *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISAPP 2008*, pages 514–519, 2008.
375. C. Spampinato, D. Giordano, R. Di Salvo, Y. Chen-Burger, R. Fisher, and G. Nadarajan. Automatic fish classification for underwater species behavior understanding. *ACM international workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams, ARTEMIS 2010*, pages 45–50, 2010.
376. C. Spampinato, S. Palazzo, B. Boom, J. van Ossenbruggen, I. Kavasidis, R. Di Salvo, F. Lin, D. Giordano, L. Hardman, and R. Fisher. Understanding fish behavior during typhoon events in real-life underwater environments. *Multimedia Tools and Applications*, pages 1–38, 2014.
377. C. Spampinato, S. Palazzo, and I. Kavasidis. A texton-based kernel density estimation approach for background modeling under extreme conditions. *Computer Vision and Image Understanding, CVIU 2014*, 122:74–83, 2014.
378. P. St-Charles, G. Bilodeau, and R. Bergevin. Flexible background subtraction with self-balanced local sensitivity. *IEEE Change Detection Workshop, CDW 2014*, June 2014.
379. C. Stauffer and E. Grimson. Adaptive background mixture models for real-time tracking. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 1999*, pages 246–252, 1999.
380. E. Stergiopoulou, K. Sgouropoulos, N. Nikolaou, N. Papamarkos, and N. Mitianoudis. Real-time hand detection in a complex background. *Engineering Applications of Artificial Intelligence*, 35:54–70, 2014.
381. Y. Sugaya and K. Kanatani. Extracting moving objects from a moving camera video sequence. *Symposium on Sensing via Image Information, SSII2004*, pages 279–284, 2004.
382. Z. Szpak and J. Tapamo. Maritime surveillance: Tracking ships inside a dynamic background using a fast level-set. *Expert Systems with Applications*, 38(6):6669–6680, 2011.
383. G. Szwoch. Performance evaluation of the parallel codebook algorithm for background subtraction in video stream. *International Conference on Multimedia Communications, Services and Security, MCSS 2011*, June 2011.
384. H. Tabkhi, R. Bushey, and G. Schirner. Algorithm and architecture co-design of mixture of Gaussian (MoG) background subtraction for embedded vision processor. *Asilomar Conference on Signals, Systems, and Computers, Asilomar 2013*, November 2013.
385. B. Tamas. Detecting and analyzing rowing motion in videos. *BME Scientific Student Conference*, 2016.
386. A. Taneja, L. Ballan, and M. Pollefeys. Modeling dynamic scenes recorded with freely. *Asian Conference on Computer Vision, ACCV 2010*, 2010.
387. Y. Tao, P. Palasek, Z. Ling, and I. Patras. Background modelling based on generative Unet. *IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2017*, September 2017.

388. A. Tavakkoli, A. Ambardekar, M. Nicolescu, and S. Louis. A genetic approach to training support vector data descriptors for background modeling in video data. *International Symposium on Visual Computing, ISVC 2007*, November 2007.
389. A. Tavakkoli, M. Nicolescu, and G. Bebis. Novelty detection approach for foreground region detection in videos with quasi-stationary backgrounds. *International Symposium on Visual Computing, ISVC 2006*, pages 40–49, November 2006.
390. E. Tejada and P. Rodriguez. Moving object detection in videos using principal component pursuit and convolutional neural networks. *GlobalSIP 2017*, pages 793–797, 2017.
391. M. Teutsch, W. Krueger, and J. Beyerer. Evaluation of object segmentation to improve moving vehicle detection in aerial videos. *IEEE Advanced Video and Signal-based Surveillance, AVSS 2014*, August 2014.
392. D. Thirde, M. Borg, and J. Ferryman. A real-time scene understanding system for airport apron monitoring. *International Conference on Computer Vision Systems, ICVS 2006*, 2006.
393. Q. Tian, L. Zhang, Y. Wei, W. Zhao, and W. Fei. Vehicle detection and tracking at night in video surveillance. *International Journal on Online Engineering*, 9:60–64, 2013.
394. P. Tiefenbacher, M. Hofmann, D. Merget, and G. Rigoll. Pid-based regulation of background dynamics for foreground segmentation. *IEEE International Conference on Image Processing, ICIP 2014*, pages 3282–3286, 2014.
395. B. Toreyin, Y. Dedeoglu, U. Gdkbay, and A. Cetin. Computer vision based method for real-time fire and flame detection. *Pattern Recognition Letters*, 2007.
396. E. Toropov, L. Gui, S. Zhang, S. Kottur, and J. Moura. Traffic Flow from a Lowframe Rate City Camera. *International Conference on Image Processing, ICIP 2015*, 2015.
397. K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. *International Conference on Computer Vision*, pages 255–261, September 1999.
398. T. Tran and T. Le. Vision based boat detection for maritime surveillance. *International Conference on Electronics, Information, and Communications, ICEIC 2016*, pages 1–4, 2016.
399. G. Tu, H. Karstoft, L. Pedersen, and E. Jorg. Segmentation of sows in farrowing pens. *IET Image Processing*, January 2014.
400. G. Tu, H. Karstoft, L. Pedersen, and E. Jorg. Illumination and reflectance estimation with its application in foreground detection. *MDPI Sensors 2015*, 15:21407–21426, 2015.
401. G. Tzanidou. Carried baggage detection and recognition in video surveillance with foreground segmentation. *Loughborough University, UK*, 2014.
402. L. Unzueta, M. Nieto, A. Cortes, J. Barandiaran, O. Otaegui, and P. Sanchez. Adaptive multicue background subtraction for robust vehicle counting and classification. *IEEE Transaction on Intelligent Transportation Systems*, 13(2):527–540, June 2012.
403. J. Varona, J. Gonzalez, I. Rius, and J. Villanueva. Importance of detection for video surveillance applications. *Optical Engineering*, pages 1–9, 2008.
404. N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurth. Robust PCA and Robust Subspace Tracking: A Comparative Evaluation. *Statistical Signal Processing Workshop,SSP 2018*, June 2018.
405. N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurth. Robust Subspace Learning: Robust PCA, Robust Subspace Tracking and Robust Subspace Recovery. *IEEE Signal Processing Magazine*, 35(4):32–55, July 2018.
406. A. Virginas-Tar, M. Baba, V. Gui, D. Pescaru, and I. Jian. Vehicle counting and classification for traffic surveillance using wireless video sensor networks. *Telecommunications Forum Telfor, TELFOR 2014*, pages 1019–1022, 2014.
407. I. Vujovic, M. Jurcevic, and I. Kuzmanic. Traffic video surveillance in different weather conditions. *Transactions on Maritime Science*, 3(1), 2014.
408. Wahyono, A. Filonenko, and K. Jo. Illegally parked vehicle detection using adaptive dual background model. *Annual Conference of the IEEE Industrial Electronics Society, IECON 2015*, pages 2225–2228, 2015.
409. C. Wang and Z. Song. Vehicle detection based on spatial-temporal connection background subtraction. *IEEE International Conference on Information and Automation*, pages 320–323, 2011.
410. G. Wang, J. Hwang, K. Williams, F. Wallace, and C. Rose. Shrinking encoding with two-level codebook learning for fine-grained fish recognition. *Workshop on Computer Vision for Analysis of Underwater Imagery in conjuction with ICPR 2016, CVAUI 2016*, 2016.
411. J. Wang, G. Bebis, M. Nicolescu, M. Nicolescu, and R. Miller. Improving target detection by coupling it with tracking. *Machine Vision and Application*, pages 1–19, 2008.
412. K. Wang, Y. Liu, C. Gou, and F. Wang. A multi-view learning approach to foreground detection for traffic surveillance applications. *IEEE Transactions on Vehicular Technology*, 2015.

413. K. Wang, Y. Liu, C. Gou, and F. Wang. A multi-view learning approach to foreground detection for traffic surveillance applications. *IEEE Transactions on Vehicular Technology*, 65(6):4144–4158, June 2016.
414. X. Wang, D. Zhao, G. Sun, X. Liu, and Y. Wu. Target detection algorithm based on improved Gaussian mixture model. *International Conference on Electrical, Computer Engineering and Electronics, ICECEE 2015*, 2015.
415. Y. Wang, P. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar. CDnet 2014: an expanded change detection benchmark dataset. *IEEE Workshop on Change Detection, CDW 2014 in conjunction with CVPR 2014*, June 2014.
416. Y. Wang, Z. Luo, and P. Jodoin. Interactive deep learning method for segmenting moving objects. *Pattern Recognition Letters*, 2016.
417. Z. Wang, J. Gao, X. Chen, and W. Zhao. Organic light emitting diode pixel defect detection through patterned background removal. *Sensor Letters*, 11(2):356–361, February 2013.
418. R. Waranusast, N. Bundon, and P. Pattanathaburt. Machine vision techniques for motorcycle safety helmet detection. *International Conference of Image and Vision Computing New Zealand, IVCNZ 2013*, pages 35–40, 2013.
419. J. Warren. Unencumbered full body interaction in video game. *Master Thesis, MFA Design and Technology Parsons School of Design, New York*, April 2003.
420. J. Wei, J. Zhao, Y. Zhao, and Z. Zhao. Unsupervised Anomaly Detection for Traffic Surveillance based on Background Modeling. *IEEE CVPR 2018*, 2018.
421. B. Weinstein. Motionmeerkat: integrating motion video detection and ecological monitoring. *Methods in Ecology and Evolution*, 2014.
422. B. Weinstein. A computer vision for animal ecology. *Journal of Animal Ecology*, October 2017.
423. B. Weinstein. Scene-specific convolutional neural networks for video-based biodiversity detection. *Methods in Ecology and Evolution*, 2018.
424. M. Wilber, W. Scheirer, P. Leitner, B. Heflin, J. Zott, D. Delaney D. Reinke, and T. Boult. Animal recognition in the mojave desert: Vision tools for field biologists. *IEEE WACV 2013*, pages 206–213, 2013.
425. K. Williams, R. Towler, and C. Wilson. Cam-trawl: a combination trawl and stereo-camera system. *Sea Technology*, 51(12):45–50, 2010.
426. C. Wren and A. Azarbayejani. Pfinder : Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780 –785, July 1997.
427. C. Wren and A. Azarbayejani. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780 –785, July 1997.
428. M. Xiao, C. Han, and X. Kang. A background reconstruction for dynamic scenes. *International Conference on Information Fusion, ICIF 2006*, pages 1–6, July 2006.
429. L. Xie, B. Li, Z. Wang, and Z. Yan. Overloaded ship identification based on image processing and kalman filtering. *Journal of Convergence Information Technology*, 7(22):425–433, December 2012.
430. L. Xu, Y. Li, Y. Wang, and E. Chen. Temporally adaptive restricted boltzmann machine for background modeling. *American Association for Artificial Intelligence, AAAI 2015*, January 2015.
431. P. Xu, M. Ye, X. Li, Q. Liu, Y. Yang, and J. Ding. Dynamic background learning through deep auto-encoder networks. *ACM International Conference on Multimedia*, November 2014.
432. P. Xu, M. Ye, Q. Liu, X. Li, L. Pei, and J. Ding. Motion detection via a couple of auto-encoder networks. *International Conference on Multimedia and Expo, ICME 2014*, 2014.
433. X. Xu. Online robust principal component analysis for background subtraction: A system evaluation on Toyota car data. *Master thesis, University of Illinois, Urbana-Champaign, USA*, 2014.
434. M. Yamazaki, G. Xu, and Y. Chen. Detection of moving objects by independent component analysis. *Asian Conference on Computer Vision, ACCV 2006*, pages 467–478, 2006.
435. K. Yamazawa and N. Yokoya. Detecting moving objects from omnidirectional dynamic images based on adaptive background subtraction. *IEEE International Conference on Image Processing, ICIP 2003*, 2003.
436. Z. Yan. A new background subtraction method for on road traffic. *Journal of Image and Graphics*, 3:593–599, December 2008.
437. C. Yang. The use of video to detect and measure pollen on bees entering a hive. *PhD Thesis, Auckland University of Technology*, 2017.
438. T. Yang, X. Wang, B. Yao, J. Li, Y. Zhang, Z. He, and W. Duan. Small moving vehicle detection in a satellite video of an urban area. *MDPI Sensors*, 2016.
439. M. Yazdi and T. Bouwmans. New Trends on Moving Object Detection in Video Images Captured by a Moving Camera: A Survey. *Computer Science Review*, 28:157–177, 2018.

440. B. Yeo, W. Lim, and H. Lim. Scalable-width temporal edge detection for recursive background recovery in adaptive background modeling. *Applied Soft Computing*, January 2013.
441. B. Yeo, W. Lim, H. Lim, and W. Won. Extended fuzzy background modeling for moving vehicle detection using infrared vision. *IEICE Electronics Express*, pages 340–345, 2011.
442. H. Yousif, J. Yuan, R. Kays, and Z. He. Fast human-animal detection from highly cluttered camera-trap images using joint background modeling and deep learning classification. *IEEE ISCAS 2017*, pages 1–4, September 2017.
443. H. Yousif, J. Yuan, R. Kays, and Z. He. Animal Scanner: Software for classifying humans, animals, and empty frames in camera trap images. *Wiley Ecology and Evolution*, October 2018.
444. X. Yu, J. Wang, R. Kays, P. Jansen, T. Wang, and T. Huang. Automated identification of animal species in camera trap images. *EURASIP Journal on Image and Video Processing*, 2013.
445. G. Yuan, Y. Gao, D. Xu, and M. Jiang. A new background subtraction method using texture and color information. *ICIC 2011*, pages 541–558, 2011.
446. A. Zaharescu and M. Jamieson. Multi-scale multi-feature codebook-based background subtraction. *IEEE International Conference on Computer Vision Workshops, ICCV 2011*, pages 1753–1760, November 2011.
447. D. Zeng, X. Chen, M. Zhu, M. Goesele, and A. Kuijper. Background subtraction with real-time semantic segmentation. *IEEE Access*, 2019.
448. D. Zeng and M. Zhu. Combining background subtraction algorithms with convolutional neural network. *Preprint*, 2018.
449. C. Zhang, X. Li, and F. Lei. A novel camera-based drowning detection algorithm. *Chinese Conference on Image and Graphics Technologies*, pages 224–233, June 2015.
450. D. Zhang, E. OConnor, K. McGuinness, N. OConnor, F. Regan, and A. Smeaton. A visual sensing platform for creating a smarter multi-modal marine monitoring network. *MAED 2012*, pages 53–56, 2012.
451. R. Zhang, X. Liao, and J. Xu. A Background Subtraction Algorithm Robust to Intensity Flicker Based on IP Camera. *Journal of multimedia*, 9(10), October 2014.
452. R. Zhang, X. Liu, J. Hu, K. Chang, and K. Liu. A fast method for moving object detection in video surveillance image. *Signal, Image and Video Processing*, 2017.
453. S. Zhang, Z. Qi, and D. Zhang. Ship tracking using background subtraction and inter-frame correlation. *International Congress on Image and Signal Processing, CISP 2009*, pages 1–4, 2009.
454. S. Zhang, H. Yao, and S. Liu. Dynamic background subtraction based on local dependency histogram. *IEEE International Workshop on Visual Surveillance, VS 2008*, pages 1–8, October 2008.
455. X. Zhang, T. Huang, Y. Tian, and W. Gao. Background-modeling-based adaptive prediction for surveillance video coding. *IEEE Transactions on Image Processing*, 23(2):769–784, 2014.
456. X. Zhang, L. Liang, and Q. Huang. An efficient coding scheme for surveillance videos captured by stationary cameras. *Journal of Visual Communication and Image Representation*, 2010.
457. X. Zhang, Y. Tian, T. Huang, S. Dong, and W. Gao. Optimizing the hierarchical prediction and coding in HEVC for surveillance and conference videos with background modeling. *IEEE Transactions on Image Processing*, October 2014.
458. X. Zhang, Y. Tian, T. Huang, and W. Gao. Low-complexity and high-efficiency background modeling for surveillance video coding. *IEEE International Conference on Visual Communication and Image Processing, VCIP 2012*, November 2012.
459. X. Zhang, Y. Tian, L. Liang, T. Huang, and W. Gao. Macro-block-level selective background difference coding for surveillance video. *IEEE International Conference on Multimedia and Expo, ICME 2012*, pages 1067–1072, July 2012.
460. X. Zhang, X. Zhou, S. Yuan, and C. Wang. Moving target detection based on improved codebook algorithm. *Ordnance Industry Automation*, 2018.
461. Y. Zhang, X. Li, Z. Zhang, F. Wu, and L. Zhao. Deep learning driven blockwise moving object detection with binary scene modeling. *Neurocomputing*, June 2015.
462. Y. Zhang, C. Zhao, J. He, and A. Chen. Vehicles detection in complex urban traffic scenes using Gaussian mixture model with confidence measurement. *IET Intelligent Transport Systems*, 2016.
463. Y. Zhang, C. Zhao, J. He, and A. Chen. Vehicles detection in complex urban traffic scenes using Gaussian mixture model with confidence measurement. *IET Intelligent Transport Systems*, 2016.
464. Y. Zhang, X. Zhao, F. Li, J. Sun, S. Jiang, and C. Chen. Robust object tracking based on simplified codebook masked camshift algorithm. *Mathematical Problems in Engineering*, 2015.
465. J. Zhao, S. Pang, B. Hartill, and A. Sarrafzadeh. Adaptive Background Modeling for Land and Water Composition Scenes. *International Conference on Image Analysis and Processing, ICIAP 2015*, 2015.

466. K. Zhao and D. He. Target detection method for moving cows based on background subtraction. *International Journal of Agricultural and Biological Engineering*, 2015.
467. L. Zhao, Y. Tian, and T. Huang. Background-foreground division based search for motion estimation in surveillance video coding. *IEEE International Conference on Multimedia and Expo, ICME 2014*, 2014.
468. L. Zhao, X. Zhang, Y. Tian, R. Wang, and T. Huang. A background proportion adaptive Lagrange multiplier selection method for surveillance video on high HEVC. *International Conference on Multimedia and Expo, ICME 2013*, July 2013.
469. X. Zhao, Y. Chen, M. Tang, and J. Wang. Joint background reconstruction and foreground segmentation via a two-stage convolutional neural network. *Preprint*, 2017.
470. X. Zhao, X. Cheng, and X. Li. Illegal vehicle parking detection based on online learning. *World Congress on Multimedia and Computer Science*, 2013.
471. Y. Zhao and W. Liu. Fast robust foreground-background segmentation based on variable rate codebook method in Bayesian framework for detecting objects of interest. *International Congress on Image and Signal Processing*, 2014.
472. J. Zheng, Y. Wang, N. Nihan, and E. Hallenbeck. Extracting roadway background image: A mode based approach. *Journal of Transportation Research Report*, 1944:82–88, 2006.
473. J. Zheng, Y. Wang, N. Nihan, and E. Hallenbeck. Extracting roadway background image: A mode based approach. *Journal of Transportation Research Report,*, pages 82–88, 2006.
474. Y. Zheng, W. Wu, H. Xu, L. Gan, and J. Yu. A novel ship-bridge collision avoidance system based on monocular computer vision. *Research Journal of Applied Sciences, Engineering and Technology*, 4(6):647–653, 2013.
475. J. Zhong and S. Sclaroff. Segmenting foreground objects from a dynamic textured background via a robust Kalman filter. *International Conference on Computer Vision, ICCV 2003*, pages 44–50, 2003.
476. A. Zhou, W. Xie, and J. Pei. Background modeling in the fourier domain for maritime infrared target detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.
477. C. Zhou, B. Zhang, K. Lin, D. Xu, C. Chen, X. Yang, and C. Sun. Near-infrared imaging to quantify the feeding behavior of fish in aquaculture. *Computers and Electronics in Agriculture*, 135:233–241, 2017.
478. X. Zhou, X. Liu, A. Jiang, B. Yan, and C. Yang. Improving Video Segmentation by Fusing Depth Cues and the ViBe Algorithm. *MDPI Sensors*, May 2017.
479. Z. Zhou, L. Shangguan, X. Zheng, L. Yang, and Y. Liu. Design and Implementation of an RFID-Based Customer Shopping Behavior Mining System. *IEEE/ACM Transactions on Networking*, 2017.
480. Z. Zivkovic. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognition Letters*, 27(7):773–780, January 2006.
481. Z. Zivkovic and F. Heijden. Recursive unsupervised learning of finite mixture models. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 5(26):651–656, 2004.