



**HAL**  
open science

## **Moving objects detection with a moving camera: A comprehensive review**

Marie-Neige Chapel, Thierry Bouwmans

### ► **To cite this version:**

Marie-Neige Chapel, Thierry Bouwmans. Moving objects detection with a moving camera: A comprehensive review. *Computer Science Review*, 2020, 38, pp.100310. <10.1016/j.cosrev.2020.100310>. <hal-02962156>

**HAL Id: hal-02962156**

**<https://hal.science/hal-02962156v1>**

Submitted on 17 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

# Moving Objects Detection with a Moving Camera: A Comprehensive Review

Marie-Neige Chapel<sup>a</sup>, Thierry Bouwmans<sup>b</sup>

<sup>a</sup>Lab. L3I, LRU<sup>univ.</sup>, Avenue Albert Einstein, 17000 La Rochelle, France

<sup>b</sup>Lab. MIA, LRU<sup>univ.</sup>, Avenue Albert Einstein, 17000 La Rochelle, France

---

## Abstract

During about 30 years, a lot of research teams have worked on the big challenge of detection of moving objects in various challenging environments. First applications concern static cameras but with the rise of the mobile sensors studies on moving cameras have emerged over time. In this survey, we propose to identify and categorize the different existing methods found in the literature. For this purpose, we propose to classify these methods according to the choice of the scene representation: one plane or several parts. Inside these two categories, the methods are grouped according to eight different approaches: panoramic background subtraction, dual cameras, motion compensation, subspace segmentation, motion segmentation, plane+parallax, multi planes and split image in blocks. A reminder of methods for static cameras is provided as well as the challenges with both static and moving cameras. Publicly available datasets and evaluation metrics are also surveyed in this paper.

*Keywords:* Moving object detection, Moving camera, Background subtraction, Motion analysis

*2010 MSC:* 00-01, 99-00

---

## 1. Introduction

Cameras are more and more present in our daily lives whether it is in the streets, in our homes and even in our pockets with smart-phones. Many real applications [1] are based on videos taken either by static or moving cameras such

5 as in video surveillance of human activities [2], visual observation of animals  
[3, 4, 5], home care [2], optical motion capture [6] and multimedia applications  
[7]. In their process, these applications often require a moving objects detec-  
tion step followed by tracking and recognition steps. Since 30 years, moving  
objects detection is thus surely among the most investigated field in computer  
10 vision providing a big amount of publications. First, methods were developed  
for static cameras but, in the last two decades with the expansion of sensors,  
approaches with moving cameras have been of many interests giving more chal-  
lenging situations to handle. However, many challenges have been identified in  
the literature and are related either to the cameras, to the background or to the  
15 moving objects of the filmed scenes.

A lot of surveys in the literature are about moving objects detection in the  
case of static cameras. In 2000, Mc Ivor [8] surveyed nine algorithms allowing  
a first comparison of the models. However, this survey is mainly limited on a  
description of the algorithms. In 2004, Piccardi [9] provided a review on seven  
20 methods and an original categorization based on speed, memory requirements  
and accuracy. This review allows the readers to compare the complexity of the  
different methods and effectively helps them to select the most adapted method  
for their specific application. In 2005, Cheung and Kamath [10] classified sev-  
eral methods into non-recursive and recursive techniques. Following this clas-  
sification, Elhabian et al.[11] provided a large survey in background modeling.  
25 However, this classification in terms of non-recursive and recursive techniques is  
more suitable for the background maintenance scheme than for the background  
modeling one. In their review in 2010, Cristiani et al. [12] distinguished the  
most popular background subtraction algorithms by means of their sensor uti-  
lization: single monocular sensor or multiple sensors. In 2014, Elgammal [12]  
30 provided a chapter on background subtraction for static and moving cameras  
over 120 papers. Since 2008, Bouwmans et al. [13] initiated several comprehen-  
sive surveys classifying each approaches following the employed models that can  
be classified into the following main chronological categories: traditional mod-  
els, recent models and prospective models that employed both mathematical,  
35

machine learning and signal processing models. These different surveys concern either all the categories [14, 15, 16], sub-categories (i.e. statistical models [13], fuzzy models [17], decomposition into low-rank plus additive matrices [18] or part of sub-categories (i.e. Mixture of Gaussian models (GMM) [19], subspace  
40 learning models [20], Robust Principal Component Analysis (RPCA) models [21], dynamic RCPA models [22], and deep learning models [23]).

Sometimes, these previous surveys presented in a sub-part extensions of background subtraction methods to static cameras for moving cameras. One can also find sub-parts that concern moving cameras in object tracking and  
45 surveillance surveys [24, 25, 26, 27]. However, the techniques addressing the case of moving cameras are more and more numerous and can be the target of whole study as proven by recent reviews [12, 28, 29]. In 2014, Elgammal [12] give an entire chapter on background subtraction techniques for moving camera classifying them into traditional and recent methods. In 2018, Komagal and  
50 Yogameena [28] chose to review foreground segmentation approaches with a Pan Tilt Zoom (PTZ) camera but those techniques cannot usually be employed with freely moving cameras. In 2018, Yazdi and Bouwmans [29] presented the most complete survey on the subject, to the best of our knowledge. The methods are presented according to challenges and a classification into four broad categories  
55 are employed but the review suffers from a lack of completeness. Thus, there is a need of a full comprehensive survey for moving objects detection with moving cameras.

In this context, we propose to fully review methods about moving objects detection with a moving camera. The aim is thus to present a review of the  
60 traditional and recent techniques used by categorizing them and making the assessment of the methods regarding the challenges. It is dedicated for students, engineers, young researchers and confirmed researchers in the field. It could serve as basis for courses too and considered as the reference in the field. The paper is organized as follows. First, we define notions of moving objects and  
65 moving cameras in Section 2 in order to delimit the scope of this survey. Second, we investigate the different challenges met in videos taken by static and moving

cameras in Section 3. In Section 4, we carefully present the general process of background subtraction method with a static camera by providing a background knowledge to well understand extensions of background subtraction methods in  
70 the case of moving cameras. In Section 5, we provide an original classification of the methods about moving objects detection with a moving camera. Then, **publicly available datasets, evaluation metrics and f-score results on the most used sequences** are presented in Section 6. Finally, we conclude the paper by a discussion and perspectives for future work.

## 75 **2. Preliminaries**

In this section, we clearly state notions of moving objects and moving cameras that defined the kind of methods that are reviewed in this paper.

### *2.1. Moving objects*

In physics, a motion is described by a change in position of an object over  
80 time according to a frame of reference attached to an observer. In our case, the observer is the camera and we will describe the observations for two kind of cameras: stationary and moving. For a stationary camera, the background appears static in the video stream of the camera and a moving object appears moving. Displacements of an object in the scene is called the *local motion*. In  
85 the case of a moving camera, both of them appear moving. The background appears moving because of the *global motion* and the distinction between a moving object and the static scene is complicated.

The range of moving objects is large, ranging from pedestrians to waving trees. But among of these objects, only a subpart has to be labeled as moving.  
90 The objects like waving trees, ocean waves or escalators are part of so-called *dynamic background* and have to be labeled as background. Conversely, pedestrians, cars or animals are objects with "significantly" motions and the subjects of applications about which we interest in this paper.

An object can be represented in many different ways [25]. In this survey,  
95 we are going to see that two kinds of representation are generally used for the

moving object detection: a bounding box or a silhouette. The bounding box is usually used in tracking methods where only a rough region of the moving object is needed. The bounding box contains pixels from the background and from the moving object. Conversely, a silhouette provides accuracy information  
100 on the moving object position since every pixel in the silhouette has to belong to the object. Silhouette results are needed for some applications like motion capture.

## 2.2. Moving cameras

The specificity of a moving camera compared to a static one, is that a static  
105 object appears moving in the video stream. This motion is caused by the motion of the camera also called the *ego motion*. As well as a moving object, the physics definition of motion can be applied to a camera. In addition to displacements in the 3D space, the camera can also perform rotations, named pan, tilt and roll.

110 Among moving cameras, there are two types of cameras: freely moving camera and constrained moving camera. As its name suggest, freely moving camera performs any kind of motion without any constraint. This camera is hand-held camera, smartphone or drone. In the category of constrained cameras, the most famous example is the PTZ camera. This camera can only perform rotations  
115 since its optical center is fixed. Even if this camera doesn't change in position, rotations are enough to defined it as a moving camera.

## 3. Challenges

Background subtraction is still an open issue with several scientific obstacles to overcome. In 1999, Toyama et al. [30] propose a list of 10 challenges  
120 about background maintenance for video surveillance systems. In this section we provide an extended list about the background subtraction challenges. Each challenge is illustrated by the figures 1, 2 and 3.

- 125 • **Bootstrapping** The construction of the background model is based on several frames of a video sequence but it is possible that each frame contains moving objects. It is necessary to take into account this case to not contaminate the background model and create false detections.
- **Camouflage** Foreground objects can have the same color than the background. Since the detectors are generally based on appearance, it is complicated to differentiate foreground from background.
- 130 • **Dynamic background** The background can contain some elements that are not completely static as water surface or waving trees. Even if they are not static, these elements are part of the background. The background model requires to represent these appearance variations in order to limit false alarms.
- 135 • **Foreground aperture** Changes inside a uniform colored region of a moving object cannot be detected and the corresponding pixels are classified as background rather than foreground.
- **Illumination changes** Illumination changes can be gradual (a cloud in front of the sun) or abrupt (light switch). The difference of illumination  
140 between the current frame and the background model causes false detections.
- **Low frame rate** According to the application, a low frame rate camera can be used. In this case, background changes appear more abrupt and it is more difficult to distinguish background and foreground changes.
- 145 • **Motion blur** Images taken by the camera can be blurred by an abrupt camera motion or by camera jittering. The boundaries of objects are fuzzy and create imprecision in the detection.
- **Motion parallax** The creation or the update of a background model and foreground detection for sequences taken by a moving camera required  
150 frame alignment. 3D scenes with large depth variations present parallax

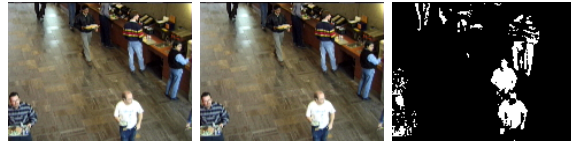
in images. This parallax can cause image registration errors and create non uniform background optical flow and thus affect the detection on several steps.

- 155 • **Moving background** In a stationary camera, static objects appear static and moving objects appear moving. In the case of a moving camera, everything appears moving because of the camera displacement, also called the ego-motion. In these conditions it is more complicated to separate moving objects from the static ones. Moreover, with a moving camera some parts of the scene appear and disappear over time. The background  
160 construction and its maintenance have to deal with this aspect.
- **Moved background object** Static objects can be moved and they should not be considered as foreground. When the object is moved, a new portion of the scene is revealed and when the object is static again an other part of the scene is occluded. The background model should be updated according  
165 to these changes to avoid misclassification.
- **Night video** Images taken at night time present low brightness, low contrast and few color information which reduce appearance changing when an object is moving.
- **Noisy images** Noise in image depends on the quality of the camera components like sensors, lenses, resolutions. The noise affects the appearance  
170 and consequently the background model.
- **Shadows** Every objects create shadows by the interception of light rays. For a moving object, its shadow is moving and it modifies the background appearance. Those shadows should not be integrated into the background  
175 model but it must not be detected as foreground and it must not be integrated into the background model.
- **Sleeping foreground object** It is possible that a moving object stop for a while and this object becomes a static object. Depending on the

application, it is necessary to continue to label this object as moving or  
180 on the contrary incorporate it into the background.

- **Waking foreground object** A static object can starts to move a long  
time after the beginning of the video. A new part of the scene that never  
appears in frames before is revealed. This portion of the scene may cre-  
ates an erroneous detection, called a ghost, since it is not part of the  
185 background.

Following these remarks, we can categorize the challenges by level of difficul-  
ties [31]. In addition, these challenges are less or more predominant depending  
on the real-applications [1]. For example in surveillance in natural environments  
like in maritime and aquatic environments, illumination changes and dynamic  
190 changes in the background are very challenging requiring more robust back-  
ground methods than the top methods of CDnet 2014 as developed by Prasad  
et al. [32, 33, 34]. However, several authors provided tools to visualize and  
analyze the variations causes by theses challenges in the temporal history of the  
pixel [35, 36].



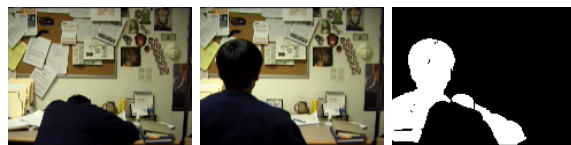
(a) Bootstrapping



(b) Camouflage



(c) Dynamic background



(d) Foreground aperture

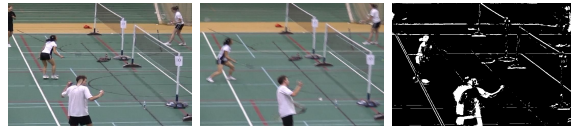


(e) Illumination changes



(f) Low frame rate

Figure 1: Illustrations of background subtraction challenges. Images come from the Wallflower (Bootstrap, Camouflage, ForegroundAperture, WavingTrees, LightSwitch sequences) dataset and the ChangeDetection.net (port.0.17fps sequence) dataset. The last column is the result of Gaussian mixture-based background/foreground segmentation in the OpenCV library.



(a) Motion blur



(b) Motion parallax (\*)



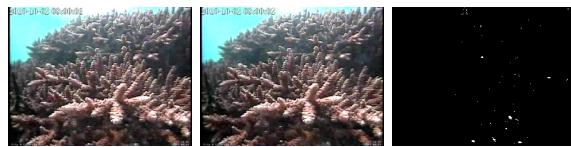
(c) Moved background object



(d) Moving background



(e) Night video



(f) Noisy images

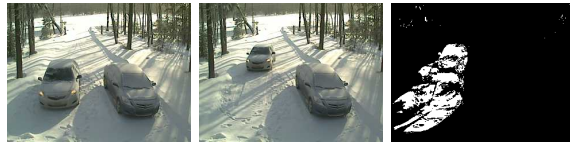
Figure 2: Illustrations of background subtraction challenges. Images come from the Wallflower (MovedObject sequence), the ChangeDetection.net (badminton, continuousPan, busyBoulevard sequences) and the ComplexBackground (Forest) dataset and the Fish4Knowledge (site NPP-3, camera 3, 10/02/2010 sequence) dataset. The last column is the result of Gaussian mixture-based background/foreground segmentation in the OpenCV library. (\* To illustrate the Motion Parallax challenge, frames are register to the first one with a homography estimated by RANSAC on feature points.)



(a) Shadows



(b) Sleeping foreground object



(c) Waking foreground object

Figure 3: Illustrations of background subtraction challenges. Images come from the ChangeDetection.net (PeopleInShade, parking, winterDriveway sequences) dataset. The last column is the result of Gaussian mixture-based background/foreground segmentation in the OpenCV library.

195 **4. Static Cameras**

There are three main categories of approaches to detect moving objects: consecutive frame difference, background subtraction, and optical flow. Consecutive frame difference methods [37, 38, 39] are very simple to implement but they are too sensitive to the challenges. Optical flow methods are more robust  
 200 but are still too time consuming to reach real-time requirements. Background subtraction which is the most popular method to detect moving objects offers the best compromise between robustness and real-time requirements. In the literature, there exist a plenty of methods to detect moving objects by background subtraction and we let readers refer to books [40, 41] and surveys that cover this  
 205 problematic for more details [14, 15, 16, 42, 43]. In this section, we describe the general process of background subtraction, survey the corresponding methods, and also investigate the current and unsolved challenges. This part is crucial to well understand extensions of background subtraction methods in the case of moving cameras.

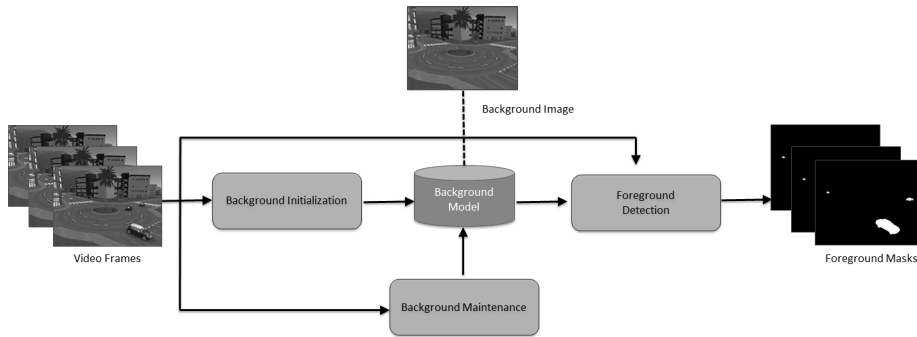


Figure 4: Background subtraction with a static camera, general scheme.

210 As defined in Section 2, from a static point of view only moving objects are moving. From this statement, background subtraction methods follow the general process (See Figure 4). Here, we describe the main process of each step.

#### 4.1. Background Modeling

The background model describes the model use to represent the background.  
215 A big variety of models coming from mathematical theories, machine learning  
and signal processing have been used for background modeling, including crisp  
models [44, 45, 46], statistical models [47, 48, 49, 50], fuzzy models [51, 52, 53],  
Dempster-Schafer models [54], subspace learning models [55, 56, 57, 58, 59],  
robust learning models [60, 61, 62, 63], neural networks models [64, 65, 66] and  
220 filter based models [67, 68, 69, 70].

##### 4.1.1. Mathematical models

Based from mathematical theories, the simplest way to model a background  
is to compute the temporal average [44], the temporal median [45] or the his-  
togram over time [46]. These methods were widely used in traffic surveillance  
225 in 1990s owing to their simplicity but are not robust to the challenges faced in  
video surveillance such as camera jitter, changes in illumination, and dynamic  
backgrounds. To consider the imprecision, uncertainty and incompleteness in  
the observed data (i.e. video), statistical models began being introduced in  
1999 such as single Gaussian [71], Mixture of Gaussians (MOG) [48, 49] and  
230 Kernel Density Estimation [47, 72]. These methods based on a Gaussian distri-  
bution model proved to be more robust to dynamic backgrounds [73, 74]. More  
advanced statistical models were after developed in the literature and can be  
classified into those based on another distribution that alleviate the strict Gaus-  
sian constraint (i.e. general Gaussian distribution [75], Student's t-distribution  
235 [76, 77], Dirichlet distribution [78, 79], Poisson distribution [80, 81]), those based  
on co-occurrence [82, 83, 84] and confidence [85, 86], free-distribution models  
[87, 88, 89], and regression models [90, 91]. These approaches have improved  
the robustness to various challenges over time. The most accomplished meth-  
ods in this statistical category are ViBe [87], PAWCS [89] and SubSENSE [88].  
240 Another theory that allows the handling of imprecision, uncertainty, and incom-  
pleteness is based on the fuzzy concept. In 2006-2008, several authors employed  
concepts like Type-2 fuzzy sets [52, 92, 93], Sugeno integral [94, 95] and Cho-

quet integral [96, 51, 97]. These fuzzy models show robustness in the presence of dynamic backgrounds [92]. Dempster-Schafer concepts were also be employed  
245 in foreground detection [54].

#### 4.1.2. Machine learning models

Based on machine learning, background modeling has been investigated by representation learning (also called subspace learning), support vector machines, and neural networks modeling (conventional and deep neural networks).

- 250 • **Representation learning:** In 1999, reconstructive subspace learning models like Principal Component Analysis (PCA) [55] has been introduced to learn the background in an unsupervised manner. Subspace learning models handle illumination changes more robustly than statistical models [21]. In further approaches, discriminative [56, 57, 58] and mixed [59]  
255 subspace learning models have been used to increase the performance for foreground detection. However, each of these regular subspace methods presents a high sensitivity to noise, outliers, and missing data. To address these limitations, since 2009, a robust PCA through decomposition into low-rank plus sparse matrices [60, 61, 62, 63] has been widely used  
260 in the field. These methods are not only robust to changes in illumination but also to dynamic backgrounds [98, 99, 100, 101]. However, they require batch algorithms, making them impractical for real-time applications. To address this limitation, dynamic robust PCA as well as robust subspace tracking [102, 22, 103] have been designed to achieve a real-time  
265 performance of RPCA-based methods. The most accomplished methods in this subspace learning category are GRASTA [104], incPCP [105], Re-ProCS [106] and MEROP [107]. However, tensor RPCA based methods [108, 109, 110, 111] allow to take into account spatial and temporal constraints making them more robust against noise.
- 270 • **Neural networks modeling:** In 1996, Schofield et al. [66] were the first to use neural networks for background modeling and foreground detec-

tion through the application of a Random Access Memory (RAM) neural network. However, a RAM-NN requires the images to represent the background of the scene correctly, and there is no background maintenance stage because once a RAM-NN is trained with a single pass of background images, it is impossible to modify this information. In 2005, Tavakkoli [112] proposed a neural network approach under the concept of novelty detector. During the training step, the background is divided in blocks. Each block is associated to a Radial Basis Function Neural Network (RBF-NN). Thus, each RBF-NN is trained with samples of the background corresponding to its associated block. The decision of using RBF-NN is because it works like a detector and not a discriminant, generating a close boundary for the known class. RBF-NN methods is able to address dynamic object detection as a single class problem, and to learn the dynamic background. However, it requires a huge amount of samples to represent general background scenarios. In 2008, Maddalena and Petrosino [113, 114, 115, 116] proposed a method called Self Organizing Background Subtraction (SOBS) based on a 2D self-organizing neural network architecture preserving pixel spatial relations. The method is considered as nonparametric, multi-modal, recursive and pixel-based. The background is automatically modeled through the neurons weights of the network. Each pixel is represented by a neural map with  $n \times n$  weight vectors. The weights vectors of the neurons are initialized with the corresponding color pixel values using the HSV color space. Once the model is initialized, each new pixel information from a new video frame is compared to its current model to determine if the pixel corresponds to the background or to the foreground. In further works, SOBS was improved in several variants such as Multivalued SOBS [117], SOBS-CF [118], SC-SOBS [119], 3dSOBS+ [120], Simplified SOM [121], Neural-Fuzzy SOM [122] and MILSOBS [123]) which allow this method to be in the leading methods on the CDnet 2012 dataset [124] during a long time. SOBS show also interesting performance for stopped object detection [125, 126, 127].

But, one of the main disadvantages of SOBS based methods is the need to manual adjust at least four parameters.

305 • **Deep Neural networks modeling:** Since 2016, DNNs have also been successfully applied to background generation [128, 129, 130, 131, 132], background subtraction [133, 134, 135, 136, 137, 138, 139, 140, 141, 142], foreground detection enhancement [143], ground-truth generation [144], and the learning of deep spatial features [145, 146, 147, 148, 149]. More  
310 practically, Restricted Boltzman Machines (RBMs) were first employed by Guo and Qi [128] and Xu et al. [130] for background generation to further achieve moving object detection through background subtraction. In a similar manner, Xu et al. [131, 132] used deep auto-encoder networks to achieve the same task whereas Qu et al. [129] used context-encoder  
315 for background initialization. As another approach, Convolutional Neural Networks (CNNs) has also been employed to background subtraction by Braham and Droogenbroeck [135], Bautista et al. [134] and Cinelli [136]. Other authors have employed improved CNNs such as cascaded CNNs [144], deep CNNs [133], structured CNNs [137] and two stage CNNs  
320 [150]. Since 2018, Sultana et al. [151, 152, 153, 154] investigated the application of Generative Adversarial Networks (GANs) to deal with dynamic backgrounds. Through another approach, Zhang et al. [149] used a Stacked Denoising Auto-Encoder (SDAE) to learn robust spatial features and modeled the background with density analysis, whereas Shafiee  
325 et al. [147] employed Neural Reponse Mixture (NeREM) to learn deep features used in the Mixture of Gaussians (MOG) model [49]. In 2019, Chan [155] 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. Valuable  
330 analysis of DNNs applied to background subtraction are available in the studies of Minematsu et al. [156, 157, 158, 159].

#### 4.1.3. Signal processing models

Based on signal processing, these models considered temporal history of a pixel as 1-D dimensional signal. Thus, several signal processing methods can be used: 1) signal estimation models (i.e. filters), 2) transform domain functions, and 3) sparse signal recovery models (i.e. compressive sensing).

- **Estimation filter:** In 1990, Karmann et al. [160] proposed a background estimation algorithm based on the Kalman filter. Any pixel that deviates significantly from its predicted value is declared foreground. Numerous variants were proposed to improve this approach in the presence of illumination changes and dynamic backgrounds [161, 69, 162]. In 1999, Toyama et al. [70] proposed in their algorithm called Wallflower a pixel-level algorithm which makes probabilistic predictions about what background pixel values are, expected in the next live image using a one-step Wiener prediction filter. Chang et al. [67, 163] used a Chebychev filter to model the background. All these filters approaches reveal good performance in the presence of slow illumination change but less when the scenes present complex dynamic backgrounds.
- **Transform domain models:** In 2005, Wren and Porikli [164] estimated the background model that captures spectral signatures of multimodal backgrounds using Fast Fourier Transform (FFT) features through a method called Waviz. Here, FFT features are then used to detect changes in the scene that are inconsistent over time. In 2005, Porikli and Wren [165] developed an algorithm called Wave-Back that generated a representation of the background using the frequency decompositions of pixel history. The Discrete Cosine Transform (DCT) coefficients are used as features are computed for the background and the current images. Then, the coefficients of the current image are compared to the background coefficients to obtain a distance map for the image. Then, the distance maps are fused in the same temporal window of the DCT to improve the robustness against noise. Finally, the distance maps are

thresholded to achieve foreground detection. This algorithm is efficient in the presence of waving trees.

- **Sparse signal recovery models:** In 2008, Cevher et al. [166] were the first authors who employed a compressive sensing approach for background subtraction. Instead of learning the full background, Cevher et al. [166] learned and adapted a low dimensional compressed representation of it which is sufficient to capture changes. Then, moving objects are estimated directly using the compressive samples without any auxiliary image reconstruction. But, to obtain simultaneously appearance recovery of the objects using compressive measurements, it needs to reconstruct one auxiliary image. To alleviate this constraint, numerous improvements were proposed in the literature [167, 168, 169, 170, 171] and particular good performance is obtained by Bayesian compressive sensing approaches [172, 173, 174, 175].
- **Graph Signal Processing:** In 2020, Giraldo et al. [176] proposed an algorithm of background subtraction based on GSP, this method was called GraphBGS. The algorithm models the instances on videos as nodes embedded in a high dimensional space. Afterwards, a recovery algorithm of graph signals is applied to classify if certain instance is a static or moving object. GraphBGS shows robust performance in the case of unseen videos.

#### 4.2. Background initialization

This step consists in computing the first background image and it is also called *background generation*, *background extraction* and *background reconstruction*. The background model is initialized with a set of images taken before the moving objects detection process. Several kind of models could be used to initialize the background and they are classified as methods based on temporal statistics [177, 178, 179], methods based on sub-sequences of stable intensity [180, 181, 182, 183, 184, 185, 186], methods based on missing data reconstruction problem [187, 188], methods based on iterative model completion [189],

methods based on conventional neural networks [119, 190], and methods based on optimal labeling [191]. The most accomplished methods applied to the SBM-net dataset [192] are Motion-assisted Spatio-temporal Clustering of Low-rank (MSCL) designed by Javed et al. [99], and LaBGen and its variants developed  
395 by Laugraud et al. [184, 185, 186]. For more details, the reader can refer to comprehensive surveys of Maddelena and Petrosino [193, 194, 195, 192].

#### 4.3. Updating background model

In order to overcome background changes (illumination changes, dynamic background, and so on), the background model is updated with information  
400 provided by the current frame taken by the camera. The update rules depend on the model chosen but they generally try to employ old data with the new one according to a learning rate. The choose of the learning rate allow to integrate more or less rapidly the changes to the background. The maintenance of the background model is a critical step since some parts of the foreground could be  
405 integrated in the background and create false-alarms. However, the background maintenance process requires an incremental on-line algorithm, since new data is streamed and so dynamically provided. The key issues of this step are the following ones:

- **Maintenance schemes:** In the literature, three maintenance schemes  
410 are present: the blind, the selective, and the fuzzy adaptive schemes [196]. The blind background maintenance updates all the pixels with the same rules which is usually an IIR filter. The main disadvantage of this scheme is that the value of pixels classified as foreground are used in the computation of the new background and so polluted the background image.  
415 To solve this problem, some authors used a selective maintenance scheme that consists of updating the new background image with different learning rate depending on the previous classification of a pixel into foreground or background. Here, the idea is to adapt very quickly a pixel classified as background and very slowly a pixel classified as foreground. But  
420 the problem is that erroneous classification may result in a permanent

incorrect background model. This problem can be addressed by a fuzzy adaptive scheme which takes into account the uncertainty of the classification. This can be achieved by graduating the update rule using the result of the foreground detection such as in El Baf et al. [196].

- 425 • **Learning rate:** The learning rate determines the speed of the adaptation to the scene changes. It can be fixed, or dynamically adjusted by a statistical, or a fuzzy method. In the first case, the learning rate is fixed as the same value for all the sequence. Then, it is determined carefully such as in [197] or can be automatically selected by an optimization algorithm [198]. However, it can take one value for the learning step and one for the maintenance step [199]. Additionally, the rate may change over time following a tracking feedback strategy [200]. For the statistical case, Lee [201] used different learning rates for each Gaussian in the MOG model. The convergence speed and approximation results are significantly improved. For the fuzzy case **(3)**, Sigari et al. [202, 203] computed an adaptive learning rate at each pixel with respect to the fuzzy membership value obtained for each pixel during the fuzzy foreground detection. In another way, Maddalena and Petrosino [117, 118] improved the adaptivity by introducing spatial coherence information.
- 440 • **Maintenance mechanisms:** The learning rate determines the speed of adaptation to illumination changes but also the time a background change requires until it is incorporated into the model as well as the time a static foreground object can survive before being included in the model. So, the learning rate deals with different challenges which have different temporal characteristics. To decouple the adaptation mechanism and the incorporation mechanism, some authors [204][205] used a set of counters which represents the number of times a pixel is classified as a foreground pixel. When this number is larger than a threshold, the pixel is considered as background. This gives a time limit on how long a pixel can be considered as a static foreground pixel.

- **Frequency of the update:** The aim is to update only when it is needed. The maintenance may be done every frame but in absence of any significant changes, pixels are not required to be updated at every frame. For example, Porikli [206] proposed adapting the time period of the maintenance mechanism with respect to an illumination score change. The idea is that no maintenance is needed if no illumination change is detected and a quick maintenance is necessary otherwise. In the same idea, Magee [207] used a variable adaptation frame rate following the activity of the pixel, which improves temporal history storage for slow changing pixels while running at high adaption rates for less stable pixels.

#### 4.4. Foreground detection

As the name of the technique suggests it, the foreground is detected by subtracting the background to the current frame. A too high difference, determined by a threshold, points the foreground out. The output is a binary image so-called a mask for which each pixel is classified as background or foreground. Thus, this task is a classification one, that can be achieved by crisp, statistical or fuzzy classification tools. For this, the different steps have to be achieved:

- **Pre-processing:** The pre-processing step avoids the detection of unimportant changes due to the motion of the camera or the illumination changes. This step may involve geometric and intensity adjustments. As the scenes are usually rigid in nature and the camera jitter is small, geometric adjustments can often be performed using low-dimensional spatial transformations such as similarity, affine, or projective transformations. On the other hand, there are several ways to achieve intensity adjustments. The pixel intensity values in the current image are then normalized to have the same mean and variance as those in the background image. Another way consists in using a homomorphic filter based which is based on the shading model. This approach permits to separate the illumination and the reflectance. As only the reflectance component contains information

about the objects in the scene, illumination-invariant foreground detection [209, 210, 211] can hence be performed by first filtering out the illumination component from the image.

- 485 • **Test:** The test which allows to classify pixels of the current image as background or foreground is usually the difference between the background image and the current image. This difference is then thresholded. Another way to compare two images are the significance and hypothesis tests. The decision rule is then cast as a statistical hypothesis test. The decision as to whether or not a change has occurred at a given pixel corresponds to  
490 choosing one of two competing hypotheses: the null hypothesis  $H_0$  or the alternative hypothesis  $H_1$ , corresponding to no-change and change decisions, respectively. Several significance tests can be found in the literature [212, 213, 214, 215, 216, 217, 218, 219].
  
- 495 • **Threshold:** In literature, there are several types of threshold schemes. First, the threshold can be fixed and the same for all the pixels and the sequence. This scheme is simple but not optimal. Indeed, pixels present different activities and it needs an adaptive threshold. This can be done by computing the threshold via the local temporal standard deviation of intensity between the background and the current images, and by updating  
500 it using an infinite impulse response (IIR) filter such as in Collins et al. [220]. An adaptive threshold can be statistically obtained also from the variance of the pixel such as in Wren et al. [71]. Another way to adaptively threshold is to use fuzzy thresholds such as in the studies of Chacon-Muguia and Gonzalez-Duarte [221].
  
- 505 • **Post-processing:** The idea here is to enhance the consistency of the foreground mask. This can be done firstly by deleting isolated pixels with classical or statistical morphological operators [222]. Another way is to use fuzzy concepts such as fuzzy inference between the previous and the current foreground masks [223].

510 Moreover, foreground detection is a particular case of change detection when  
(1) one image is the background and the other one is the current image, and  
(2) the changes concern moving objects. So, all the techniques developed for  
change detection can be used in foreground detection. A survey concerning  
change detection can be found in [208, 224].

#### 515 4.4.1. Solved and Unsolved Challenges

For fair evaluation and comparison on videos presenting challenges described  
in CDnet 2014 dataset [225] which was developed as part of Change Detection  
Workshop challenge (CDW 2014). This dataset includes all the videos from the  
CDnet 2012 dataset [124] plus 22 additional camera-captured videos provid-  
520 ing 5 different categories that incorporate challenges that were not addressed  
in the 2012 dataset. The categories are as follows: baseline, dynamic back-  
grounds, camera jitter, shadows, intermittent object motion, thermal, challeng-  
ing Weather, low frame-rate, night videos, PTZ and turbulence. In 2015, Jodoin  
[226] did the following remarks regarding the solved and unsolved challenges by  
525 using the experimental results available at CDnet 2014:

- Conventional background subtraction methods can efficiently deal with  
challenges met in "baseline" and "bad weather" sequences.
- The "Dynamic backgrounds", "thermal video" and "camera jitter" cate-  
gories are a reachable challenge for top-performing background subtrac-  
530 tion.
- The "Night videos", "low frame-rate", and "PTZ" video sequences repre-  
sent significant challenges.

However, Bouwmans et al. [23] analyzed the progression made over 20 years  
from the MOG model [49] designed in 1999 up to the recent deep neural net-  
works models developed in 2019. To do so, Bouwmans et al. [23] computed  
535 different key increases in the F-measure score in terms of percentage by con-  
sidering the gap between MOG [49] and the best conventional neural network

(SC-SOBS [119]), the gap between SC-SOBS [119] and the best non-parametric multi-cues methods (SubSENSE [88]), the gap between SuBSENSE [88] and Cascaded CNNs [144], the gap between SuBSENSE [88] and the best DNNs based method (FgSegNet-V2 [227]), and the gap between FgSegNet-V2 [227] and the ideal method (F-Measure= 1 in each category). The big gap has been obtained by DNNs methods against SuBSENSE with 24.31% and 32.92% using Cascaded CNN and FgSegNet-V2, respectively. The gap of 1.55% that remains between FgSegNet-V2 and the ideal method is less than the gap of 6.93% between Cascaded CNN and FgSegNet-V2. Nevertheless, it is important to note that the large gap provided by cascaded CNN and FgSegNet-V2 is mainly due to their supervised aspect, and a required drawback of training using labeling data. However, when labeling data are unavailable, efforts should be concentrated on unsupervised GANs as well as unsupervised methods based on semantic background subtraction [228, 229], and robust subspace tracking [107, 103, 230, 105, 102, 22] that are still of interest in the field of background subtraction. Furthermore, deep learning approaches detect the changes in images with static backgrounds successfully but are more sensitive in the case of dynamic backgrounds and camera jitter, although they do provide a better performance than conventional approaches [231]. In addition, several authors avoid experiments on the "IOM" and the "PTZ" categories. In addition, when the F-Measure score is provided for these categories, the score is not very high. Thus, it seems that the current deep neural networks tested face problems in these cases perhaps because they have difficulties in how to learn the duration of sleeping moving objects and how to handle changes from moving cameras. However, even if background subtraction models designed for static cameras progress for camera jitter and PTZ cameras as with several RPCA models [230, 232, 233, 234, 235, 236] and deep learning models [227, 237, 238], they can only handle small jitter movements or translation and rotation movements. Thus, detection of moving objects with moving cameras required more dedicated strategies and models that we reviewed in this survey.

## 5. Moving Cameras

The background subtraction that we have just presented here is designed for static camera cannot be applied directly to moving camera since the background is no longer static in the images. Most of the methods that we present in this paper are adaptations or inspirations of the idea of background subtraction to a moving camera.

In the case of a moving camera, the result of foreground detection depends on the background representation. We choose to categorize the methods of moving objects detection with a moving camera by the type of background representation chosen to solve the problem. The figure 5 presents the taxonomy adopted in this survey.

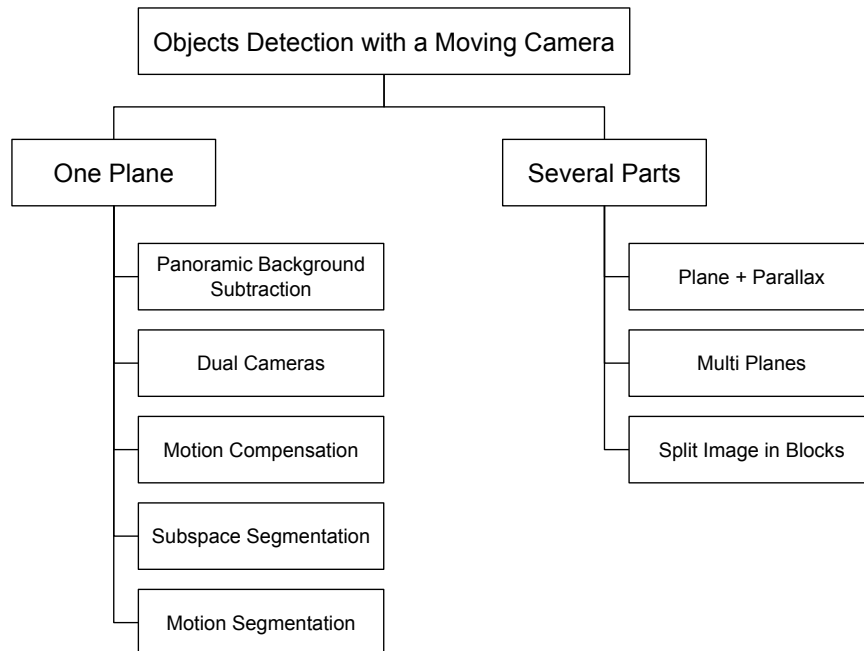


Figure 5: The taxonomy adopted in this survey.

### 5.0.1. Panoramic background subtraction

580 The images captured by a moving camera can be stitch together to form a bigger image so-called a panorama or a mosaic as shown in the figure 6 . This panorama can be used to model the background and detect moving objects as for a static camera.

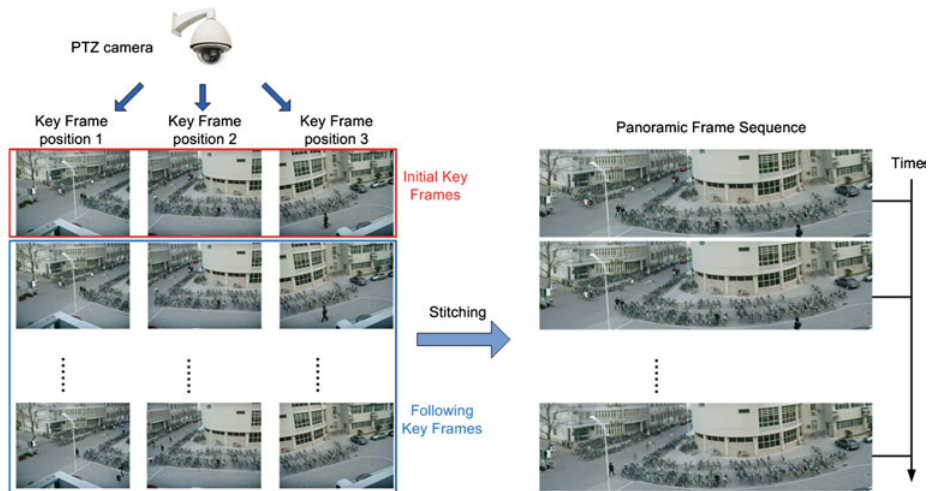


Figure 6: An example of a technique to construct a panoramic background model.

Source: Images from Xue et al. [239].

585 The construction of a panoramic view is a key step that needs high accuracy [240, 241, 242, 243]. There are three techniques to align the images to construct the mosaic:

- **Frame to frame:** alignment parameters are computed for each pair of successive frames for the entire sequence. All the frames are then aligned to a fixed coordinate system, given by the reference frame or a virtual coordinate system. The problem with this mosaic construction is that errors may accumulate during the alignment to the fixed coordinate system.
- **Frame to mosaic:** since the mosaic is larger than a frame, large displacement has to be handle to align a frame to the mosaic. To manage it, the parameters between the previous frame and the mosaic are used as

595 an initial estimation since they are closed to those between the new frame  
and the mosaic.

- **Mosaic to frame:** contrary to the two previous alignment techniques, the mosaic is aligned to the new frame. There is no static coordinate system and the current image is maintained in its input coordinate systems.

600 The two first techniques, *frame-to-frame* and *frame-to-mosaic* are widely used in the construction of a mosaic for the moving object detection problem.

In order to warp images to form a mosaic, two motions are generally used: the affine or the perspective motion model [244, 245]. The perspective transformations better fit the camera transformation but in some cases the affine motion  
605 model can be sufficient and it is also faster since there are only six parameters to estimate against eight for the perspective one. In both cases, a refinement step is generally performed to correct misalignment errors.

In 2000, Mittal et al. [246] construct the panorama by registering an image to the entire mosaic in order to limit cascading of registration errors with the frame  
610 to frame technique. The registration is performed by an affine transformation based on the Kanade-Lucas-Tomasi (KLT) feature tracker [247] which is refined by using the Levenberg-Marquardt method. In an other work, Bartoli et al. [248] combine the direct method and the feature based method to construct a panorama. The feature based method is used to obtain a first estimation of  
615 the panorama. Then a direct method refines each frame registration. To deal with real-time and accuracy, Azzari et al. [249] use a feature-based method to construct a panorama where the outlier features are filtered by a simple but efficient clustering method in order to estimate a projective transformation with only features that result to the camera ego-motion. The frame-to-frame  
620 alignment errors are fixed by a two-stage registration based on the frame-to-mosaic technique. In an other work, Xue et al. [239] choose a feature based method to construct a mosaic with key frames of which positions are manually chosen. The background model is a Panoramic GMM (PGMM), extended from

the model proposed by Friedman and Russell [250]. The method proposed in  
2007 by Brown and Lowe [243] to build a panorama is used by Xue et al. [251],  
Zhang et al. [252] and more recently by Avola et al. [253]. This method performs  
a mosaic with unordered images by using a Frame-to-Mosaic approach. In the  
work of Sugaya and Kanatani [254] feature points that belong to the background  
are selected by fitting a 2D affine space to the feature point trajectories. These  
features points are then used to estimate homographies by the re-normalization  
method [255]. While most approaches used feature points to estimate their  
transformation, the method of Amri et al. [256] operates on both regions and  
points of interest. In an other approach, Vivet et al. [257] compute the global  
motion with the Multiple Kernel Tracking [258] method on small uniformly  
selected regions. This approach is computationally light and doesn't need a lot  
of memory. Some authors use a priori knowledge or measured data to register  
a pair of image as in the work of Kang et al. [259] where the focal length  
and the size of the CCD sensor are known. When the telemetry information  
is available from their airborne sensor, Ali and Shah [260] combined the angles  
with a feature based approach and a direct method. Rather than improve the  
image alignment, Hayman et al. [261] choose to improve the GMM proposed  
by Stauffer and Grimson [262] to handle image noise and calibration errors.  
In 2019, Moore et al. [263] construct a panorama by computing homography  
between consecutive frames. To separate the background and the foreground,  
a low-rank representation is used based on the robust PCA method. In order  
to obtain better decomposition, the authors use the OptShrink, rather than the  
singular value thresholding, and a weighted total variation regularization.

In 2014, Hamid et al. [264] use two videos: one with foreground objects and  
one with only the background. Each video is recorded with a handheld camera  
and both camera paths are closed each other. The use of the background video  
is similar to a panorama: foreground frame are matched to background frames  
in order to perform a background subtraction. The authors propose a spatio-  
temporal matching with a significant computational speed-up. Rather than  
construct a panorama, Chelly et al. [265] prefer to represent the background

655 by several subspaces. Each portion of the scene is learned individually based  
on frames registered to a global coordinate system and each portion partially  
overlap each other. Thanks to this representation large scene can be handle.

After image registration, the last step to construct a mosaic is the blending  
step. It consists of mixing pixels that belong to the overlap region of images  
660 when they are warped together.

Several approaches exist, from simple ones like the triangular weighting func-  
tion used by Bhat et al. [266] to more complex ones as the multi-band blending  
used by Xue et al. [239]. In an other work, Amri et al. [256] choose to use  
the *temporal median operator*. The advantage of the temporal median scheme  
665 is that it can remove foreground from the mosaic since it supposes that moving  
objects doesn't stay at the same location more than half time during the initial-  
ization step. In 2005, Azzari et al. [249] use the alpha-update rule also known  
as the Infinite Impulse Response (IIR) filter used for background maintenance.  
In a further work, Bevilacqua and Azzari [267] reduce the seam effects on the  
670 panorama by performing a tonal alignment on gray scale images. To do that,  
the authors use an intensity mapping function on histograms.

To compute the foreground detection on the current frame, it is necessary  
to register the image to the background.

In 2000, Bhat et al. [266] make use of the panorama building step to store  
675 information that are then used to register a new frame to the mosaic. For each  
frame that constitute the panorama, the pan and the tilt angles and the affine  
parameters are stored. The rotation angles of the new frames are used with the  
stored information to obtain a first coarse registration which is refined by the  
estimation of transformation parameters between the new frame and the rough  
680 mosaic region. In an other work, Xue et al. [251] use feature points and camera  
parameters saved during the panoramic building step to register the current im-  
age to the panorama. A gray-level histogram is computed for the background  
and the current image where the pixel value distributions are previously nor-  
malized to prevent lighting changes. The Kullback-Leiber Divergence is then

685 used to obtain the foreground probabilities of each pixel and finally thresholded  
to compute the foreground mask.

The image registration with a PTZ camera is a complex task because the image can be taken at the different scale from the background. To overcome this problem, Zhang et al. [252] capture images at different focal length and  
690 these images are group according to the focal length. When the current image is register to the mosaic with the feature points, the sets of feature points attached to each group of mosaic images are enlarged with the new matched feature points. In an other approach, Xue et al. [239] propose a new multi-layered propagation method that cope with the number of matching features  
695 points between the current frame and the panorama that decreases when the scale of the current frame increases. A hierarchy of image at different scales is constructed where a layer groups frames taken at the same scale and layers are linked together by matching feature points. The hierarchy of layers is then used to register the current frame to the panorama by propagating correspondences  
700 through the layers. The foreground detection is computed by thresholding the minimum Mahalanobis distance between a pixel and a block centered on the corresponding background pixel. The multi-layered system is also used by Liu et al. [268] but to represent the background and not to register the current frame to a panorama. Each layer is composed of a set of key frames where  
705 key frames are encoded with a spatio-temporal model. The current frame is registered with the pan, the tilt angle and the focal to find the nearest key frames and a homography is computed for the registration. **Recently, Chelly et al. [265] train a regression net to learn 2D transformation of each frame to a global coordinate system. The transformation is refined by using local subspaces that overlap the frame.**  
710

In 2008, Asif et al. [269] choose to analyze the global motion by block in the image. The phase correlation is used to determinate the motion of each block which permit to obtain a first foreground estimation. Foreground blocks are divided into smaller blocks to refine the label by analyzing the sum of absolute  
715 difference for each block and their neighbors. In an other work, Ali and Shah

[260] suggest to use two methods to obtain foreground objects: accumulative frame differencing and background subtraction. A histogram of log-evidence is combined with the result of a hierarchical background subtraction to detect moving objects. In an recent work, Avola et al. [253] propose to attach a spatio-temporal structure to each keypoints. The spatio-temporal information is used to track background feature points and label them as background or foreground. A clustering stage is also applied on keypoints to validate the foreground labeling. When two objects are represented by only one blob, because of noise or shadows, Kang et al. [259] analyze the vertical projection histogram and use it to correct the segmentation.

References	Main contribution	FF	FM	AB	FB	DM	AM	PM
Mittal et al. (2000) [246]	Mosaic building with moving objects	×	✓	×	✓	✓	✓	×
Bhat et al. (2000) [266]	Mosaic building & background modeling	×	×	✓	×	×	×	×
Bartoli et al. (2002) [248]	Mosaic building	✓	×	×	✓	✓	×	×
Hayman et al. (2003) [261]	Background modeling	✓	×	×	✓	×	×	✓
Kang et al. (2003) [259]	Reduce segmentation noise	×	×	✓	✓	×	×	×
Sugaya et al. (2004) [254]	Mosaic building	×	✓	×	✓	×	✓	×
Azzari et al. (2005) [249]	Mosaic building	✓	✓	×	✓	×	×	✓
Ali and Shah (2006) [260]	Methods combined	✓	×	✓	✓	✓	×	✓
Bevilacqua et al. (2006) [267]	Tonal alignments	✓	✓	×	✓	×	×	✓
Asif et al. (2008) [269]	Moving objects detection	✓	×	×	✓	✓	×	×
Vivet et al. (2009) [257]	Mosaic building	×	✓	×	×	×	×	✓
Amri et al. (2010) [256]	Temporal median operator	✓	×	×	✓	×	×	✓
Xue et al. (2010) [251]	Mosaic building	×	✓	×	✓	×	×	✓
Zhang et al. (2010) [252]	Large zoom	×	✓	×	✓	×	×	✓
Xue et al. (2013) [239]	Large zoom	×	✓	×	✓	×	×	×
Hamid et al. (2014) [264]	Speed-up frame matching	×	×	×	✓	×	×	✓
Liu et al. (2015) [268]	Hierarchical background model	×	×	✓	✓	×	×	✓
Avola et al. (2017) [253]	Spatio-temporal keypoints tracking	×	✓	×	✓	×	✓	×
Moore et al. (2019) [263]	Panoramic low-rank background	✓	×	×	✓	×	×	✓
Chelly et al. (2020) [265]	Partially-overlapping local subspaces	×	✓	×	×	✓	✓	×

Table 1: Panoramic methods summary. **FF**: Frame-to-Frame, **FM**: Frame-to-Mosaic, **AB**: Angle Based, **FB**: Feature Based, **DM**: Direct Method, **AM**: Affine Model, **PM**: Projective Model.

### 5.0.2. Dual cameras

Instead of construct a panorama, some methods use a dual-camera system where one of the two cameras has a wide focal of view to observe the whole scene.

730 The camera calibration is an important step to make use of information provided by several cameras. Autocalibration is generally used contrary to calibration which necessitate some device whose the best-known example is the chessboard.

In 2008, Chen et al. [270] propose a homography calibration with polyno-  
735 mials without prior knowledge but at the cost of a slightly degraded mapping accuracy. In an other work, Horaud et al. [271] estimate the intrinsic parameters of both cameras and use 3D patterns for the stereo calibration. Another calibration step, named the kinematic calibration and based on the epipolar geometry, is used to rotate the PTZ camera. To achieve real-time computation, Kumar et al. [272] construct an offline look-up table with different pan and tilt  
740 angles. A neural network is then trained offline with the look-up table and the result is used to interpolate any PTZ orientation during the online image registration process. In an other work, Lim et al. [273] first compute zero-positions between the static camera and the PTZ ones. The pan and tilt angles needed  
745 to track an object are derived from the projective geometry equations and image point trajectories. Several static and PTZ cameras are used in the work of Krahnstoeber et al. [274]. To calibrate their cameras in the same coordinate system, the authors use the *foot-to-head homology* combined with a Bayesian formulation to handle measurement uncertainties [275, 276].

750 Motion detection is usually performed in two step, firstly in the static camera to indicate where the moving camera has to look before it performs moving objects detection too.

In 2003, Lim et al. [273] use the method proposed by Elgammal et al. [278] designed for stationary cameras. This method is based on non-parametric back-  
755 ground representation which handle dynamic background and shadows. In 2014,

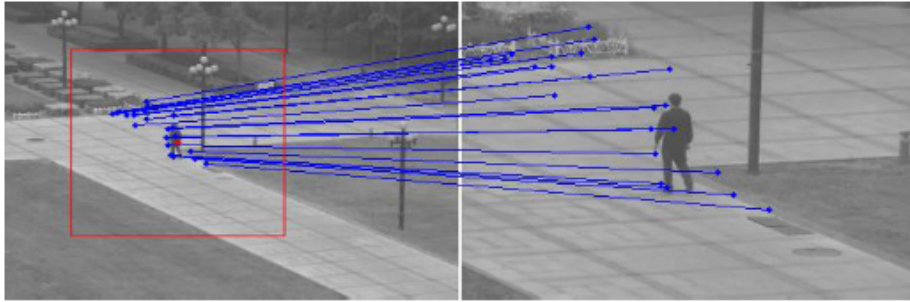


Figure 7: An example of image registration between a large-view static camera and local-view PTZ camera.

Source: Images from Cui et al. [277].

Cui et al. [277] use two cameras: a large-view static camera at low resolution and a local-view PTZ camera at high resolution. The images from the static camera are used for the background model and moving objects are detected in the images of the PTZ camera. Images are registered in three steps: a rough region is obtained with mean-shift, a 2D transformation is computed from features points with the RANdom SAMple Consensus (RANSAC) algorithm [279], the transformation is refined with the Sum Squared Difference (SSD) method. To refine the foreground area, Horaud et al. [271] compare three aligned images.

References	Main contribution	APK	A	FB	DM	AM	PM
Lim et al. (2003) [273]	Image based control technique	×	✓	✓	×	×	✓
Horaud et al. (2006) [271]	Calibration with epipolar geometry	✓	✓	×	✓	×	✓
Chen et al. (2008) [270]	Two spatial mapping methods	×	✓	×	×	×	✓
Krahnstoever et al. (2008) [274]	Combine several cameras	×	✓	×	×	×	×
Kumar et al. (2009) [272]	Real time rectification method	×	✓	✓	×	×	✓
Cui et al. (2014) [277]	A three-step image registration	×	×	✓	×	✓	×

Table 2: Dual camera methods summary. **APK**: A Priori Knowledge, **A**: Autocalibration, **FB**: Feature Based, **DM**: Direct Method, **AM**: Affine Model, **PM**: Projective Model.

### 5.0.3. Motion compensation

765 One simplest technique to adapt the background subtraction method to a  
moving camera is to compensate the motion of the camera in order to realize  
the subtraction as in a stationary camera case. Those methods used *Motion*  
*Compensation* techniques to register the current image with the background  
model with a 2D parametric transformation [280, 281]. After the registration  
770 step, images are configured as with a static camera and background subtraction  
techniques can be applied on the registered frame. Nevertheless the global  
estimation of the 2D transformation of the current frame with a previous one  
or a background model lead to foreground false alarms due to the registration  
errors as shown by the figure 8 and generally a refinement step is necessary.

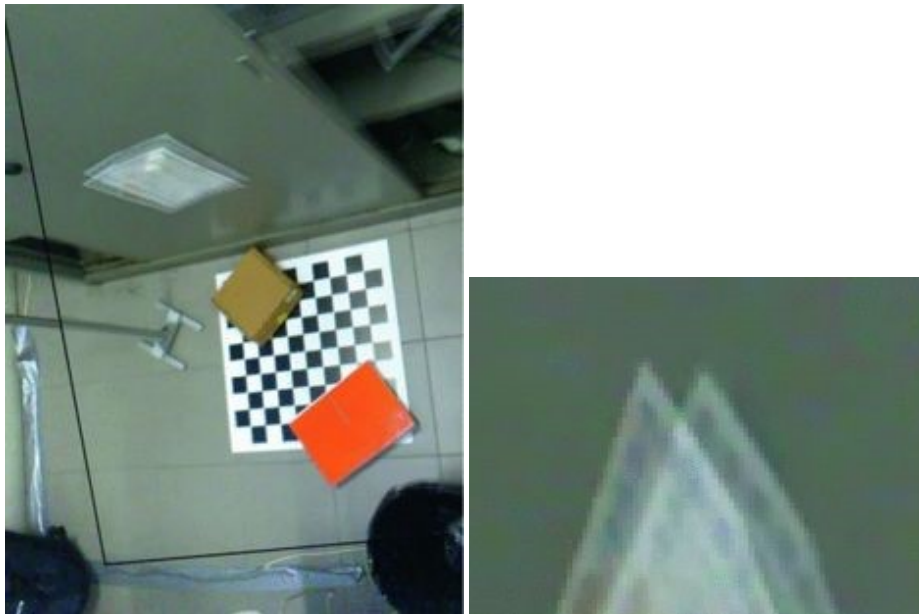


Figure 8: An example after image registration with a homography. The 2D transformation is based on the floor and we observe that the closet is misaligned. The second picture clearly shows this misalignment on the sheet paper.

*Source:* Images from Romanoni et al. [282].

775 Contrary to previous methods presented in section 5.0.1, the background

model is not an extended image as a panorama but an image with the same resolution as a frame taken by the moving camera. From one frame to another the visible part of the background changes over time since the camera is moving. The background image at a time  $t$  is composed of previous scene parts still visible  
780 in the camera field of view and new scene parts that appear in the current image. The background subtraction with motion compensation can also be used with a PTZ camera [283, 284, 285]. In that case, instead of creating a panorama with several images, the background model has the size of a frame. This reduces the computation time and the memory allocation needed for the whole subtraction  
785 process.

To reduce errors in the final mask, some authors choose to use two models [286, 287, 288, 289, 290, 291]. In 2011, Wu et al. [286] compute background and foreground maps in a joint spatial-color domain with the Kernel Density Estimation (KDE) method applies on the previous pixel classification. The  
790 spatial-color cue is used with contrast and motion cues to obtain a segmentation by a Conditional Random Field (CRF) energy minimization. In an other work, Wan et al. [287] construct two GMM for each feature points, based on the mean and the variance of background and foreground clusters. A foreground feature point is removed from the foreground set if its probability to belong to  
795 the foreground model is less than belong to the background model. In a recent work, Zhao et al. [291] use two confidence images: the foreground confidence image preserved the proximity captured by a GMM while the background confidence image preserves set of background spatio-temporal features. Two models are used in the work of Lopez-Rubio et al. [288] for two different tasks: one to  
800 estimate the motion of the camera and the other one to compute the foreground. For both models, one Gaussian component represents the background and one uniform component represents the foreground. The first model is in the RGB space while the second one uses 24 features. In 2016, Kurnianggoro et al. [289] and, in a recent work by the same authors, Yu et al. [290] use a background  
805 model and a candidate background model. The candidate background guarantees that a pixel is stable on a given period before add it in the background

model. In 2014, Ivanov [292] uses two background models. The first model is used for the pixel classification of the current frame while the second one is used to reduce error in the classification. The second model is used to transform  
810 the first model into a coordinate system closer to the current frame in order to overlap a larger region.

In 2014, Ferone and Maddalena [293] propose to use a neural map as background model. This map is an enlarged version of a frame where each pixel is represented by  $n \times n$  weight vectors. When a pixel find a match with the back-  
815 ground model, the corresponding neuron in the map is updated and also its neighborhood in order to take into account spatial relationship. More recently, Avola et al. [294] perform the background subtraction on a modified version of the neural background subtraction proposed by Maddalena and Petrosino [295]. To save computation time and reduce the noise, only areas of foreground  
820 clusters, computed with foreground feature points, are used in the background subtraction process.

Image registration is done by estimating a 2D transformation between the current image and a previous one or the background model. In 1994, Murray and Basu [283] use the focal length and the pan and tilt rotations given by  
825 potentiometers to estimate the position of a pixel in the previous frame. In an other work, Tzanidou et al. [296] use the telemetry information provided by GPS and the intrinsic parameters of an Unmanned Aerial Vehicles (UAV). Those information are completed by a global registration computed with a Discrete Fourier Transform (DFT) in order to improve the translational accuracy.  
830 Rather than using a priori knowledge, the computation of alignment parameters can be performed with feature-based [297] or direct [298] methods. Generally the feature based method is preferred (see 3) because it is fast to compute and the features usually used are the well-known feature points [299, 247]. To save computation time and reach real-time performance, Micheloni and Foresti [300]  
835 use a Fast Feature Selection (FFS) which select good feature points based on the quality criterion of Tomasi and a map of good feature points is maintained rather than extract features from scratch. In 2014, Kryjak et al. [301] esti-

mate the global displacement by matching feature points according to a grid of regular blocks. The motion model between two consecutive images is the median of the flow vectors. In an other approach, Robinault et al. [284] estimate a homography with a minimization algorithm and accelerate the computation time by using a cost function based on the location of feature points. To reject bad homography estimation, Lopez-Rubio et al. [288] propose to find "minor errors" which occur when the model is too large or too small. A new homography is then computed based on new features points. If 10 consecutive minor errors occur then it is a severe error and the current frame is skipped. With three consecutive severe errors, both models are reset with the current frame. Since the camera is moving, some images can be blurred by the motion and this affects the accuracy of feature points detection and matching. To prevent that, Kadim et al. [285] find vertical edges compute the average absolute edge magnitude to evaluate the blurriness level of the current image and only keep images taken when the camera is in a stable position. Recently, Allebosch et al. [302] propose a robust method to compensate the motion of a PTZ camera by estimating the camera parameters. First the focal and the tilt angle are estimated by tracking feature points and using a non-linear estimation. Two motion models are then used, fixed tilt angle and simultaneous pan/tilt, to estimate and compensate the camera motion. The authors show by experimentation that their method is robust when few feature are matches. In order to save more computational time, some authors choose to select points on a grid and track them with optical flow or with well-known track methods as the KLT [303, 289, 304, 305, 290].

Feature points that belong to foreground should not be used to compute the 2D transformation and Wan et al. [287] propose a two-layer iteration to estimate the transformation parameters. In the inner layer, the RANSAC algorithm is used to obtain a transformation model while in the outer layer the transformation parameters are used to classify feature points as background or foreground. The new background feature points are used to estimate a new transformation model until the classification converge. In 2010, Guillot et al. [306] reduce matching candidates for a feature point by using a small search

870 window to match more points. In a further work, Guillot et al. [307] propose a modified version of the SURF descriptor for low textured areas. In 2018, Avola et al. [294] compute a first background/foreground classification on feature points according to a homography estimated on matches. Points that belong to the background are then used to estimate an affine transformation used for the camera motion compensation.

875 In theory, after the registration step, the background model and the current frame are aligned and a foreground detection used with static camera can be applied. In practice the current frame is not perfectly aligned because of parallax generated by 3D objects that do not belong to the 2D plane described by the 2D transformation.

880 A common way to handle the parallax is to use the neighborhood of a pixel to classify. In 1997, Odobez and Boutemy [280] use only motion measurements rather than intensity change measurements. These measurements are embedded in a multiscale Markov Random Field (MRF) framework to encourage neighboring pixels to have the same label. A voting technique is proposed by Paragios and Tziritas [308] to choose the regularization parameter of the cost function to minimize to obtain a binary mask. In an other work, Ren et al. [309] propose a Spatial Distribution of Gaussians (SDG) model to provide a temporal and spatial distribution of the background where the authors assume that the intensity distribution of each pixel can be modeled by a two-component MOG. The methods proposed by Kim et al. [310] and Viswanath et al. [311] compared the intensity of a pixel labeled as foreground and the intensities of its neighborhood in the background model. A low difference between intensities means a false alarm but the silhouette of a moving object can be affected by this refinement. Kim et al. used PID control-based tracking and probabilistic morphology refinement step to recover the silhouette. In the approach proposed by Romanoni et al. [282], two histograms are computed: one on the neighborhood of a pixel and another one based on the neighborhood and the intensities history of the same pixel. The Bhattacharyya distance is used with a threshold to detect mov-

ing objects. In 2014, Setyawan et al. [312] classify the pixel as foreground or  
900 background with a two-step method. First, the pixels of the current warped  
image are compared to the background based on their color intensity and a  
threshold is used to label pixels. Then, foreground pixels are tested by compar-  
ing them to their neighborhood in the background model to be sure that they  
belong to moving objects.. In an other work, Minematsu et al. [304] proposed  
905 to find an intensity match between a pixel and another one in a search region.  
This region represents the neighborhood of a pixel where the size of the region  
depends on re-projection errors. Later, the authors proposed to update the  
background model by selecting background pixels based on a similarity measure  
and the re-projection error. Instead of building and maintaining a background  
910 model, Kadim et al. [285] choose to detect moving objects by using successive  
frames. The Wronskian detector [313] is used to detect moving objects between  
the current and the previous frame. The authors also use the neighborhood  
to refine their motion map and they remove false moving blobs by validating  
only those that are detected for at least two successive frames. More recently,  
915 Zhao et al. [291] work with superpixel at different level. A competition between  
background and foreground cues is organized. The result gives the classification  
of the corresponding superpixel. To counteract error alignment accumulations,  
a strong updating strategy is applied on background pixels. In 2019, Yu et al.  
[290] align the two previous frames to the current one and to save computation  
920 time they compute the frame difference on the average on the pixel and its 8-  
neighborhood. To remove shadows from the foreground, the consistency of local  
changes is checked. A consistency points out a shadow area while there is no  
consistency for a moving object. In addition, a lighting influence threshold is  
used to managed illumination changes in the entire frame.

925 Rather than statistical models, Gong et al. [314] and more recently Andreev et  
al. [315] have chosen a codebook model. After the motion compensation with a  
homography, Gong et al. [314] propose compare pixels to codewords on the color  
distorsion and the brightness. The comparison is performed on a small region  
centered on the pixel position in the previous frame to handle warping errors.

930 In a recent work, Andreev [315] chooses the codebook method to represent the  
background with a dictionary of codewords containing pixel intensity informa-  
tion over time. This representation is more robust to background short term  
variations and produces less false detections compare to K-nearest and mixture  
of Gaussians methods. In 2014, Kryjak et al. [301] use the Visual Background  
935 Extractor (ViBE) algorithm [316, 317] to model the background with a modified  
update rule to reduce errors in the classification due to the camera movement.

In the case where the application domain is constrained, the segmentation of  
the scene can be an additional information to moving object detection. Perera et  
al. [318] and Huang et al. [319] both work on aerial images and try to segment  
940 vehicles on roads. Perera et al. [318] choose to use scene understanding to  
segment the image into region and attribute a predefined class, as road or tree,  
to each region. Huang et al. [319] segment images into regions and road regions  
are identified by the size and the straight line property of the region contour. In  
both methods, a region is a moving object according to its position relative to  
945 a road region. Perera et al. also use the scene understanding to remove feature  
points on trees to obtain a better estimation of the homography. As for Huang  
et al. combine the result of image segmentation with the one of frame difference  
to obtain a better foreground segmentation.

Since the camera is moving, some parts of the scene disappear while oth-  
950 ers appear. Parts that disappear do not need special treatment and they are  
just remove when the background is updated. However, new parts have to be  
integrated in the classification and in some methods, they are initialized as back-  
ground [310]. In their method, Lopez-Rubio et al. [288], find the closest labeled  
pixel of a new one. If the new pixel belongs to the background model of its  
955 closest neighbor, then this background model is used as initialization, otherwise  
with a neutral state.

A traditional approach to reduce noise in the binary mask is morphologi-  
cal operation (see 3). This technique can remove small groups of pixels falsely  
labeled as foreground and fill small holes in the foreground segmentation. To  
960 remove noise pixels connected to foreground, Solehah et al. [320] propose to

compare the histogram of the current image with the one of the warped background and threshold it to re-classify the pixels.

Gong et al. [314] use a codebook to model the background. After the motion compensation with a homography, pixels are compared to codewords on the color distortion and the brightness. The comparison is performed on a small region centered on the pixel position in the previous frame to handle warping errors.

#### 5.0.4. Subspace segmentation

In this section, moving objects detection methods use the trajectories of feature points to separate the background and the foreground. Contrary to the previous approaches, there is no registration between images to apply a background subtraction technique. The features points are labeled according to the analysis of their trajectories and the label information is propagated to the whole image to obtain a pixel-wise segmentation.

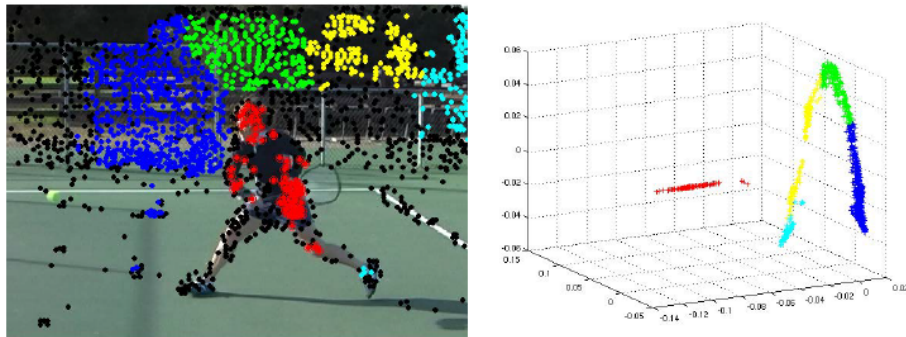


Figure 9: An example of clustering trajectories into a subspace (right) and the result on the image (left).

Source: Images from Elqursh and Elgammal et al. [321].

In 2009, Sheikh et al. [322] use three long term trajectories to construct a 3D subspace. Feature points whose trajectories belong to this subspace are considered as part of the background while the others are foreground. In the proposed method of Elqursh and Elgammal [321] a subspace is constructed with trajec-

References		Main contribution	FB	DM	AM	PM	MF
Murray and Basu	(1994) [283]	Real time motion detection	×	×	×	✓	✓
Odobez and Bouthemy	(1997) [280]	Statistical regularization framework	×	✓	✓	×	×
Paragios and Tziritas	(1999) [308]	Regularization parameter by a voting technique	×	✓	✓	×	×
Ren et al.	(2003) [309]	Spatial distribution of Gaussians	✓	×	✓	×	✓
Michelsoni and Foresti	(2006) [300]	Real time	✓	×	×	×	×
Perera et al.	(2006) [318]	Use scene understanding	✓	×	×	✓	✓
Robinault et al.	(2009) [284]	Real time	✓	✓	×	✓	×
Guillot et al.	(2010) [306]	Feature points matching	✓	×	×	✓	×
Guillot et al.	(2010) [307]	Modified SURF descriptor	✓	×	×	✓	×
Huang et al.	(2010) [319]	Combine frame difference and image segmentation	✓	×	×	✓	×
Wu et al.	(2011) [286]	Spatial-color cue for CRF	✓	×	×	✓	×
Solehah et al.	(2012) [320]	Refine foreground with local histogram processing	✓	×	×	✓	✓
Kadim et al.	(2013) [285]	Avoid blurred images	✓	×	×	✓	×
Kim et al.	(2013) [310]	Spatio-temporal update scheme	✓	×	×	✓	✓
Ferone and Maddalena	(2014) [293]	Self-organizing background subtraction	✓	×	×	✓	×
Ivanov	(2014) [292]	Two background models	✓	×	×	✓	×
Kryjak et al.	(2014) [301]	Modified ViBE	✓	×	×	×	×
Romanoni et al.	(2014) [282]	Temporal + Spatio-Temporal Histograms algorithm	✓	×	×	✓	✓
Setyawan et al.	(2014) [312]	Background subtraction	✓	×	×	✓	×
Wan et al.	(2014) [287]	Two-layer iteration	✓	×	✓	×	×
Lopez-Rubio et al.	(2015) [288]	Two probabilistic models	✓	×	×	✓	×
Minematsu et al.	(2015) [304]	Re-projection error	✓	×	×	✓	×
Tzanidou et al.	(2015) [296]	Telemetry registration	×	✓	✓	×	×
Viswanath et al.	(2015) [311]	Spatio-temporal Gaussian model	✓	×	×	✓	×
Kurnianggoro et al.	(2016) [303]	Using dense optical flow	✓	×	×	✓	×
Kurnianggoro et al.	(2016) [289]	Candidate background model	✓	×	×	✓	✓
Gong et al.	(2017) [314]	Codebook background model	✓	×	×	✓	×
Minematsu et al.	(2017) [305]	Improved updating background models	✓	×	×	✓	✓
Avola et al.	(2018) [294]	Neural background subtraction	✓	×	✓	×	×
Alleboschi et al.	(2019) [302]	Robust to few feature matches	✓	×	×	✓	×
Yu et al.	(2019) [290]	Improve background subtraction	✓	×	×	✓	✓
Zhao et al.	(2019) [291]	Integration of foreground and background cues	✓	×	×	✓	×
Andreev	(2020) [315]	Codebook background model	✓	×	✓	×	×

Table 3: Motion compensation methods summary. **FB**: Feature Based, **DM**: Direct Method, **AM**: Affine Model, **PM**: Projective Model, **MF**: Morphological Filtering.

tory affinities computed on motion and spatial location. The trajectories in the  
980 embedded subspace are then clustered and labeled foreground or background by  
minimizing an energy function which combine multiple cues. The result of this  
segmentation is presented in the figure 9. In an other work, Nonaka et al. [323]  
cluster the trajectories by using three different distances and label the cluster  
based on the shape and the size. To reduce the computation time and the  
985 memory resource, the trajectories from two consecutive frames are used rather  
than long term trajectories. In 2014, Berger and Seversky [324] managed the  
changing number of trajectories over time by a dynamic subspace tracking. At  
each frame, the camera parameters are updated and used to update the shape  
of the trajectories. More recently, Sajid et al. [325] propose to combine mo-  
990 tion and appearance. The motion module performs a low-rank approximation  
of the background dense motion with an iterative method. The probability of  
each pixel belongs to the foreground is estimated from the pixel-wise motion er-  
ror between the background motion approximation and the one observed. The  
appearance module models background and foreground with GMM.

995 In order to obtain a binary mask, the sparse label information is propagated  
to the whole image. The common method to propagate the information is  
to segment the image by constructing a pairwise MRF and minimizing the  
energy generally with the graph-cut algorithm. A pairwise MRF is a graph  
where vertices represent the pixels and the edges connect the vertices with their  
1000 neighborhood as a grid structure over the image. The energy of a MRF is  
composed of two terms: the unary term and the binary term. The unary term is  
used to assign a label to a vertex while the binary term encourages to assign the  
same label to vertices connected by an edge in order to smooth the segmentation.  
A cut is then found in the graph by minimizing the energy to obtain an image  
1005 segmentation.

In 2009, Sheikh et al.[322] use the kernel density estimation method to obtain  
two models, one for the background and one for the foreground. The graph-  
cut algorithm is then used to minimize an energy function on a MRF. In the

method of Elqursh and Elgammal [321] the motion model is propagated to each  
 1010 pixel with a pairwise MRF and estimate the labels with a Bayesian filtering. In  
 an other approach, Nonaka et al. [323] propose to use a case database, which  
 described the foreground with the color and the location, in the segmentation  
 step for the next frame.

References		Main contribution	LTT	FB	DOP	MRF	GCA
Sheikh et al.	(2009) [322]	Three dimensional subspace	✓	✓	×	✓	✓
Elqursh and Elgammal	(2012) [321]	Appearance and motion models	×	✓	×	✓	✓
Nonaka et al.	(2013) [323]	Reduce time computation	×	✓	×	✓	✓
Berger and Seversky	(2014) [324]	Dynamic subspace tracking	✓	✓	×	✓	✓
Sajid et al.	(2019) [325]	Combine motion and appearance	×	×	✓	✓	✓

Table 4: Subspace segmentation methods summary. **LTT**: Long Term Trajectory, **FB**: Feature Based, **DOP**: Dense Optical Flow, **MRF** : Markov Random Field, **GCA**: Graph Cut Algorithm

### 5.0.5. Motion segmentation

1015 The same way as the previous section, the methods presented here uses the  
 trajectories of the feature points to segment each frame of the video as static or  
 moving but without using a subspace (see figure 10). Those methods are inspired  
 by the methods called Motion Segmentation in the literature which segment  
 the image according to the apparent motions. Here the methods presented  
 1020 go further than just segment each frame of the video by the 2D motions by  
 proposing a background/foreground labeling.

In 2013, Narayana et al. [327] segment images based on optical flow ori-  
 entations. In the case where the camera performs only translation, the depth  
 variations do not take part in the optical flow orientations. In an other approach,  
 1025 Yin et al. [328] cluster feature points according to their trajectory similarity  
 and reject false trajectories by using the PCA algorithm. In an other work,  
 Bideau et al. [329] use the translational flow obtained by the subtraction of the  
 dense optical flow and the rotational flow. The angle field is then estimated

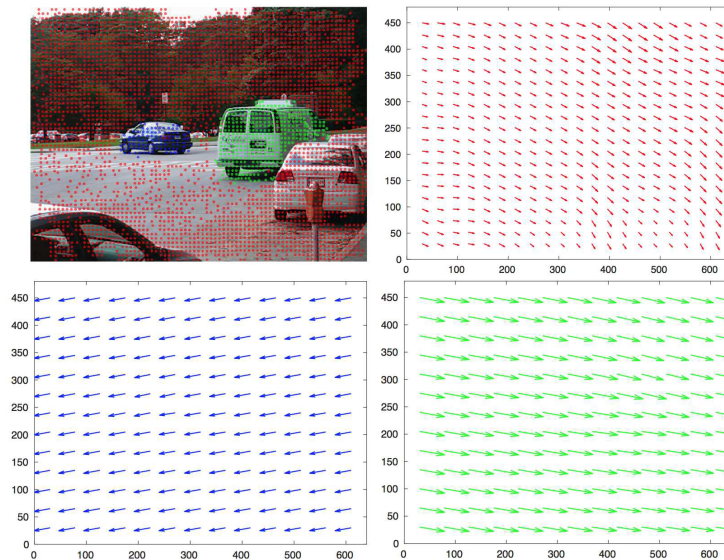


Figure 10: An example motion segmentation on the top left image. The three other images represent the optical flow of the three motions observed in the image.

*Source:* Images from Zhu and Elgammal [326].

from the translational flow according to the magnitude which indicates the re-  
1030 liability of the flow angle. Then the conditional flow angle likelihood estimate  
the probability that the flow direction of a pixel corresponds to the one esti-  
mated. Finally, the Bayes' rule is used to obtain the posterior probability for  
each pixel which is used for the final segmentation. The authors also proposed  
to segment the first frame of the video by choosing three superpixels with a mod-  
1035 ified RANSAC algorithm in order to estimate the motion of the background.  
In an other approach, Kao et al. [330] recover the 3D motions from the 2D  
motions observed by using motion vanishing point and the estimated depth of  
the scene. The final segmentation is applied on the 3D motions. The method  
proposed by Zhu and Elgammal [326] first clusters trajectories based on their  
1040 affinities and propagate the label of trajectories dynamically. The clusters auto-  
matically adapt to the number of foreground object in the frames by computing  
intra-cluster variation. In a recent work, Sugimura et al. [331] use the OneCut  
algorithm to segment frames. Rather than manually select seeds by hand for

the OneCut segmentation, the authors propose to find automatically the seeds  
1045 by using motion boundaries computed by the Canny detector on the magni-  
tude and direction flow fields. Foreground seeds are selected inside enclosed  
motion boundaries while background seeds are selected on rectangles that en-  
close motions boundaries. Recently, Huang et al. [332] estimate a dense optical  
flow by using FlowNet2.0 [333] an optical flow estimation algorithm with deep  
1050 networks. The background optical flow is estimated by a quadratic transforma-  
tion function with the Constrained RANSAC Algorithm (CRA). The CRA is a  
modified version of the RANSAC algorithm to avoid overfitting and improving  
the searching efficiency. In 2020, Zhang et al. [334] reconstruct the background  
orientation by solving the Poisson equation with the five-point interpolation  
1055 method. The reconstruction is then compared to the original orientation field  
to obtain moving objects.

As in the previous section 5.0.4, sparse labeling information is propagated  
to the whole image to obtain a dense labeling.

In 2013, Narayana et al. [327] compute color likelihoods with KDE by sup-  
1060 posing that the region, resulting from a dense segmentation, with the largest  
number of pixels belongs to background. The authors proposed to combined  
the color likelihoods to the previous labeling, compensated according to the  
optical flow, to reduce errors in the segmentation based on optical flow orien-  
tations. In order to propagate the label information, Yin et al. [328] propose  
1065 a trajectory-controlled watershed segmentation algorithm. After applying a bi-  
lateral filtering to smooth the image and enhance the edges, gradient minima  
and the trajectory points are selected as markers. Those markers are used by  
the watershed algorithm as seeds to obtain a segmentation for which the regions  
are labeled background or foreground according to the labels of the trajectories.  
1070 Finally, the background/foreground information is propagated to the unlabeled  
regions by minimizing an energy function on a MFR with the graph-cut algo-  
rithm. The Multi-Layer Background Subtraction (MLBS) proposed by Zhu et  
al. [326] propose a multi-label segmentation rather than a binary segmentation.

Each motion cluster is associated to a layer. For each layer, a pixel-wise motion  
1075 estimation is performed by a Gaussian Belief Propagation (GaBP). Then the  
appearance model and the prior probability map are updated with the motion  
estimation and they are used to compute the posterior probability map. The  
multi-label segmentation is performed on the posterior probability map by the  
minimization of the energy of a pairwise MRF. In an recent work, Sugimura  
1080 et al. [331] prevent unreliable magnitude and direction foreground flow field by  
introducing a prediction based on the lasts foreground estimated regions. In the  
case where the magnitude and direction foreground are unreliable, the predic-  
tion is used rather than the two flow fields as the segmentation result otherwise  
the prediction is jointly used with the two others flow fields. The OneCut is  
1085 applied a second time with the appearance information in order to improve the  
final segmentation. In an other work, Kao et al. [330] obtain a binary mask  
by segmenting the 3D motions with three different clustering methods: simple  
k-means clustering, spectral clustering with a 4-connected graph and with a  
fully connected graph. In 2019, Huang et al. [332] propose a dual judgment  
1090 mechanism to separate the foreground from the background. The foreground  
is estimated by thresholding the difference of the estimated background opti-  
cal flow and the one estimated by FlowNet2.0. In order to take into account  
the case where the camera is zooming, a second judge mechanism is based on  
thresholding the difference of cosine angles. Recently, Zhang et al. [334] pro-  
1095 pose to enhance a motion saliency map in order to segment moving objects that  
are indistinguishable in the dense optical flow. The saliency map accumulate a  
weighted similarity on the motion and the appearance. The temporality is used  
to remove false positive alarms.

In 2017, Tokmakov et al. [335] propose to combine results from appearance  
1100 and motion network into a memory module. This combination prevents the clas-  
sification of a temporary stopped moving object as background since it cannot  
be detected as moving with the motion information unlike the appearance one.  
Moreover, the authors propose a bidirectional processing in order to overcome  
the *Waking foreground object* challenge by using to parallel process, one in the

1105 forward direction and one in the backward direction. The same year, Jain et al. [336] also propose a combination of appearance and motion network. The authors suggest to train the appearance network in order to learn a generic foreground appearance. According to the authors, the network is able to identify foreground regions from more than 3,000 object categories. The training of the  
 1110 appearance network requires a lot of annotated frames but those data does not exist currently. To overcome the problem, the authors propose a training based on weakly annotated video, i.e. video annotated with bounding boxes.

References	Main contribution	LTT	FB	DOP
Narayana et al. (2013) [327]	Segmentation based on optical flow orientation	×	×	✓
Yin et al. (2015) [328]	Trajectory-controlled watershed segmentation	✓	✓	×
Bideau et al. (2016) [329]	Combine angle and magnitude	×	×	✓
Kao et al. (2016) [330]	3D motions segmentation	×	×	✓
Jain et al. (2017) [336]	Appearance and motion network	×	×	✓
Tokmakov et al. (2017) [335]	Appearance and motion network	×	×	✓
Zhu and Elgammal (2017) [326]	Multi-label background subtraction	×	✓	×
Sugimura et al. (2018) [331]	Automatic OneCut method	×	✓	✓
Huang et al. (2019) [332]	Dual judgment mechanism	×	×	✓
Zhang et al. (2020) [334]	Background orientation reconstruction	×	×	✓

Table 5: Motion segmentation methods summary. **LTT**: Long Term Trajectory, **FB**: Feature Based, **DOP**: Dense Optical Flow

### 5.0.6. Plane+Parallax

The Plane+Parallax decomposition is a scene-centered representation [337].  
 1115 As in the previous section, this technique firstly compensates the camera motion with a 2D parametric transformation that describes the dominant plane in the scene. After the registration process, camera rotation and zoom are eliminated and misaligned pixels correspond either to the parallax caused by the camera translation or to a moving object. Then, residual displacements belong to the  
 1120 scene form a radial field centered at the epipole [338].

In 1998, Irani and Anandan [339] stratify the moving object detection problem and propose a method that handles from 2D scenes up to 3D complex

scenes. The first level of the stratification is the approximation of the scene by a 2D plane. A single 2D parametric transformation is estimated between two images and used to warp them. Misalignments correspond to moving objects. The second level handle misalignments due to the parallax. Several 2D planes are estimated successively with the same method in the previous level and regions which are inconsistent with the motion of any 2D planes are moving objects. When the scene is complex, with many small moving objects are different depths, the two previous methods cannot correctly make the detection. In this case, the third level with a Plane+Parallax scene representation is used. The authors noticed that the residual movements after the registration are due to the translation motion of the camera and they form a radial field centered at the Field Of Expansion (FOE). The estimation of the FOE can be used to apply the *Epipolar Constraint* but the estimation can be biased by moving objects as shown in the figure 11. To avoid this, the authors proposed a *Parallax-Based Rigidity Constraint* which is a consistency measure between two points over three consecutive frames. One of the two point is known static in order to evaluate the label of the second point. In an other work, Sawhney et al. [340] impose the *Shape Constancy* and the epipolar constraint over several frames to estimate a robust image alignment. The authors used the Plane+Parallax decomposition to enforce the two constraints.

In 2005, Kang et al. [341] use the consistency constraint. The advantages of this constraint are: the reference plane does not need to be the same. It could be the floor and then a wall for example. Static points are not necessary and the assumption of small camera displacement between two consecutive frames are not required. The authors combined the epipolar constraint and a structure consistency constraint to eliminate false detections due to the parallax. From the epipolar constraint an angular difference map is created and from the structure consistency constraint a depth variation map is created for each residual pixel. Rather than propose a binary mask, a likelihood map is computed on a sliding window and used directly by a tracking algorithm.

There exist one particular case where the Plane+Parallax methods do not

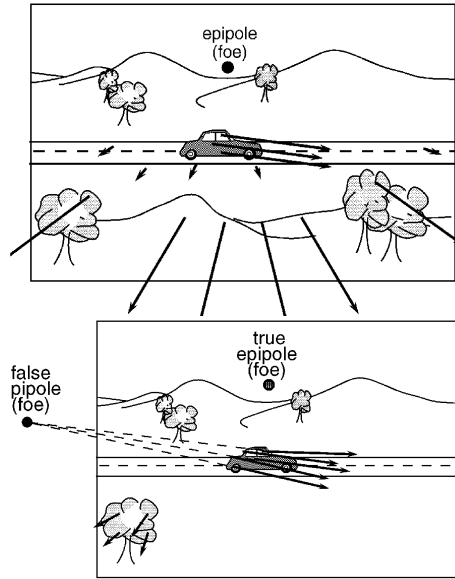


Figure 11: An illustration.

Source: Images from Irani and Anandan [339].

work: when the camera and an object both move in the same direction with  
 1155 constant velocities. The constraints defined to distinguish the parallax and a  
 moving object are verified and the object is labeled static.

References	Main contribution	FB	DM	AM	PM
Irani and Anandan (1998) [339]	Handle 2D and 3D scenes	×	✓	×	✓
Sawhney et al. (1999) [340]	Shape constancy and epipolar constraint	✓	✓	×	✓
Kang et al. (2005) [341]	Structure consistency and angular map	✓	×	×	✓

Table 6: Plane+Parallax methods summary. **FB**: Feature Based, **DM**: Direct Method, **AM**: Affine Model, **PM**: Projective Model

### 5.0.7. Multi planes

Multi planes scene representation was firstly used in motion segmentation [342, 343, 344].

1160 Contrary to Motion Compensation method where only one image alignment is computed, several alignments are estimated in the case of multi-layers ap-

proaches. Cascade of RANSAC is a very used technique to estimate several  
real planes in a scene [345, 346, 347, 348, 349, 350, 351]. Here is the general  
principle: RANSAC is used on feature points to estimate one 2D transformation  
1165 between two images in the video sequence. Feature points that fit the homogra-  
phy are removed from the process and a new transformation is estimated with  
the residual feature points. This process is repeated until a condition is reached.

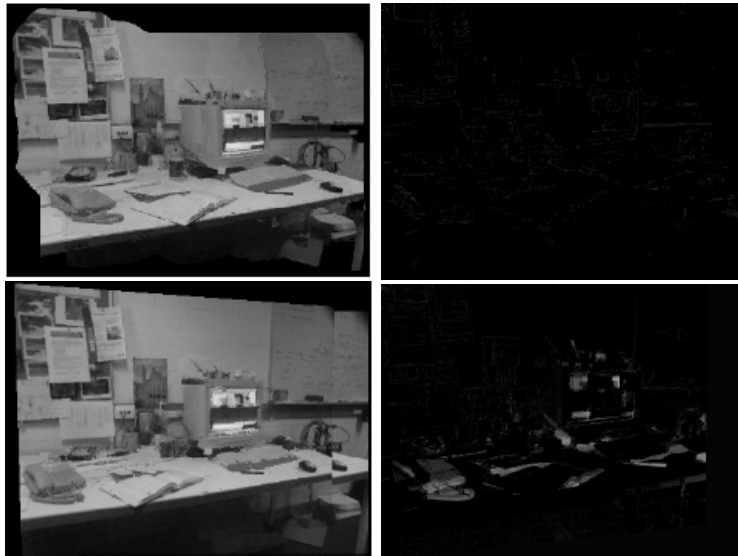


Figure 12: An example of image registration with several planes (top line) compared with  
image registration with one plane (bottom line). The left column represents the rectified  
frame after the compensation and the right column represents the disparity.

*Source:* Images from Jin et al. [345].

In 2008, Patwardhan et al. [346] use a training step to automatically ini-  
tialize the number of layers. Layers are estimated iteratively on color of pixels  
1170 by Sampling-Expectation refining process. The method of Zhang et al. [347]  
propose to adaptively adjust the parameters of RANSAC to handle simple and  
complex classes of scenes. Feature points are hierarchically clustered based on  
Euclidean distance criterion on optical flow data. A cluster is labeled as back-  
ground if it has a widespread spatial distribution. Then, the number of layers  
1175 is estimated iteratively by increasing the number layers until a consensus is

reached. In an other work, Zamalieva et al. [349] modify the GRIC score to find out if the scene can be approximated by one plane or by several planes. The modified GRIC score is computed on one homography or on the fundamental matrix. If the homography wins, it is chosen to compensate the camera motion.

1180 On the other case, a cascade of RANSAC is used to compute several homographies. In the approach of Hu et al. [350] feature points are first classified as background or foreground and use them to compensate the camera motion by a homography. The authors use one plane for the frame compensation but they approximate the scene by several planes during the feature points classification

1185 by computing the fundamental matrix and using the epipolar constraint. In a similar manner, Zamalieva et al. [348] propose to approximate the scene with the epipolar geometry to avoid alignment error due to the parallax. The authors propose a modified Temporal Fundamental Matrix (TFM) which can be used with a single camera to represent the epipolar geometry. In an other approach,

1190 Kim et al. [352] estimate several homographies by clustering trajectories into the Distance and Motion Coordinate (DMC) system. From the biggest clusters, two regression lines are derived and used to find the preliminary background clusters. Homographies are estimated with the RANSAC algorithm from those background trajectories after another clustering step. Rather than find real

1195 planes in the scene, Zamalieva et al. [353] propose to create parallel hypothetical planes based on the dominant plane in the scene. These planes are estimated with the vanishing line and the vertical vanishing point. The image registration is computed by homographies estimated for each hypothetical plane.

When several homographies are used to register the background, it is necessary to find which homography have to be applied for each pixel. In both work of

1200 Jin et al. [345] and Zamalieva et al. [349] pixel intensity similarity is computed for each homography to select a plane for the candidate pixel. In 2008, Jin et al. [345] assign non-overlap pixels to layers with *Minimal Span Tree* to represent scene smoothness. In 2014, Zamalieva et al. [349] handle occluded background

1205 pixels by performing a majority voting on neighbor pixels associated to a plane.

Foreground detection step is very close to those used for static camera thanks to the image registration step [345, 347]. Jin et al. [345] use mixture of Gaussians and a background panorama to detect moving objects while Zhang et al. [347] simply assign a pixel to background based on intensity difference thresholding. In an other work, Patwardhan et al. [346] assign pixels to one layer in the training stack or identifies them as foreground. Spatio-temporal subvolume identify candidate layers and non-parametric KDE is used to estimate the probability that the current pixel belongs to each candidate layers. **Zamalieva et al. [348] use the estimated epipolar geometry in a MAP-MRF framework for the background/foreground labeling step. Pixels are grouped by color coherency to reduce the size of the graph in the graph-cut algorithm for the energy minimization.** In a recent work, Zhou et al. [351] detect regions that became visible by motion parallax and produce false alarms. The authors combine these regions information with a codebook-based background segmentation.

References	Main contribution	RP	IP	CR	EG
Jin et al. (2008) [345]	Cascade of RANSAC	✓	×	✓	×
Patwardhan et al. (2008) [346]	Training stack of layers	✓	×	×	×
Zhang et al. (2012) [347]	Multi-classes RANSAC	✓	×	✓	×
<b>Zamalieva et al. (2014) [348]</b>	<b>Temporal fundamental matrix</b>	<b>✓</b>	<b>×</b>	<b>×</b>	<b>✓</b>
Zamalieva et al. (2014) [349]	Adaptive motion compensation	✓	×	✓	✓
Zamalieva et al. (2014) [353]	Stack of hypothetical 3D planes	×	✓	×	✓
Hu et al. (2015) [350]	Epipolar geometry	✓	×	×	✓
Kim et al. (2016) [352]	Distance and Motion Coordinate system	✓	×	×	×
Zhou et al. (2017) [351]	Regions revealed by motion parallax	✓	×	✓	×

Table 7: Multi layers methods summary. **RP**: Real Planes, **IP**: Imaginary Planes, **CR**: Cascade of RANSAC, **EG**: Epipolar Geometry.

#### 1220 5.0.8. Split image in blocks

In the literature, one identifies two ways to divide an image into blocks. The first one simply divides the image into a regular grid where each block has a predefined size. The second technique uses superpixel segmentation methods. Each block represents a region in the image whose features depend on the

1225 segmentation method.

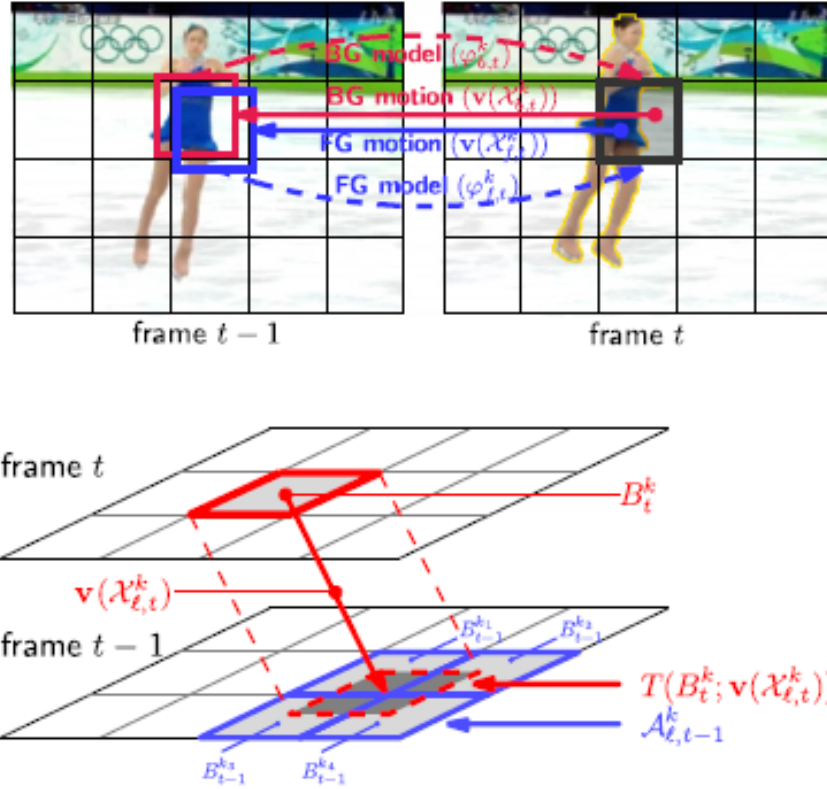


Figure 13: An example of a technique to divide the image into a regular grid and compensate the motion by blocks.

Source: Images from Lim et al. [354].

In some methods [355, 356, 357, 358, 359] the motion compensation is estimated on the whole image as in the section 5.0.3 but others compensate the camera motion by blocks. Rather than compute one homography for the whole image, one homography for each grid cell could be computed to register images [360, 361]. Some authors propose to estimate two types of motion for each block: one for the background and one for the foreground [362, 354, 363]. In 2011, Kwak et al. [362] choose to use non-parametric Belief Propagation to reduce the noise in optical flow and recover the missing background motion. To

estimate background and foreground motion, Lim et al. [354] simply use sparse  
1235 optical flow. In an other method, Kim et al. [364] propose a multi-resolution  
motion propagation to compensate the camera motion on blocks. If a block  
does not have background feature points to estimate its transformation, the  
parameters are propagated from the blocks at a higher level. In the method  
of Lim and Han [363], the previous segmentation mask is warped with dense  
1240 optical flow and use the warped mask to compute dense motion for background  
and foreground independently. In 2016, Sun et al. [360] compute two kinds of  
motion. The first one is computed over a regular grid with the *As Similar As  
Possible* method. The motion of the whole image is a set of homography. The  
second motion is computed over superpixels with the KLT technique. These  
1245 two motions are then used to obtain a background/foreground segmentation  
from motions. Recently, Ben Khalifa et al. [365] propose a pedestrian detection  
with a camera located on top of a vehicle’s windshield based on moving objects  
detection. The camera motion is estimated by matching blocks between two  
consecutive frame located in a ROI. This ROI is extracted on the top of the  
1250 images because this portion should not contain foreground object according to  
the camera location.

In 2008 Unger et al. [366] propose to use motion models to generate an artificial  
background image for each frame of the video and use it into a background  
subtraction process. The artificial background is constructed by a weighted  
1255 accumulation of pixels of the neighboring frames. The weights are computed  
according a block-wise motion deviation and its the Rayleigh distribution modeling  
in order to accumulate background pixels only.

In 2018, Kazantzidis et al. [367] propose a method based on vide-omics, a  
genomics inspired paradigm. The frame is not divided into square blocks but by  
1260 line. The scanline alignment is based on the analogy between protein sequence  
alignment and image pair correspondence. Pixels are matched according to a  
scoring function and gap are inserted to model mutations corresponding to pixel  
insertions, substitutions and deletions.

The blocks are also used to model the scene. Rather than model each pixel  
1265 in the image, each block is represented by one model which reduce computation  
time.

In 2013, Yi et al. [355] choose to model each block with a Single Gaussian  
Model (SGM). After motion compensation, one block generally overlap several  
blocks in the previous frame. In order to update block models in the current  
1270 frame, the overlap block models are mixed together where each block is weighted  
proportionally to the overlapping area. The same mixing blocks is used by Lim  
et al. [354] for their temporal model propagation step and they additionally  
use a spatial step to enforce the spatial coherence. The methods of Kwak et  
al. [362] and Lim and Han [363] also combine motion and appearance models.  
1275 In 2015, Yun and Choi [356] propose to improve the method of Yi et al. [355]  
with a selectively update step based on a sampling map. Only some pixels  
are chosen according to temporal and spatial properties to update the model.  
In a further work, Chung et al. [357] regulate the background model of Yi et  
al. [355] by including foreground cues coming from frame differencing. In an  
1280 other work, Zheng et al. [358] propose to use two background models: local  
and global. The first one is used to obtain a first pixel classification after a  
motion compensation based on optical flow. This rough detection is used to  
construct a superpixel-level background regions which are used to estimate the  
GMM background appearance. This second background modeling, called the  
1285 global background, is used to reduce noise and false detection by using the GMM  
background on pixels classified as foreground with the local background.

Once the models are updated, the data are combined together to create the  
final segmentation mask for the current frame.

In 2011, Kwak et al. [362] predict the appearance model of each block by  
1290 a weighted sum of Gaussian-blurred blocks of the previous frame. In order to  
reduce segmentation errors, some methods [362, 354, 363, 357] propose to iter-  
ate the process on motion and appearance models until the models converge.  
The method of Lim et al. [354] and the one of Lim and Han [363] both iterate

on motion and appearance estimations to obtain a segmentation mask at each  
1295 frame. In their approach, Lim and Han [363] choose to use superpixel rather  
than a grid because this kind of pixel groups has color and motion consistency.  
In an other work, Yi et al.[355] use two background models with ages to reduce  
foreground and noise contamination. Models are swapped when the candidate  
model is older than the current model and the new candidate model is initialized  
1300 to remove contaminations. In 2017, Makino et al. [368] use the method of Yi  
et al. [355] as a baseline to compute an anomaly score map. The authors also  
compute a motion score map based on optical flow angles after motion compen-  
sation. The two score maps are merged in the moving object detection step. In  
order to manage slow moving objects, Yun et al. [359] update the SGM block-  
1305 based model of Yi et al. [355] according to the foreground velocity. In the case  
where the foreground moves less than a block size during several frames, the  
SGM mean is updated with the illumination change and the average intensity  
of the block. The SGM variance is increased according to the current block  
intensity and the mean the previous and current time. The authors also reduce  
1310 false positives by combining threshold labeling and watershed segmentation. In  
an other work, Kim et al. [364] combine sparse optical flow clustering with  
the Delaunay triangulation method in order to complete the missing detection  
information of the Frame Differencing method. The optical flow clustering is  
computed on blocks with the K-means method. In an other approach, Sun et al.  
1315 [360] create two segmentations, one from motion and one from appearance. The  
motion one is created from the difference between the camera motion estima-  
tion and the superpixels motion estimation. Identical motions on superpixels  
come from the background and they are used as seeds for a region growing  
propagation. The appearance segmentation is based on color and Local Binary  
1320 Similarity Patterns (LBSP). The two segmentations are then combined with  
MRF and the final segmentation is obtained by graph-cut. In 2017, Wu et al.  
[361] use a coarse-to-fine method to detect foreground objects. Each block of the  
regular grid is warped according to its dominant motion over a sliding window.  
The Mean Squared Error (MSE) is then used as a threshold to obtain a coarse

1325 foreground region. The motion of the coarse foreground region is decomposed  
into background and foreground motions thanks to inpainting method. The fine  
foreground is obtained by an adaptive thresholding method. After compensat-  
ing the camera motion by a Hierarchical Block-Matching algorithm, Szolgay et  
al. [369] build a Modified Error Image (MEI) from the result of the frame dif-  
1330 ference. A spatio-temporal background Probability Density Function (PDF) for  
each pixel of the MEI is computed with the Kernel Density Estimation (KDE).  
Pixels are then labeled as background or foreground according to the PDFs and  
pixels are finally clustered with their motion, color and location. In a recent  
work, Komagal and Yogameena [370] propose a combination of Region-based  
1335 Mixture of Gaussian (RMoG) and the Extended Center Symmetric Local Bi-  
nary Pattern (XCS-LBP) for moving objects detection with a PTZ camera. The  
RMoG algorithm is a block-based method which reduce noise from the camera  
motion and illumination changes. Rather than work at pixel level, the RMoG  
use region of pixel and its neighborhood to describe pixel intensity information.  
1340 To reduce false positives, the pixel classification obtained with the RMoG algo-  
rithm is used with the XCS-LBP. Finally a morphological operator is used to  
obtain a better classification. Recently, Ben Khalifa et al. [365] propose a block  
classification by comparing a global motion estimated on blocks that belongs to  
the background to the motion estimated for each individual block. The result  
1345 is then used to identify a detected object as pedestrian or non-pedestrian with  
a Support Vector Machine (SVM) classifier.

In 2018, Minematsu et al. [371] propose a method based on patch reconstruc-  
tion which does not require motion compensation. A first network is trained to  
reconstruct the masked center part of a patch by using the surrounding region.  
1350 The moving objects detection is realized by a second network trained to com-  
pare the reconstructed patch to the current one. This network is also able to  
classify shadows as background thanks to the supervised learning.

The proposed method based on genomics paradigm by Kazantzidis et al.  
[367] differentiate unmatched pixels regions on scalines into foreground and  
1355 background. These unmatched regions come from moving objects, occluded

or non-overlapping areas. To obtain a robust classification, unmatched region  
a foreground likelihood computed as the mean of scanline comparisons over  
several frames. Thus, region with few matches are foreground candidates. A  
final classification refinement is applied with morphological operators and the  
1360 GrowCut [372] algorithm.

## 6. Datasets and evaluation

This section introduces the publicly available datasets and the quantitative  
evaluation metrics that be used on these datasets to measure the performance  
of a method and compare them. An overview of f-score results is also provided.

### 1365 6.1. Existing datasets

In order to test the performance of a moving object detection method with  
a moving camera, it is necessary to have video sequences whose each pixel of  
each frame are annotated. This section presents the datasets that can be used  
to evaluate and compare methods. In the same way one writes this paper, only  
1370 datasets that contain videos taken by a moving camera are presented.

- The **Hopkins 155+16**<sup>1</sup> dataset was firstly introduced by Tron and Vidal  
[373] and known as the Hopkins 155 dataset. This dataset was originally  
created to evaluate motion segmentation algorithms but the data can also  
be used for moving objects detection algorithms. There are 57 different  
1375 videos, mostly taken by a moving camera and 114 sequences derived from  
these videos. The derived sequences differ from the original ones by their  
ground truth which represent a subset of motions in the video. For each  
sequence, complete trajectories of feature points and ground truth on the  
points are provided. For the 16 additional sequences, the trajectories  
1380 contain missing data and outliers. Moving objects are chessboards in two-  
thirds of sequences and the last third contains cars and people.

---

<sup>1</sup><http://www.vision.jhu.edu/data/hopkins155>

References	Main contribution	RG	S	MCI	MCB	BFM	MAM	IM
Unger et al. (2008) [366]	Artificial background image	✓	×	×	✓	×	×	×
Kwak et al. (2011) [362]	Hybrid inference motion/appearance	✓	×	×	✓	✓	✓	✓
Szolgay et al. (2011) [369]	Modified Error Image	✓	×	×	✓	×	×	×
Lim et al. (2012) [354]	Combine spatial/temporal models	✓	×	×	✓	✓	✓	✓
Kim et al. (2013) [364]	Optical flow clustering and triangulation	✓	×	×	✓	×	×	×
Yi et al. (2013) [355]	Dual background model	✓	×	✓	×	×	×	×
Lim and Han (2014) [363]	Superpixel segmentation	×	✓	×	✓	✓	✓	✓
Yun and Choi (2015) [356]	Selectively update	✓	×	✓	×	✓	×	×
Chung et al. (2016) [357]	Reduce background model errors	✓	×	✓	×	✓	✓	✓
Sun et al. (2016) [360]	Motion/appearance segmentations	✓	✓	×	✓	×	×	×
Zheng et al. (2016) [358]	Local-to-global background modeling	×	✓	✓	×	×	×	×
Makino et al. (2017) [368]	Score maps	✓	×	✓	×	×	✓	×
Wu et al. (2017) [361]	Coarse to fine strategy	✓	×	×	✓	×	✓	×
Yun et al. (2017) [359]	Slow moving objects	✓	×	✓	×	×	×	×
Kazantzidis et al. (2018) [367]	Genomics inspired paradigm	×	×	×	✓	×	×	×
Komagal and Yogameena (2018) [370]	RMoG and XCS-LBP combination	✓	×	×	✓	×	×	×
Minematsu et al. (2018) [371]	Patch reconstruction	✓	×	×	×	×	×	×
Ben Khalifa et al. (2020) [365]	Block motion segmentation	✓	×	✓	×	×	×	×

Table 8: Split image in blocks methods summary. **RG**: Regular Grid, **S**: Superpixels, **MCI**: Motion Compensation on Image, **MCB**: Motion Compensation on Blocks, **BFM**: Background and Foreground Models, **MAM**: Motion and Appearance Models, **IM**: Iterative Method.

- 1385
1390
 • **FBMS-59**<sup>2</sup> dataset proposed by Ochs et al. [374] (Freiburg-Berkeley Motion Segmentation dataset) is an extension of the **BMS-26** dataset of Brox and Malik [375] (Berkeley Motion Segmentation dataset). The BMS-26 consists of 26 sequences, where 12 sequences come from the Hopkins 155 dataset, taken by a moving camera where most video sequences present high camera movements. Brox and Malik provided ground truth masks on some frames of the BMS-26 dataset, accumulating a total of 189 frames annotated. Annotations are masks where each moving object is pixel-accurate identified by a grayscale value. The FBMS-59 dataset extended the BMS-26 dataset with 33 additional video sequences with a total of 720 frames annotated. This dataset is decomposed into training and test sets. The masks provided can be easily used to evaluate moving objects detection algorithms.
- 1395
1400
 • **ChangeDetection.net**<sup>3</sup> called CDnet. There exist two versions of this dataset: CDnet 2012 [376] and CDnet 2014 [377]. Almost all sequences are taken by a static camera but in CDnet 2014, four sequences are taken by a PTZ camera. For each sequence, a ground truth mask is provided. The mask contains five labels: static, hard shadow, outside region of interest, unknown motion (usually around moving objects, due to semi-transparency and motion blur) and motion. Each label is associated to a gray color and a simple filter can be used on this mask to obtain a binary mask which can be used to evaluate a method.
- 1405
 • The Densely Annotated Video Segmentation **DAVIS**<sup>4</sup> was proposed by Perazzi et al. [378]. Three versions of the dataset were proposed: [378], [379], [380]. The first version [378] contains 50 different videos where only 5 videos were taking by a static or a shaking camera. For each video, a binary ground truth mask is given for each frame. In the two other

---

<sup>2</sup><https://lmb.informatik.uni-freiburg.de/resources/datasets/moseg.en.html>

<sup>3</sup><http://changedetection.net>

<sup>4</sup><https://davischallenge.org>

versions of the dataset [379] and [380], 40 videos were added. Among  
1410 those 90 video sequences, only 10 were taken by a static or a shaking  
camera. In the same manner than for the first dataset version, for each  
frame of a video, a mask is given. The mask is not a binary mask but  
moving objects are classified into categories like human or bike according  
to colors. The background is still identified by the black color and it can  
1415 be used to differentiate background from foreground.

- The **SegTrack v2**<sup>5</sup> dataset [381] is an updated version of the SegTrack  
dataset [382]. This second version propose 14 sequences where 10 se-  
quences are taken by a moving camera. A binary ground truth mask is  
provided for each frame of each sequence and the ground truth of each  
1420 moving object is provided separately.
- The **ComplexBackground**<sup>6</sup> is a dataset proposed by Narayana et al.  
[327] and contains five video sequences taken by a hand-held camera. Each  
video contains 30 frames and 7 frames are used for the ground truth as a  
binary mask. These videos contain one or several moving objects and the  
1425 static scene presents significant depth variations.

## 6.2. Evaluation metrics

Thanks to the publicly available datasets and their associated ground truth,  
quantitative metrics are used to evaluate the performance of background/foreground  
segmentation approaches and compare them together.

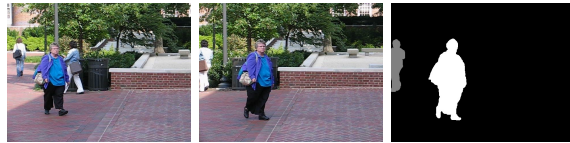
1430 According to the ground truth, the pixels are categorized into one of these  
four categories:

- True Positive (TP): the number of pixels correctly labeled as foreground.  
Also known as *hit*.

---

<sup>5</sup><https://web.engr.oregonstate.edu/~lif/SegTrack2/dataset.html>

<sup>6</sup><http://vis-www.cs.umass.edu/motionSegmentation/complexBgVideos.html>



(a) Hopkins 155+16 dataset, sequence people2



(b) FBMS-59, sequence giraffes01



(c) ChangeDetection.net, sequence continuousPan



(d) DAVIS, sequence bmx-trees



(e) ComplexBackground, sequence forest

Figure 14: Illustrations of datasets with input images and their ground truth. The two first columns are images taken by a moving camera and the third column is the ground truth of images from the second column.

1435

- True Negative (TN): the number of pixel correctly labeled as background. Also known as *correct rejection*.
- False Positive (FP): the number of pixel incorrectly labeled as foreground. Also known as *false alarm* or *Type I error*.
- False Negative (FN): the number of pixel incorrectly labeled as background. Also known as *miss* or *Type II error*.



Figure 15: An example of background/foreground segmentation on the people01 sequence the from the Hopkins dataset. Left: the original image. Center: the ground truth. Right: an example of background/foreground segmentation where green pixels are labeled as background and red pixels are labeled as foreground.

1440

Three measures are commonly used to evaluate background subtraction algorithms: the precision, the recall and the F-score.

- The *precision* (also known as *positive predictive value*) is the proportion of pixels that are correctly detected as moving among all pixels detected as moving by the algorithm.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

1445

- The *recall* (also known as *sensitivity*, *hit rate* or *true positive rate*) is the proportion of pixels that are correctly detected as moving among all pixels that belong to moving objects in the ground truth.

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

- The *F-score* (also known as *F1 score* or *F-measure*) is the combination of precision and recall. It is the harmonic mean of precision and recall

1450

measures:

$$F - score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (3)$$

Several other measure metrics are also used:

- The *Accuracy* is the proportion of pixels detected as moving among all the labeled pixels.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

1455

- The *Specificity* (also known as *selectively* or *true negative rate*) is the proportion of pixels that are correctly detected as static among all pixels that belong to static objects in the ground truth.

$$Specificity = TN / (TN + FP) \quad (5)$$

- The *false positive rate* (also known as *fall-out*) is the proportion of pixels that are incorrectly detected as moving among all pixels that belong to static objects in the ground truth.

$$FalsePositiveRate : FP / (FP + TN) \quad (6)$$

1460

- The *false negative rate* (also known as *miss rate*) is the proportion of pixels that are incorrectly static as moving among all pixels that belong to moving objects in the ground truth.

$$FalseNegativeRate : FN / (TP + FN) \quad (7)$$

### 6.3. Quantitative results

The tables 9 and 10 present F-score obtained on five sequences from the Hopkins 155 dataset for different methods presented in this survey. Those sequences are among the most used for evaluation and the values are reported from the original papers. Here we have chosen to present the F-score, since its a popular metric, but it worth to note that the evaluations are not computed

on the same data for each method. For example, Minematsu et al. [304] have  
1470 manually constructed the ground truth for all frames while others as Elqursh  
and Elgammal have chosen to use the ground truth provided in the FBMS-59  
dataset by Ochs et al. [374] on a set of frames.

The five sequences from the Hopkins 155 dataset are short sequences (be-  
tween 19 and 54 frames), with one or two moving objects progressing in a simple  
1475 scene and with a smooth camera movement. Those sequences are part of simple  
cases where three challenges are present: moving background, small motion  
parallax (quite small for these sequences) and the shadows.

Among all methods reported in the tables 9 and 10, only two methods have  
a f-score exceeded 0.9 for the five sequences, Berger and Seversky [324] and  
1480 Sajid et al. [325], both in the subspace segmentation category. Most f-score are  
between 0.7 and 0.9 and no category really stands out from the others. Those  
results show that the moving objects detection with a moving camera is still a  
challenging problem even for simple sequences.

## 7. Discussion

### 1485 7.1. Solved and unsolved challenges

Among all challenges presented in the section 3, two concerns more specifi-  
cally this survey: *Motion parallax* and *Moving background*. The other challenges  
are generally overcome by using or adapting solutions which come from methods  
with a static camera. More information about those challenges are provided by  
1490 Garcia-Garcia et al. [1].

The moving objects detection methods presented in this survey are divided  
into two big categories: one plane and several planes. This categorization re-  
flects the level of motion parallax that the methods can handle. The approaches  
that approximate the scene by one plane suppose a zero or a near-zero paral-  
1495 lax while the parallax is present in those that use several planes for the scene  
approximation. The distribution of the methods between the two categories is

References		people1	people2	cars1	cars4	cars6
<b>Panoramic background subtraction</b>						
Avola et al.	(2017) [253]	0.840	-	-	-	0.840
<b>Motion compensation</b>						
Ferone and Maddalena	(2014) [293]	0.940	0.950	0.862	-	0.913
Minematsu et al.	(2015) [304]	0.875	0.896	-	0.822	0.802
Minematsu et al. (1)	(2017) [305]	0.873	0.890	0.833	0.789	0.819
Minematsu et al. (1)	(2017) [305]	0.864	0.897	0.824	0.748	0.667
Avola et al.	(2018) [294]	0.876	0.903	-	-	0.871
Zhao et al. (IFB SU)	(2018) [291]	0.53	0.68	0.65	0.81	0.64
Zhao et al. (IFB KA)	(2018) [291]	0.63	0.72	0.66	0.79	0.61
Zhao et al. (IFB SI)	(2018) [291]	0.53	0.80	0.70	0.88	0.63
<b>Subspace segmentation</b>						
Elqursh and Elgammal (1)	(2012) [321]	0.890	0.770	0.910	-	-
Elqursh and Elgammal (2)	(2012) [321]	0.920	0.880	0.900	-	-
Nonaka et al.	(2013) [323]	0.570	0.780	0.850	-	-
Berger and Seversky	(2014) [324]	0.999	0.999	0.944	0.992	0.998
Sajid et al. - SVM	(2019) [325]	0.95	0.92	0.94	0.87	0.90
Sajid et al. - GMM	(2019) [325]	0.93	0.93	0.92	0.85	0.91

Table 9: F-score comparison on the Hopkins 155 dataset for the *Panoramic background subtraction*, the *Motion compensation* and the *Subspace segmentation* approaches. Values are reported from the original papers.

References		people1	people2	cars1	cars4	cars6
<b>Motion segmentation</b>						
Narayana et al.	(2013) [327]	0.695	0.884	0.508	0.480	0.843
Yin et al.	(2015) [328]	-	-	0.880	0.926	-
Zhu and Elgammal	(2017) [326]	0.814	0.943	0.920	0.916	0.923
Sugimura et al.	(2018) [331]	0.802	0.882	0.901	0.874	0.918
<b>Multi Planes</b>						
Zamalieva et al.	(2014) [348]	0.728	0.901	0.747	0.831	0.890
Zamalieva et al.	(2014) [349]	0.866	0.916	0.822	0.895	0.903
Zamalieva et al. (N=1+MRF)	(2014) [353]	0.52	0.65	-	-	-
Zamalieva et al. (N=50+MRF)	(2014) [353]	0.80	0.77	-	-	-
<b>Split image in blocks</b>						
Kwak et al.	(2011) [362]	0.94	0.87	0.88	-	-
Lim et al.	(2012) [354]	0.93	0.79	0.98	0.92	0.88
Sun et al.	(2016) [360]	0.912	0.965	0.852	0.942	0.923
Wu et al.	(2017) [361]	-	-	-	0.87	0.91

Table 10: F-score comparison on the Hopkins 155 dataset for the *Motion segmentation*, the *Multi Planes* and the *Split image in blocks* approaches. Values are reported from the original papers.

unequal: there are more methods in the first category. There are two popular approaches: panoramic background subtraction and motion compensation. The first has the advantage to construct an exhaustive background model while the second is simple and fast. In the same manner, the video sequences used in general for experimentation are mostly videos where the scene is quite flat, with low depth variations and thus low motion parallax, or far enough from the camera.

According to the f-score results presented in the section 6.3 on simple sequences, the moving background appears to be still challenging. The motion parallax challenge necessitate a comparison of several methods on the same sequences to evaluate its level of completion.

## 7.2. Future directions

*Panoramic background subtraction* and *Dual cameras* approaches are more dedicated to surveillance system with PTZ cameras since the observable scene

is limited by the camera motion. Construct and maintain a panorama with a freely moving camera could suffer from an accumulation of errors during the frame registration process.

Single plane approaches limit the scope of applications while the *Multi planes* and *Split image in blocks* approaches appear to be the more appropriate to handle a larger variety of scene configuration. Those approaches should be tested on more complex environment and more investigated to reach real-time performances for practical applications. The *Motion segmentation* and *Subspace segmentation* are interesting approaches since some of them are based on dense optical flow and represent an accurate scene representation but they suffer from a quite general assumption that the dominant motion or subspace belongs to background and thus reduce the scene approximation to one place.

The comparison of the methods is quite difficult because there is no benchmark designed for background subtraction with a moving camera. The method evaluations provided in the papers are not always computed on the same baseline: different number of frame are used, sometimes a modified version of the methods are used and some video sequences are not publicly available. With this in mind, it could be interesting to create a complete dataset where the proposed sequences have different level of scene complexity, different type of camera motions and different duration.

## 8. Conclusion

We have proposed in this paper a review of methods for moving objects detection with a moving camera categorized into eight different approach groups divided into two big categories. We have chosen to separate the methods into these two categories, one plane and several planes, since the approach to use depends on the scene configuration. For each group, the following conclusions can be made:

- For the approaches based **panoramic background subtraction**, a panorama of the observed scene is first constructed. Then, the current image is reg-

1540 istered to the background model in order to do the subtraction and obtain  
the moving objects. These approaches are often used in the context of  
video surveillance with a PTZ camera. The *panoramic background sub-*  
*traction* approach is well suited for this kind of camera since the part of  
the scene that the camera can observed is limited because it cannot per-  
1545 form a translation. A special attention must be paid on the construction  
of the panorama because errors can be accumulated and caused errors in  
the background subtraction step.

- When several cameras are used, static and moving, it could be interesting  
to couple the information to detect moving objects. In the **dual cameras**  
1550 approaches, when a moving object is detected in the static camera, gener-  
ally with a large-view, the moving camera, generally a PTZ camera, will  
move to detect the moving object. The advantage of using the large-view  
image, compared to the panoramic background subtraction, is that the  
whole background model is updated with the new frames.
- The background subtraction with a **motion compensation** approach is  
1555 the most popular in the literature, as shown by the table 3. The two ad-  
vantages of this method are the ease of implementation and its low time  
computation. The compensation is a 2D transformation which approxi-  
mates the scene by a plane. When the parallax is small, it can be handle  
1560 after the compensation but when the parallax is too large, this approach  
cannot be used.
- Contrary to the three previous approaches, the motion of the camera is  
not compensated to compute the background subtraction. The **subspace**  
**segmentation** approach is based on the apparent motion, computed by  
1565 optical flow algorithms on feature points or on the entire image. These  
trajectories are then clustered or segmented into a subspace representa-  
tion. The clusters segmented as background generally reflect a plane in  
the scene.

- 1570 • The **motion segmentation** approach is also based on trajectories. The motions are analyzed and segmented according to their similarities. The methods presented in this survey go further than just segmented the motions, a background or foreground label is associated to the sets of motions. In the same manner as the subspace representation approach, the background motions generally reflect a plane in the scene.
- 1575 • The **Plane+Parallax** approach was not much studied in the context of detecting moving objects. To the best of our knowledge, only three different methods relate about the *Plane+Parallax* decomposition. This scene representation performs well when the scene contains few parallax and difficulties arise when the scene is composed of several planes.
- 1580 • In the **multi planes** approaches, the scene is approximated by several planes, reals or not. With such a scene representation, most of the parallax effect is directly handled. Nevertheless, if a moving object is big enough in the images, it can be approximated by a plane and considered as a part of the background.
- 1585 • Rather than representing the scene by several planes, the **split image in blocks** approaches divided the image into several blocks. Each block is processed individually in order to find the foreground objects. As in the multi planes approaches, a moving object as to be small in a block in order to approximate the block as a plane.

## 1590 **Acknowledgments**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

- [1] B. Garcia-Garcia, T. Bouwmans, A. Silva, Background subtraction in real applications: Challenges, current models and future directions, Computer Science Review 35 (2020) 100204. doi:10.1016/j.cosrev.2019.100204.

- [2] J. Zheng, Y. Wang, N. Nihan, E. Hallenbeck, Extracting roadway background image: A mode based approach, *Journal of Transportation Research Report*, (2006) 82–88.
- [3] B. Weinstein, Motionmeerkat: integrating motion video detection and ecological monitoring, *Methods in Ecology and Evolution*.  
1600
- [4] B. Weinstein, A computer vision for animal ecology, *Journal of Animal Ecology*.
- [5] E. Sheehan, D. Bridger, S. Nancollas, S. Pittman, PelagiCam: a novel underwater imaging system with computer vision for semi-automated monitoring of mobile marine fauna at offshore structures, *Environmental Monitoring and Assessment*.  
1605
- [6] J. Carranza, C. Theobalt, M. Magnor, H. Seidel, Free-viewpoint video of human actors, *ACM Transactions on Graphics* 22 (3) (2003) 569–577.
- [7] F. E. Baf, T. Bouwmans, Comparison of background subtraction methods for a multimedia learning space, *International Conference on Signal Processing and Multimedia, SIGMAP 2007*.  
1610
- [8] A. M. Ivor, Background subtraction techniques, *International Conference on Image and Vision Computing, New Zealand, IVCNZ 2000*.
- [9] M. Piccardi, Background subtraction techniques: a review, *IEEE International Conference on Systems, Man and Cybernetics*.  
1615
- [10] S. Cheung, C. Kamath, Robust background subtraction with foreground validation for urban traffic video, *Journal of Applied Signal Processing, EURASIP 2005*.
- [11] S. Elhabian, K. El-Sayed, S. Ahmed, Moving object detection in spatial domain using background removal techniques - state-of-art, *Patents on Computer Science* 1 (1) (2008) 32–54.  
1620

- [12] M. Cristani, M. Farenzena, D. Bloisi, V. Murino, Background subtraction for automated multisensor surveillance: A comprehensive review, *EURASIP Journal on Advances in Signal Processing* 2010 (2010) 24.
- 1625 [13] T. Bouwmans, F. E. Baf, 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 (2010) 181–199.
- [14] T. Bouwmans, Traditional Approaches in Background Modeling for Video  
1630 Surveillance, *Handbook Background Modeling and Foreground Detection for Video Surveillance*, Taylor and Francis Group, T. Bouwmans, B. Hoferlin, F. Porikli, A. Vacavant.
- [15] T. Bouwmans, Recent Approaches in Background Modeling for Video  
1635 Surveillance, *Handbook Background Modeling and Foreground Detection for Video Surveillance*, Taylor and Francis Group, T. Bouwmans, B. Hoferlin, F. Porikli, A. Vacavant.
- [16] T. Bouwmans, Traditional and recent approaches in background modeling for foreground detection: An overview, *Computer Science Review* 11 (31–66).
- 1640 [17] 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 (2012) 103–139.
- [18] T. Bouwmans, A. Sobral, S. Javed, S. Jung, E. Zahzah, Decomposition  
1645 into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset, *Computer Science Review*.
- [19] T. Bouwmans, F. E. Baf, B. Vachon, Background Modeling using Mixture

- of Gaussians for Foreground Detection - A Survey, Recent Patents on  
1650 Computer Science, RPCS 2008 1 (3) (2008) 219–237.
- [20] T. Bouwmans, Subspace Learning for Background Modeling: A Survey,  
Recent Patents on Computer Science, RPCS 2009 2 (3) (2009) 223–234.
- [21] T. Bouwmans, E. Zahzah, Robust PCA via principal component pur-  
suit: A review for a comparative evaluation in video surveillance, Special  
1655 Issue on Background Models Challenge, Computer Vision and Image Un-  
derstanding, CVIU 2014 122 (2014) 22–34.
- [22] N. Vaswani, T. Bouwmans, S. Javed, P. Narayanamurthy, Robust Sub-  
space Learning: Robust PCA, Robust Subspace Tracking and Robust  
Subspace Recovery, IEEE Signal Processing Magazine 35 (4) (2018) 32–  
1660 55.
- [23] T. Bouwmans, Z. Javed, M. Sultana, S. Jung, Deep neural network con-  
cepts for background subtraction: A systematic review and comparative  
evaluation, Neural Networks.
- [24] T. Moeslund, E. Granum, A survey of computer vision-based human mo-  
1665 tion capture, Computer Vision and Image Understanding 81 (3) (2001)  
231–268. doi:10.1006/cviu.2000.0897.
- [25] A. Yilmaz, O. Javed, M. Shah, Object tracking, ACM Computing Surveys  
38 (4) (2006) 13–es. doi:10.1145/1177352.1177355.
- [26] M. Cristani, M. Farenzena, D. Bloisi, V. Murino, Background subtrac-  
1670 tion for automated multisensor surveillance: A comprehensive review,  
EURASIP Journal on Advances in Signal Processing 2010 (1) (2010)  
343057. doi:10.1155/2010/343057.
- [27] K. Joshi, D. Thakore, A survey on moving object detection and tracking  
in video surveillance system, International Journal of Soft Computing and  
1675 Engineering 2 (3) (2012) 44–48.

- [28] E. Komagal, B. Yogameena, Foreground segmentation with PTZ camera: a survey, *Multimedia Tools and Applications* 77 (17) (2018) 22489–22542. doi:10.1007/s11042-018-6104-4.
- [29] M. Yazdi, T. Bouwmans, New trends on moving object detection in video images captured by a moving camera : A survey, *Computer Science Review*.  
1680
- [30] K. Toyama, J. Krumm, B. Brumitt, B. Meyers, Wallflower: Principles and practice of background maintenance, *Proceedings of the Seventh IEEE International Conference on Computer Vision* 1 (1999) 255–261. doi:10.1109/ICCV.1999.791228.  
1685
- [31] S. Sanches, C. Oliveira, A. Sementille, V. Freire, Challenging situations for background subtraction algorithms, *Applied Intelligence* (2018) 1–4.
- [32] D. Prasad, C. Prasath, D. Rajan, L. Rachmawati, E. Rajabally, 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).  
1690
- [33] D. Prasad, D. Rajan, L. Rachmawati, E. Rajabally, C. Quek, Video processing from electro-optical sensors for object detection and tracking in maritime environment: A survey, Preprint.
- [34] D. Prasad, D. Rajan, C. Quek, Are object detection assessment criteria ready for maritime computer vision?, Preprint.  
1695
- [35] S. Ramadan, Using time series analysis to visualize and evaluate background subtraction results in computer vision applications, Master Thesis, University of Maryland.
- [36] A. Sanchez-Rodriguez, J. Gonzalez-Castolo, O. Deniz-Suarez, TimeViewer: a Tool for Visualizing the Problems of the Background Subtraction, *Pacific-Rim Symposium, PSIVT 2013* (2013) 372–384.  
1700

- 1705 [37] R. Collins, A. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt, L. Wixson, A system for video surveillance and monitoring, *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- [38] I. Haritaoglu, D. Harwood, L. Davis, W4:Real time surveillance of people and their activities, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8 (22) (2000) 80–85.
- 1710 [39] L. Zhao, Q. Tong, H. Wang, Study on moving-object-detection arithmetic based on W4 theory, *IEEE International Conference on Artificial Intelligence, Management Science and Electronic Commerce, AIMSEC 2011* (2011) 4387–4390.
- 1715 [40] T. Bouwmans, F. Porikli, B. Horferlin, A. Vacavant, *Handbook on Background Modeling and Foreground Detection for Video Surveillance*, CRC Press, Taylor and Francis Group.
- [41] T. Bouwmans, N. Aybat, E. Zahzah, *Handbook on Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*, CRC Press, Taylor and Francis Group.
- 1720 [42] T. Bouwmans, C. Silva, C. Marghes, M. Zitouni, H. Bhaskar, C. Frelicot, On the role and the importance of features for background modeling and foreground detection, *Computer Science Review* 28 (26-91).
- [43] L. Maddalena, A. Petrosino, *Background Subtraction for Moving Object Detection in RGB-D Data: A Survey*, *MDPI Journal of Imaging*.
- 1725 [44] B. Lee, M. Hedley, Background estimation for video surveillance, *Image and Vision Computing New Zealand, IVCNZ 2002* (2002) 315–320.
- [45] P. Graszka, Median mixture model for background-foreground segmentation in video sequences, *Conference on Computer Graphics, Visualization and Computer Vision, WSCG 2014*.

- 1730 [46] S. Roy, A. Ghosh, Real-time Adaptive Histogram Min-Max Bucket (HMMB) Model for Background Subtraction, *IEEE Transactions on Circuits and Systems for Video Technology*.
- [47] A. Elgammal, L. Davis, Non-parametric model for background subtraction, *European Conference on Computer Vision, ECCV 2000 (2000)* 751–767.
- 1735 [48] R. Caseiro, P. Martins, J. Batista, Background Modelling on Tensor Field for Foreground Segmentation, *BMVC 2010 (2010)* 1–12.
- [49] C. Stauffer, E. Grimson, Adaptive background mixture models for real-time tracking, *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 1999 (1999)* 246–252.
- 1740 [50] S. Varadarajan, P. Miller, H. Zhou, Spatial mixture of Gaussians for dynamic background modelling, *IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2013 (2013)* 63–68.
- [51] F. E. Baf, T. Bouwmans, B. Vachon, Fuzzy integral for moving object detection, *IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2008 (2008)* 1729–1736.
- 1745 [52] F. E. Baf, T. Bouwmans, B. Vachon, Type-2 fuzzy mixture of Gaussians model: Application to background modeling, *International Symposium on Visual Computing, ISVC 2008 (2008)* 772–781.
- [53] F. E. Baf, T. Bouwmans, 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 (2009)* 60–65.
- 1750 [54] O. Munteanu, T. Bouwmans, E. Zahzah, R. Vasiu, The detection of moving objects in video by background subtraction using Dempster-Shafer theory, *Transactions on Electronics and Communications* 60 (1).
- 1755

- [55] N. Oliver, B. Rosario, A. Pentland, A Bayesian computer vision system for modeling human interactions, International Conference on Vision Systems, ICVS 1999.
- 1760 [56] D. Farcas, T. Bouwmans, Background modeling via a supervised subspace learning, International Conference on Image, Video Processing and Computer Vision, IVPCV 2010 (2010) 1–7.
- [57] D. Farcas, C. Marghes, T. Bouwmans, Background subtraction via incremental maximum margin criterion: A discriminative approach, Machine  
1765 Vision and Applications 23 (6) (2012) 1083–1101.
- [58] C. Marghes, T. Bouwmans, Background modeling via incremental maximum margin criterion, International Workshop on Subspace Methods, ACCV 2010 Workshop Subspace 2010.
- [59] C. Marghes, T. Bouwmans, R. Vasiu, Background modeling and foreground  
1770 detection via a reconstructive and discriminative subspace learning approach, International Conference on Image Processing, Computer Vision, and Pattern Recognition, IPCV 2012.
- [60] E. Candes, X. Li, Y. Ma, J. Wright, Robust principal component analysis?, International Journal of ACM 58 (3).
- 1775 [61] A. Sobral, T. Bouwmans, E. Zahzah, Double-constrained RPCA based on saliency maps for foreground detection in automated maritime surveillance, ISBC 2015 Workshop conjunction with AVSS 2015.
- [62] S. Javed, A. Mahmood, T. Bouwmans, S. Jung, Motion-Aware Graph Regularized RPCA for Background Modeling of Complex Scenes, Scene Background  
1780 Modeling Contest, International Conference on Pattern Recognition, ICPR 2016.
- [63] S. Javed, A. Mahmood, T. Bouwmans, S. Jung, Spatiotemporal Low-rank Modeling for Complex Scene Background Initialization, IEEE Transactions on Circuits and Systems for Video Technology.

- 1785 [64] G. Ramirez-Alonso, M. Chacon-Murguia, Self-adaptive SOM-CNN neural system for dynamic object detection in normal and complex scenarios, *Pattern Recognition*.
- [65] J. Ramirez-Quintana, M. Chacon-Murguia, Self-organizing retinotopic maps applied to background modeling for dynamic object segmentation  
1790 in video sequences, *International Joint Conference on Neural Networks, IJCNN 2013*.
- [66] A. Schofield, P. Mehta, T. Stonham, A system for counting people in video images using neural networks to identify the background scene, *Pattern Recognition* 29 (1996) 1421–1428.
- 1795 [67] T. Chang, T. Ghandi, M. Trivedi, Vision modules for a multi sensory bridge monitoring approach, *International Conference on Intelligent Transportation Systems, ITSC 2004 (2004)* 971–976.
- [68] G. Cinar, J. Principe, Adaptive background estimation using an information theoretic cost for hidden state estimation, *International Joint Conference on Neural Networks, IJCNN 2011*.  
1800
- [69] S. Messelodi, C. Modena, N. Segata, M. Zanin, A Kalman filter based background updating algorithm robust to sharp illumination changes, *International Conference on Image Analysis and Processing, ICIAP 2005* 3617 (2005) 163–170.
- 1805 [70] K. Toyama, J. Krumm, B. Brumitt, B. Meyers, Wallflower: Principles and practice of background maintenance, *International Conference on Computer Vision, ICCV 1999 (1999)* 255–261.
- [71] C. Wren, A. Azarbayejani, Pfinder: Real-time tracking of the human body, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19 (7)  
1810 (1997) 780–785.

- [72] Z. Zivkovic, Efficient adaptive density estimation per image pixel for the task of background subtraction, *Pattern Recognition Letters* 27 (7) (2006) 773–780.
- [73] J. Pulgarin-Giraldo, A. Alvarez-Meza, D. Insuasti-Ceballos, T. Bouwmans, G. Castellanos-Dominguez, GMM background modeling using divergence-based weight updating, *Conference Ibero American Congress on Pattern Recognition, CIARP 2016*.  
1815
- [74] B. Garcia-Garcia, F. Gallegos-Funes, A. Rosales-Silva, A Gaussian-Median Filter for Moving Objects Segmentation Applied for Static Scenarios, *Intelligent Systems Conference, IntelliSys 2018 (2018)* 478–493.  
1820
- [75] T. Elguebaly, N. Bouguila, Finite asymmetric generalized Gaussian mixture models learning for infrared object detection, *Computer Vision and Image Understanding*.
- [76] D. Mukherjee, J. Wu, Real-time video segmentation using Student’s t mixture model, *International Conference on Ambient Systems, Networks and Technologies, ANT 2012 (2012)* 153–160.  
1825
- [77] L. Guo, M. Du, Student’s t-distribution mixture background model for efficient object detection, *IEEE International Conference on Signal Processing, Communication and Computing, ICSPCC 2012 (2012)* 410–414.
- [78] T. Haines, T. Xiang, Background subtraction with Dirichlet processes, *European Conference on Computer Vision, ECCV 2012*.  
1830
- [79] W. Fan, N. Bouguila, Online variational learning of finite Dirichlet mixture models, *Evolving Systems*.
- [80] A. Faro, D. Giordano, C. Spampinato, Adaptive background modeling integrated with luminosity sensors and occlusion processing for reliable vehicle detection, *IEEE Transactions on Intelligent Transportation Systems* 12 (4) (2011) 1398–1412.  
1835

- 1840 [81] T. Zin, P. Tin, T. Toriu, H. Hama, A new background subtraction method using bivariate Poisson process, *International Conference on Intelligent Information Hiding and Multimedia Signal Processing* (2014) 419–422.
- [82] D. Liang, S. Kaneko, M. Hashimoto, K. Iwata, X. Zhao, Co-occurrence Probability based Pixel Pairs Background Model for Robust Object Detection in Dynamic Scenes, *Pattern Recognition* 48 (4) (2015) 1374–1390.
- 1845 [83] D. Liang, S. Kaneko, M. Hashimoto, K. Iwata, X. Zhao, Y. Satoh, Co-occurrence-based adaptive background model for robust object detection, *International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2013*.
- [84] D. Liang, S. Kaneko, M. Hashimoto, K. Iwata, X. Zhao, Y. Satoh, Robust object detection in severe imaging conditions using co-occurrence background model, *International Journal of Optomechatronics* (2014) 14–29.
- 1850 [85] J. Rosell-Ortega, G. Andreu-Garcia, A. Rodas-Jorda, V. Atienza-Vanacloig, Background Modelling in Demanding Situations with Confidence Measure, *IAPR International Conference on Pattern Recognition, ICPR 2008*.
- 1855 [86] J. Rosell-Ortega, G. Andreu, V. Atienza, F. Lopez-Garcia, Background modeling with motion criterion and multi-modal support, *International Conference on Computer Vision Theory and Applications, VISAPP 2010*.
- [87] O. Barnich, M. V. Droogenbroeck, ViBe: a powerful random technique to estimate the background in video sequences, *International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2009* (2009) 945–948.
- 1860 [88] P. St-Charles, G. Bilodeau, R. Bergevin, Flexible background subtraction with self-balanced local sensitivity, *IEEE Change Detection Workshop, CDW 2014*.

- 1865 [89] P. St-Charles, G. Bilodeau, R. Bergevin, A self-adjusting approach to change detection based on background word consensus, IEEE Winter Conference on Applications of Computer Vision, WACV 2015.
- [90] F. Tombari, A. Lanza, L. D. Stefano, S. Mattoccia, Non-linear Parametric Bayesian Regression for Robust Background Subtraction, IEEE Workshop on Motion and Video Computing, MOTION 2009.
- 1870 [91] A. Lanza, F. Tombari, L. D. Stefano, Accurate and efficient background subtraction by monotonic second-degree polynomial fitting, IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2010.
- [92] T. Bouwmans, F. E. Baf, Modeling of Dynamic Backgrounds by Type-2 Fuzzy Gaussians Mixture Models, MASAUM Journal of Basic and Applied Sciences 1 (2) (2009) 265–277.
- 1875 [93] Z. Zhao, T. Bouwmans, X. Zhang, Y. Fang, A Fuzzy Background Modeling Approach for Motion Detection in Dynamic Backgrounds, International Conference on Multimedia and Signal Processing.
- [94] H. Zhang, D. Xu, Fusing color and gradient features for background model, International Conference on Signal Processing, ICSP 2006 2 (7).
- 1880 [95] H. Zhang, D. Xu, Fusing color and texture features for background model, International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2006 4223 (7) (2006) 887–893.
- 1885 [96] F. E. Baf, T. Bouwmans, B. Vachon, Foreground detection using the Choquet integral, International Workshop on Image Analysis for Multimedia Interactive Integral, WIAMIS 2008 (2008) 187–190.
- [97] P. Chiranjeevi, S. Sengupta, Interval-valued model level fuzzy aggregation-based background subtraction, IEEE Transactions on Cybernetics.

- 1890 [98] S. Javed, S. Oh, A. Sobral, T. Bouwmans, S. Jung, Background subtraction via superpixel-based online matrix decomposition with structured foreground constraints, Workshop on Robust Subspace Learning and Computer Vision, ICCV 2015.
- [99] S. Javed, A. Mahmood, T. Bouwmans, S. Jung, Background-foreground  
1895 modeling based on spatiotemporal sparse subspace clustering, IEEE Transactions on Image Processing.
- [100] B. Rezaei, S. Ostadabbas, Background Subtraction via Fast Robust Matrix Completion, International Workshop on RSL-CV in conjunction with ICCV 2017.
- 1900 [101] B. Rezaei, S. Ostadabbas, Moving Object Detection through Robust Matrix Completion Augmented with Objectness, IEEE Journal of Selected Topics in Signal Processing.
- [102] N. Vaswani, T. Bouwmans, S. Javed, P. Narayanamurthy, Robust PCA and Robust Subspace Tracking: A Comparative Evaluation, Statistical  
1905 Signal Processing Workshop, SSP 2018.
- [103] S. Prativadibhayankaram, H. Luong, T. Le, A. Kaup, Compressive online video backgroundforeground separation using multiple prior information and optical flow, MDPI Journal of Imaging.
- [104] J. He, L. Balzano, A. Szlam, Incremental gradient on the grassmannian  
1910 for online foreground and background separation in subsampled video, International on Conference on Computer Vision and Pattern Recognition, CVPR 2012.
- [105] P. Rodriguez, B. Wohlberg, Incremental principal component pursuit for video background modeling, Journal of Mathematical Imaging and Vision  
1915 55 (1) (2016) 1–18.
- [106] H. Guo, C. Qiu, N. Vaswani, Practical ReProCS for separating sparse and low-dimensional signal sequences from their sum, Preprint.

- [107] P. Narayanamurthy, N. Vaswani, A Fast and Memory-efficient Algorithm for Robust PCA (MEROP), IEEE International Conference on Acoustics, Speech, and Signal, ICASSP 2018.
- 1920
- [108] S. Javed, T. Bouwmans, S. Jung, Stochastic decomposition into low rank and sparse tensor for robust background subtraction, ICDP 2015.
- [109] A. Sobral, S. Javed, S. Jung, T. Bouwmans, E. Zahzah, Online stochastic tensor decomposition for background subtraction in multispectral video sequences, Workshop on Robust Subspace Learning and Computer Vision, ICCV 2015.
- 1925
- [110] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, S. Yan, Tensor robust principal component analysis with a new tensor nuclear norm, IEEE Transactions on Pattern Analysis and Machine Intelligence.
- [111] D. Driggs, S. Becker, J. Boyd-Graberz, Tensor robust principal component analysis: Better recovery with atomic norm regularization, Preprint.
- 1930
- [112] A. Tavakkoli, Foreground-background segmentation in video sequences using neural networks, Intelligent Systems: Neural Networks and Applications.
- [113] L. Maddalena, A. Petrosino, A self-organizing approach to detection of moving patterns for real-time applications, Advances in Brain, Vision, and Artificial Intelligence 4729 (2007) 181–190.
- 1935
- [114] L. Maddalena, A. Petrosino, A self-organizing neural system for background and foreground modeling, International Conference on Artificial Neural Networks, ICANN 2008 (2008) 652–661.
- 1940
- [115] L. Maddalena, A. Petrosino, Neural model-based segmentation of image motion, KES 2008 (2008) 57–64.
- [116] L. Maddalena, A. Petrosino, A self organizing approach to background subtraction for visual surveillance applications, IEEE Transactions on Image Processing 17 (7) (2008) 1168–1177.
- 1945

- [117] L. Maddalena, A. Petrosino, Multivalued background/foreground separation for moving object detection, International Workshop on Fuzzy Logic and Applications, WILF 2009 (2009) 263–270.
- [118] L. Maddalena, A. Petrosino, A fuzzy spatial coherence-based approach to background/foreground separation for moving object detection, Neural Computing and Applications, NCA 2010 (2010) 1–8.
- [119] L. Maddalena, A. Petrosino, The SOBS algorithm: What are the limits?, IEEE Workshop on Change Detection, CVPR 2012.
- [120] L. Maddalena, A. Petrosino, The 3dSOBS+ algorithm for moving object detection, Computer Vision and Image Understanding, CVIU 2014 122 (2014) 65–73.
- [121] M. Chacon-Mugua, S. Gonzalez-Duarte, P. Vega, Simplified SOM-neural model for video segmentation of moving objects, International Joint Conference on Neural Networks, IJCNN 2009 (2009) 474–480.
- [122] M. Chacon-Murguia, G. Ramirez-Alonso, S. Gonzalez-Duarte, Improvement of a neural-fuzzy motion detection vision model for complex scenario conditions, International Joint Conference on Neural Networks, IJCNN 2013.
- [123] G. Gemignani, A. Rozza, A novel background subtraction approach based on multi-layered self organizing maps, IEEE International Conference on Image Processing.
- [124] N. Goyette, P. Jodoin, F. Porikli, J. Konrad, P. Ishwar, Changedetection.net: A new change detection benchmark dataset, IEEE Workshop on Change Detection, CDW 2012 in conjunction with CVPR 2012.
- [125] L. Maddalena, A. Petrosino, 3D neural model-based stopped object detection, International Conference on Image Analysis and Processing, ICIAP 2009 (2009) 585–593.

- [126] L. Maddalena, A. Petrosino, Self organizing and fuzzy modelling for parked vehicles detection, *Advanced Concepts for Intelligent Vision Systems, ACVIS 2009* (2009) 422–433.  
1975
- [127] L. Maddalena, A. Petrosino, Stopped object detection by learning foreground model in videos, *IEEE Transactions on Neural Networks and Learning Systems* 24 (5) (2013) 723–735.
- [128] R. Guo, H. Qi, Partially-sparse restricted Boltzmann machine for background modeling and subtraction, *International Conference on Machine Learning and Applications, ICMLA 2013* (2013) 209–214.  
1980
- [129] Z. Qu, S. Yu, M. Fu, Motion background modeling based on context-encoder, *IEEE International Conference on Artificial Intelligence and Pattern Recognition, ICAIPR 2016*.
- [130] L. Xu, Y. Li, Y. Wang, E. Chen, Temporally Adaptive Restricted Boltzmann Machine for Background Modeling, *American Association for Artificial Intelligence, AAAI 2015*.  
1985
- [131] P. Xu, M. Ye, Q. Liu, X. Li, L. Pei, J. Ding, Motion detection via a couple of auto-encoder networks, *International Conference on Multimedia and Expo, ICME 2014*.  
1990
- [132] P. Xu, M. Ye, X. Li, Q. Liu, Y. Yang, J. Ding, Dynamic background learning through deep auto-encoder networks, *ACM International Conference on Multimedia*.
- [133] M. Babaei, D. Dinh, G. Rigoll, A deep convolutional neural network for background subtraction, *Preprint*.  
1995
- [134] C. Bautista, C. Dy, M. Manalac, R. O. and M. Cordel, Convolutional neural network for vehicle detection in low resolution traffic videos, *TENCON 2016*.

- [135] M. Braham, M. V. Droogenbroeck, Deep background subtraction with scene-specific convolutional neural networks, International Conference on Systems, Signals and Image Processing, IWSSIP 2016 (2016) 1–4.  
2000
- [136] L. P. Cinelli, Anomaly detection in surveillance videos using deep residual networks, Master Thesis, Universidade de Rio de Janeiro.
- [137] K. Lim, W. Jang, C. Kim, Background subtraction using encoder-decoder structured convolutional neural network, IEEE International Conference on Advanced Video and Signal based Surveillance, AVSS 2017.  
2005
- [138] S. Choo, W. Seo, D. Jeong, N. Cho, Multi-scale recurrent encoder-decoder network for dense temporal classification, IAPR International Conference on Pattern Recognition, ICPR 2018 (2018) 103–108.
- [139] S. Choo, W. Seo, D. Jeong, N. Cho, Learning background subtraction by video synthesis and multi-scale recurrent networks, Asian Conference on Computer Vision, ACCV 2018.  
2010
- [140] A. Farnoosh, B. Rezaei, S. Ostadabbas, DeepPBM: deep probabilistic background model estimation from video sequences, Preprint.
- [141] S. Ammar, T. Bouwmans, N. Zaghden, M. Neji, Moving Objects Segmentation based on DeepSphere in Video Surveillance, International Symposium on Visual Computing, ISVC 2019.  
2015
- [142] S. Ammar, T. Bouwmans, N. Zaghden, M. Neji, A Deep Detector Classifier (DeepDC) for moving objects segmentation and classification in video surveillance, IET Image Processing.  
2020
- [143] D. Zeng, M. Zhu, Combining background subtraction algorithms with convolutional neural network, Preprint.
- [144] Y. Wang, Z. Luo, P. Jodoin, Interactive deep learning method for segmenting moving objects, Pattern Recognition Letters.

- 2025 [145] S. Lee, D. Kim, Background subtraction using the factored 3-way restricted boltzmann machines, Preprint.
- [146] T. Nguyen, C. Pham, S. Ha, J. Jeon, Change detection by training a triplet network for motion feature extraction, *IEEE Transactions on Circuits and Systems for Video Technology*.
- 2030 [147] M. Shafiee, P. Siva, P. Fieguth, A. Wong, Embedded motion detection via neural response mixture background modeling, *International Conference on Computer Vision and Pattern Recognition, CVPR 2016*.
- [148] M. Shafiee, P. Siva, P. Fieguth, A. Wong, Real-time embedded motion detection via neural response mixture modeling, *Journal of Signal Processing Systems*.
- 2035 [149] Y. Zhang, X. Li, Z. Zhang, F. Wu, L. Zhao, Deep learning driven blockwise moving object detection with binary scene modeling, *Neurocomputing*.
- [150] X. Zhao, Y. Chen, M. Tang, J. Wang, Joint background reconstruction and foreground segmentation via a two-stage convolutional neural network, Preprint.
- 2040 [151] M. Sultana, A. Mahmood, S. Javed, S. Jung, Unsupervised deep context prediction for background estimation and foreground segmentation, *Machine Vision and Applications*.
- [152] M. Sultana, S. Jung, Illumination Invariant Foreground Object Segmentation using ForeGANs, *International Workshop on Frontiers of Computer Vision, IW-FCV 2019*.
- 2045 [153] M. Sultana, A. Mahmood, T. Bouwmans, S. Jung, Unsupervised Adversarial Learning for Dynamic Background Modeling, *International Workshop on Frontiers of Computer Vision, IW-FCV 2020*.
- 2050 [154] M. Sultana, A. Mahmood, T. Bouwmans, S. Jung, Dynamic Background Subtraction Using Least Square Adversarial Learning, *IEEE International Conference on Image Processing, ICIP 2020*.

- [155] Y. Chan, Deep learning-based scene-awareness approach for intelligent change detection in videos, *Journal of Electronic Imaging* 28 (1) (2019) 013038.  
2055
- [156] T. Minematsu, A. Shimada, R. Taniguchi, Analytics of deep neural network in change detection, *IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2017*.
- [157] T. Minematsu, H. Uchiyama, A. Shimada, R. Taniguchi, Analytics of deep neural network-based background subtraction, *MDPI Journal of Imaging*.  
2060
- [158] T. Minematsu, A. Shimada, R. Taniguchi, Simple background subtraction constraint for weakly supervised background subtraction network, *IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2019*.
- [159] T. Minematsu, A. Shimada, R. Taniguchi, Rethinking background and foreground in deep neural network-based background subtraction, *IEEE International Conference on Image Processing, ICIP 2020*.  
2065
- [160] K. Karmann, A. V. Brand, Moving object recognition using an adaptive background memory, *Time-Varying Image Processing and Moving Object Recognition, Elsevier*.  
2070
- [161] M. Boninsegna, A. Bozzoli, A tunable algorithm to update a reference image, *Signal Processing: Image Communication* 16 (4) (2000) 1353–365.
- [162] D. Fan, M. Cao, C. Lv, An updating method of self-adaptive background for moving objects detection in video, *International Conference on Audio, Language and Image Processing, ICALIP 2008 (2008)* 1497–1501.  
2075
- [163] T. Chang, T. Ghandi, M. Trivedi, Computer vision for multi-sensory structural health monitoring system, *International Conference on Intelligent Transportation Systems, ITSC 2004*.

- 2080 [164] C. Wren, F. Porikli, Waviz: Spectral similarity for object detection, IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, PETS 2005.
- [165] F. Porikli, C. Wren, Change detection by frequency decomposition: Wave-back, International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2005.
- 2085 [166] V. Cevher, D. Reddy, M. Duarte, A. Sankaranarayanan, R. Chellappa, R. Baraniuk, Compressive sensing for background subtraction, European Conference on Computer Vision, ECCV 2008.
- [167] J. Mota, L. Weizman, N. Deligiannis, Y. Eldar, M. Rodrigues, Reference-based compressed sensing: A sample complexity approach, IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2090 2016.
- [168] G. Warnell, D. Reddy, R. Chellappa, Adaptive rate compressive sensing for background subtraction, IEEE International Conference on Acoustics, Speech, and Signal Processing.
- 2095 [169] G. Warnell, S. Bhattacharya, R. Chellappa, T. Basar, Adaptive-rate compressive sensing via side information, IEEE Transactions on Image Processing 24 (11) (2015) 3846–3857.
- [170] R. Davies, L. Mihaylova, N. Pavlidis, I. Eckley, The effect of recovery algorithms on compressive sensing background subtraction, Workshop Sensor Data Fusion: Trends, Solutions, and Applications. 2100
- [171] H. Xiao, Y. Liu, M. Zhang, Fast  $l_1$ -minimization algorithm for robust background subtraction, EURASIP Journal on Image and Video Processing.
- [172] D. Kuzin, O. Isupova, L. Mihaylova, Compressive sensing approaches for autonomous object detection in video sequences, Sensor Data Fusion: Trends, Solutions, Applications, SDF 2015 (2015) 1–6. 2105

- [173] D. Kuzin, O. Isupova, L. Mihaylova, Compressive sensing approaches for autonomous object detection in video sequences, Preprint.
- [174] D. Kuzin, O. Isupova, L. Mihaylova, Spatio-Temporal Structured Sparse Regression with Hierarchical Gaussian Process Priors, *IEEE Transactions on Signal Processing* 66 (17) (2018) 4598–4611.
- [175] D. Kuzin, Sparse machine learning methods for autonomous decision making, PhD Thesis, University of Sheffield.
- [176] J. Giraldo, T. Bouwmans, Semi-supervised background subtraction of unseen videos: Minimization of the total variation of graph signals, *IEEE International Conference on Image Processing, ICIP 2020*.
- [177] M. Molinier, T. Hame, H. Ahola, Connected components analysis for traffic monitoring in image sequences acquired from a helicopter, *Scandinavian Conference, SCIA 2005* (2005) 141.
- [178] Y. Chung, J. Wang, S. Cheng, Progressive background image generation, *IPPR Conference on Computer Vision, Graphics and Image Processing, CVGIP 2002* (2002) 858–865.
- [179] R. M. Colque, G. Camara-Chavez, Progressive background image generation of surveillance traffic videos based on a temporal histogram ruled by a reward/penalty function, *SIBGRAPI 2011*.
- [180] W. Long, Y. Yang, Stationary background generation: An alternative to the difference of two images, *Pattern Recognition* 12 (23) (1990) 1351–1359.
- [181] H. Wang, D. Suter, A novel robust statistical method for background initialization and visual surveillance, *Asian Conference on Computer Vision, ACCV 2006* (2006) 328–337.
- [182] D. Gutchess, M. Trajkovic, E. Cohen, D. Lyons, A. Jain, A background model initialization for video surveillance, *International Conference on Computer Vision, ICCV 2001* (2001) 733–740.

- 2135 [183] C. Chen, J. Aggarwal, An adaptive background model initialization algorithm with objects moving at different depths, International Conference on Image Processing, ICIP 2008 (2008) 2264–2267.
- [184] B. Laugraud, S. Pierard, M. V. Droogenbroeck, LaBGen-P: Apixel-level stationary background generation method based on LaBGen, Scene Background Modeling Contest in conjunction with ICPR 2016.
- 2140 [185] B. Laugraud, S. Pierard, M. V. Droogenbroeck, A method based on motion detection for generating the background of a scene, Pattern Recognition Letters.
- [186] B. Laugraud, S. Pierard, M. V. Droogenbroeck, LaBGen-P-Semantic: A First Step for Leveraging Semantic Segmentation in Background Generation, MDPI Journal of Imaging 4 (7).
- 2145 [187] A. Sobral, T. Bouwmans, E. Zahzah, Comparison of matrix completion algorithms for background initialization in videos, SBMI 2015 Workshop in conjunction with ICIAP 2015.
- [188] A. Sobral, E. Zahzah, Matrix and tensor completion algorithms for background model initialization: A comparative evaluation, Special Issue on Scene Background Modeling and Initialization, Pattern Recognition Letters.
- 2150 [189] H. Lin, T. Liu, J. Chuang, A probabilistic SVM approach for background scene initialization, International Conference on Image Processing, ICIP 2002 3 (2002) 893–896.
- [190] M. Gregorio, M. Giordano, Background estimation by weightless neural networks, Pattern Recognition Letters.
- [191] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, M. Cohen, Interactive digital photomontage, ACM Transactions on Graphics 23.
- 2160

- [192] P. Jodoin, L. Maddalena, A. Petrosino, Extensive benchmark and survey of modeling methods for scene background initialization, *IEEE Transactions on Image Processing* (2017) 5244–5256.
- 2165 [193] L. Maddalena, 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.
- [194] L. Maddalena, A. Petrosino, Towards benchmarking scene background initialization, *Workshop on Scene Background Modeling and Initialization in conjunction with ICIAP 2015 1* (2015) 469–476.
- 2170 [195] T. Bouwmans, L. Maddalena, A. Petrosino, Scene background initialization: a taxonomy, *Pattern Recognition Letters*.
- [196] F. E. Baf, T. Bouwmans, B. Vachon, A Fuzzy Approach for Background Subtraction, *IEEE International Conference on Image Processing, ICIP 2008* (2008) 2648–2651.
- 2175 [197] Q. Zang, R. Klette, Evaluation of an adaptive composite Gaussian model in video surveillance, *CITR Technical Report 114*, Auckland University.
- [198] B. White, M. Shah, Automatically tuning background subtraction parameters using particle swarm optimization, *IEEE International Conference on Multimedia and Expo, ICME 2007* (2007) 1826–1829.
- 2180 [199] P. KaewTraKulPong, R. Bowden, An improved adaptive background mixture model for real-time tracking with shadow detection, *AVBS 2001*.
- [200] A. Pnevmatikakis, L. Polymenakos, 2D person tracking using Kalman filtering and adaptive background learning in a feedback loop, *Proceedings of the CLEAR Workshop 2006 4122* (2006) 151–160.
- 2185 [201] D. Lee, Improved adaptive mixture learning for robust video background modeling, *IAPR Workshop on Machine Vision for Applications, MVA 2002* (2002) 443–446.

- 2190 [202] M. Sigari, N. Mozayani, H. Pourreza, Fuzzy Running Average and Fuzzy Background Subtraction: Concepts and Application, *International Journal of Computer Science and Network Security* 8 (2) (2008) 138–143.
- [203] M. Sigari, Fuzzy Background Modeling/Subtraction and its Application in Vehicle Detection, *World Congress on Engineering and Computer Science, WCECS 2008*.
- 2195 [204] Y. Zhang, Z. Liang, Z. Hou, H. Wang, M. Tan, An adaptive mixture Gaussian background model with online background reconstruction and adjustable foreground merge time for motion segmentation, *International Conference on Industrial Technology, ICIT 2005* (2005) 23–27.
- 2200 [205] H. Wang, D. Suter, A re-evaluation of mixture-of-Gaussians background modeling, *International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2005* (2005) 1017–1020.
- [206] F. Porikli, Human body tracking by adaptive background models and mean-shift analysis, *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, PETS 2003*.
- 2205 [207] D. Magee, Tracking multiple vehicles using foreground, background and motion models, *Image and Vision Computing* 22 (2004) 143–155.
- [208] R. Radke, S. Andra, O. Al-Kofahi, B. Roysam, Image Change Detection Algorithms: A Systematic Survey, *IEEE Transactions on Image Processing* 14 (3) (2005) 294–307.
- 2210 [209] D. Toth, T. Aach, V. Metzler, Illumination-invariant change detection, *IEEE Southwest Symposium on Image Analysis and Interpretation, SSIAP 2000* (2000) 3–7.
- 2215 [210] D. Toth, T. Aach, V. Metzler, Bayesian spatio-temporal motion detection under varying illumination, *European Signal Processing Conference, EUSIPCO 2000* (2000) 2081–2084.

- [211] G. Pajares, J. Ruz, J. M. de la Cruz, Performance analysis of homomorphic systems for image change detection, *IBPRIA 2005* (2005) 563–570.
- [212] B. Xie, V. Ramesh, T. Boult, Sudden illumination change detection using order consistency, *Image and Vision Computing* 22 (2) (2004) 117–125.
- 2220 [213] M. Singh, V. Parameswaran, V. Ramesh, Order consistent change detection via fast statistical significance testing, *IEEE Computer Vision and Pattern Recognition Conference, CVPR 2008*.
- [214] T. Aach, A. Kaup, R. Mester, Statistical model-based change detection in moving video, *Signal Processing*, (1993) 165–180.
- 2225 [215] T. Aach, A. Kaup, R. Mester, Change detection in image sequences using Gibbs random fields: a Bayesian approach, *IEEE Workshop Intelligent Signal Processing and Communications Systems*.
- [216] T. Aach, A. Kaup, Bayesian algorithms for adaptive change detection in image sequences using Markov random fields, *Signal Processing Image Communication* 7 (1995) 147–160.
- 2230 [217] R. Mester, T. Aach, L. Duembgen, Illumination-invariant change detection using a statistical colinearity criterion, *DAGM 2001* (2001) 170–177.
- [218] T. Aach, L. Dumbgen, R. Mester, D. Toth, Bayesian illumination-invariant motion detection, *IEEE International Conference on Image Processing, ICIP 2001* 3 (2001) 640–643.
- 2235 [219] T. Aach, D. Toth, R. Mester, Motion estimation in varying illumination using a total least squares distance measure, *Picture Coding Symposium, PCS 2003* (2003) 145–148.
- [220] R. Collins, A. Lipton, T. Kanade, H. Fujiyoshi, D. Duggin, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, A system for video surveillance and monitoring, *Technical Report CMU-RI-TR-00-12*, Robotics Institute, Carnegie Mellon University.
- 2240

- 2245 [221] M. Chacon-Muguia, S. Gonzalez-Duarte, An adaptive neural-fuzzy approach for object detection in dynamic backgrounds for surveillance systems, *IEEE Transactions on Industrial Electronics*.
- [222] E. Stringa, Morphological change detection algorithms for surveillance applications, *British Machine Vision Conference, BMVC 2000*.
- 2250 [223] F. Rahman, A. Hussain, W. Zaki, H. Zaman, N. Tahir, Enhancement of background subtraction techniques using a second derivative in gradient direction filter, *Journal of Electrical and Computer Engineering* 2013 (2013) 12.
- [224] P. Rosin, E. Ioannidis, Evaluation of global image thresholding for change detection, *Pattern Recognition Letters* 24 (2003) 2345–2356.
- 2255 [225] Y. Wang, P. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, P. Ishwar, CDnet 2014: an expanded change detection benchmark dataset, *IEEE Workshop on Change Detection, CDW 2014 in conjunction with CVPR 2014*.
- [226] P. Jodoin, Motion detection: Unsolved issues and [potential] solutions, *Invited Talk, SBMI 2015 in conjunction with ICIAP 2015*.
- 2260 [227] L. Lim, H. Keles, Foreground segmentation using a triplet convolutional neural network for multiscale feature encoding, *Preprint*.
- [228] M. Braham, S. Pierard, M. V. Droogenbroeck, Semantic Background Subtraction, *IEEE International Conference on Image Processing, ICIP 2017*.
- [229] D. Zeng, X. Chen, M. Zhu, M. Goesele, A. Kuijper, Background Subtraction with Real-time Semantic Segmentation, *Preprint*.
- 2265 [230] P. Rodriguez, B. Wohlberg, Translational and rotational jitter invariant incremental principal component pursuit for video background modeling, *IEEE International Conference on Image Processing, ICIP 2015*.

- 2270 [231] O. Karadag, O. Erdas, Evaluation of the robustness of deep features on the change detection problem, IEEE Signal Processing and Communications Applications Conference, SIU 2018 (2018) 1–4.
- [232] G. Silva, P. Rodriguez, Jitter invariant incremental principal component pursuit for video background modeling on the TK1, Asilomar Conference on Signals, Systems, and Computers, ACSSC 2015.
- 2275 [233] G. Chau, P. Rodriguez, Panning and jitter invariant incremental principal component pursuit for video background modeling, International Workshop on RSL-CV in conjunction with ICCV 2017.
- [234] J. He, D. Zhang, L. Balzano, T. Tao, Iterative grassmannian optimization for robust image alignment, Image and Vision Computing.
- 2280 [235] J. He, D. Zhang, L. Balzano, T. Tao, Iterative online subspace learning for robust image alignment, IEEE Conference on Automatic Face and Gesture Recognition, FG 2013.
- [236] B. Wohlberg, Endogenous convolutional sparse representations for translation invariant image subspace models, IEEE International Conference on Image Processing, ICIP 2014.
- 2285 [237] L. Lim, H. Keles, Foreground segmentation using convolutional neural networks for multiscale feature encoding, Pattern Recognition Letters 112 (2018) 256–262.
- [238] L. Lim, I. Ang, H. Keles, Learning multi-scale features for foreground segmentation, Preprint.
- 2290 [239] K. Xue, Y. Liu, G. Ogunmakin, J. Chen, J. Zhang, Panoramic Gaussian mixture model and large-scale range background subtraction method for PTZ camera-based surveillance systems, Machine Vision and Applications 24 (3) (2013) 477–492. doi:10.1007/s00138-012-0426-4.

- 2295 [240] M. Irani, P. Anadan, J. Bergen, R. Kumar, S. Hsu, Efficient representations of video sequences and their application, *Signal Processing: Image Communication* 8 (4) (1996) 327–351.
- [241] O. Faugeras, *Panoramic vision: Sensors, theory, and applications*doi: 10.1007/978-1-4757-3482-9.
- 2300 [242] M. Brown, D. Lowe, Recognising panoramas, *Proceedings Ninth IEEE International Conference on Computer Vision* (2003) 1218–1225 vol.2doi: 10.1109/ICCV.2003.1238630.
- [243] M. Brown, D. Lowe, Automatic panoramic stitching using invariant features, *International Journal on Computer Vision (IJCV)* 74 (1) (2007) 59–73. doi:10.1007/s11263-006-0002-3.
- 2305 [244] L. Brown, A survey of image registration techniques, *ACM Computing Surveys* 24 (4) (1992) 325–376. doi:10.1145/146370.146374.
- [245] B. Zitova, J. Flusser, Image registration methods: A survey, *Image and Vision Computing* 21 (11) (2003) 977–1000. doi:10.1016/S0262-8856(03)00137-9.
- 2310 [246] A. Mittal, D. Huttenlocher, Scene modeling for wide area surveillance and image synthesis, *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)* 2 (2000) 160–167. doi:10.1109/CVPR.2000.854767.
- [247] J. Shi, C. Tomasi, Good features to track, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR-94* (1994) 593–600.
- 2315 [248] A. Bartoli, N. Dalal, B. Bose, R. Horaud, From video sequences to motion panoramas, *Proceedings - Workshop on Motion and Video Computing, MOTION 2002* (2002) 201–207doi:10.1109/MOTION.2002.1182237.
- [249] P. Azzari, L. D. Stefano, A. Bevilacqua, An effective real-time mosaicing algorithm apt to detect motion through background subtraction using a
- 2320

- PTZ camera, IEEE International Conference on Advanced Video and Signal Based Surveillance 2005 (2005) 511–516. doi:10.1109/AVSS.2005.1577321.
- [250] N. Friedman, S. Russell, Image segmentation in video sequences: A probabilistic approach, UAI'97 Proceedings of the Thirteenth conference on  
2325 Uncertainty in artificial intelligence (1997) 175–181.
- [251] K. Xue, Y. Liu, J. Chen, Q. Li, Panoramic background model for PTZ camera, 2010 3rd International Congress on Image and Signal Processing 1 (2010) 409–413. doi:10.1109/CISP.2010.5647998.
- [252] J. Zhang, Y. Wang, J. Chen, K. Xue, A framework of surveillance system  
2330 using a PTZ camera, 2010 3rd International Conference on Computer Science and Information Technology 1 (2010) 658–662. doi:10.1109/ICCSIT.2010.5565067.
- [253] D. Avola, L. Cinque, G. Foresti, C. Massaroni, D. Pannone, A keypoint-  
2335 based method for background modeling and foreground detection using a PTZ camera, Pattern Recognition Letters 96 (2017) 96–105. doi:10.1016/j.patrec.2016.10.015.
- [254] Y. Sugaya, K. Kanatani, Extracting moving objects from a moving camera video sequence, Proceedings of the 10th Symposium on Sensing via  
2340 Imaging Information 39 (2) (2004) 279–284.
- [255] K. Kanatani, N. Ohta, Y. Kanazawa, Optimal homography computation with a reliability measure, IEICE Transactions on Information and Systems E83-D (7) (2000) 13691374.
- [256] S. Amri, W. Barhoumi, E. Zagrouba, A robust framework for joint background/foreground segmentation of complex video scenes filmed with  
2345 freely moving camera, Multimedia Tools and Applications 46 (2-3) (2010) 175–205. doi:10.1007/s11042-009-0348-y.

- 2350 [257] M. Vivet, B. Martínez, X. Binefa, Real-time motion detection for a mobile observer using multiple kernel tracking and belief propagation, *Pattern Recognition and Image Analysis* (2009) 144–151.
- [258] G. Hager, M. Dewan, C. Stewart, Multiple kernel tracking with SSD, *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004. CVPR 2004. 1 (2004) I–I. doi: 10.1109/CVPR.2004.1315112.
- 2355 [259] S. Kang, J. Paik, A. Koschan, B. Abidi, M. Abidi, Real-time video tracking using PTZ cameras, *Proceedings of the International Conference on Quality Control by Artificial Vision* 5132 (2003) 103–111. doi: 10.1117/12.514945.
- 2360 [260] S. Ali, M. Shah, Cocoa: tracking in aerial imagery, *Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications III* 6209 (2006) 6209D. doi:10.1117/12.667266.
- [261] E. Hayman, J. Eklundh, Statistical background subtraction for a mobile observer, *IEEE International Conference on Computer Vision* (2003) 67–74 vol.1doi:10.1109/ICCV.2003.1238315.
- 2365 [262] C. Stauffer, W. Grimson, Adaptive background mixture models for real-time tracking, *Proceedings 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* Cat No PR00149 2 (1999) 246–252.
- 2370 [263] B. Moore, C. Gao, R. Nadakuditi, Panoramic robust pca for foreground-background separation on noisy, free-motion camera video, *IEEE Transactions on Computational Imaging* 5 (2) (2019) 195–211.
- [264] R. Hamid, A. Sarma, D. Decoste, N. Sundaresan, Fast approximate matching of cell-phone videos for robust background subtraction, *Computing Research Repository*.

- 2375 [265] I. Chelly, V. Winter, D. Rosen, O. Freifeld, Ja-pols : a moving-camera background model via joint alignment and partially-overlapping local subspaces, *Computer Vision and Pattern Recognition*.
- [266] K. S. Bhat, M. Saptharishi, P. Khosla, Motion detection and segmentation using image mosaics, 2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing  
2380 World of Multimedia (Cat. No.00TH8532) 3 (2000) 1–5. doi:10.1109/ICME.2000.871070.
- [267] A. Bevilacqua, P. Azzari, High-quality real time motion detection using PTZ cameras, 2006 IEEE International Conference on Video and Signal  
2385 Based Surveillance (2006) 23doi:10.1109/AVSS.2006.57.
- [268] N. Liu, H. Wu, L. Lin, Hierarchical ensemble of background models for PTZ-based video surveillance, *IEEE Transactions on Cybernetics* 45 (1) (2015) 89–102. doi:10.1109/TCYB.2014.2320493.
- [269] M. Asif, J. Soraghan, Video analytics for panning camera in dynamic  
2390 surveillance environment, 2008 50th International Symposium ELMAR 1 (2008) 79–82.
- [270] C. Chen, Y. Yao, D. Page, B. Abidi, A. Koschan, M. Abidi, Heterogeneous fusion of omnidirectional and PTZ cameras for multiple object tracking, *IEEE Transactions on Circuits and Systems for Video Technology* 18 (8)  
2395 (2008) 1052–1063.
- [271] R. Horaud, D. Knossow, M. Michaelis, Camera cooperation for achieving visual attention, *Machine Vision and Applications* 16 (6) (2006) 1–2. doi:10.1007/s00138-005-0182-9.
- [272] S. Kumar, C. Micheloni, C. Piciarelli, Stereo localization using dual PTZ  
2400 cameras, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5702 LNCS (2009) 1061–1069. doi:10.1007/978-3-642-03767-2\_129.

- [273] S. Lim, A. Elgammal, L. Davis, Image-based pan-tilt camera control in a multi-camera surveillance environment, IEEE International Conference on Multimedia and Expo, ICME 2003 1 (2003) I645–I648. doi:10.1109/ICME.2003.1221000.
- [274] N. Krahnstoever, T. Yu, S. Lim, K. Patwardhan, P. Tu, Collaborative real-time control of active cameras in large scale surveillance systems, Workshop on Multicamera and Multimodal Sensor Fusion Algorithms and Applications M2SFA2 2008 (2008) 1–12.
- [275] N. Krahnstoever, P. Mendonça, Bayesian autocalibration for surveillance, Proceedings of the IEEE International Conference on Computer Vision II (2005) 1858–1865. doi:10.1109/ICCV.2005.44.
- [276] N. Krahnstoever, P. Mendonça, Autocalibration from tracks of walking people, BMVC (2006) 12.1–12.10doi:10.5244/C.20.12.
- [277] Z. Cui, A. Li, K. Jiang, Cooperative moving object segmentation using two cameras based on background subtraction and image registration, Journal of Multimedia 9 (3) (2014) 363–370. doi:10.4304/jmm.9.3.363-370.
- [278] A. Elgammal, D. Harwood, L. Davis, Non-parametric model for background subtraction, European conference on computer vision (2000) 751–767doi:10.1007/3-540-45053-X\_48.
- [279] M. Fischler, R. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM 24 (6) (1981) 381 – 395.
- [280] J. Odobez, P. Bouthemy, Separation of moving regions from background in an image sequence acquired with a mobile camera, Video Data Compression for Multimedia Computing: Statistically Based and Biologically Inspired Techniques (1997) 283–311doi:10.1007/978-1-4615-6239-9\_{8}.
- [281] R. Hartley, A. Zisserman, Multiple view geometry in computer vision.

- 2430 [282] A. Romanoni, M. Matteucci, D. Sorrenti, Background subtraction by combining temporal and spatio-temporal histograms in the presence of camera movement, *Machine Vision and Applications* 25 (6) (2014) 1573–1584.
- [283] D. Murray, A. Basu, A. Basu, Motion tracking with an active camera, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16 (5)  
2435 (1994) 449–459.
- [284] L. Robinault, S. Bres, S. Miguet, Real time foreground object detection using PTZ camera, *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications* (2009) 609–614.
- [285] Z. Kadim, M. Daud, S. Radzi, N. Samudin, H. Woon, Method to detect and track moving object in non-static PTZ camera, *Int MultiConf Eng Comput Sci* 1.  
2440
- [286] M. Wu, X. Peng, Q. Zhang, Segmenting moving objects from a freely moving camera with an effective segmentation cue, *Measurement Science and Technology* 22 (2) (2011) 025108. doi:10.1088/0957-0233/22/2/025108.  
2445
- [287] Y. Wan, X. Wang, H. Hu, Automatic moving object segmentation for freely moving cameras, *Mathematical Problems in Engineering* 2014.
- [288] F. López-Rubio, E. López-Rubio, Foreground detection for moving cameras with stochastic approximation, *Pattern Recogn. Lett.* 68 (P1) (2015) 161–168. doi:10.1016/j.patrec.2015.09.007.  
2450
- [289] L. Kurnianggoro, Y. Yu, D. Hernandez, K. Jo, Online background-subtraction with motion compensation for freely moving camera, *International Conference on Intelligent Computing* (2016) 569–578.
- [290] Y. Yu, L. Kurnianggoro, K. Jo, Moving object detection for a moving camera based on global motion compensation and adaptive background model, *International Journal of Control, Automation and Systems* 17 (7)  
2455 (2019) 1866–1874. doi:10.1007/s12555-018-0234-3.

- [291] C. Zhao, A. Sain, Y. Qu, Y. Ge, H. Hu, Background subtraction based on integration of alternative cues in freely moving camera, *IEEE Transactions on Circuits and Systems for Video Technology* 29 (7) (2019) 1933–1945. doi:10.1109/TCSVT.2018.2854273.
- [292] O. Ivanov, Adaptation of known background subtraction methods in the case of a moving ptz camera mounted on a mobile platform, *Pattern Recognition and Image Analysis* 24 (2014) 318–323.
- [293] A. Ferone, L. Maddalena, Neural background subtraction for pan-tilt-zoom cameras, *IEEE transactions on systems, man, and cybernetics: systems* 44 (5) (2014) 571–579. doi:10.1109/TSMC.2013.2280121.
- [294] D. Avola, M. Bernardi, L. Cinque, G. Foresti, C. Massaroni, Combining keypoint clustering and neural background subtraction for real-time moving object detection by PTZ cameras, *International Conference on Pattern Recognition Applications and Methods* (2018) 638–645doi:10.5220/0006722506380645.
- [295] L. Maddalena, A. Petrosino, The 3dsobs+ algorithm for moving object detection, *Computer Vision and Image Understanding* 122 (2014) 65–73. doi:10.1016/j.cviu.2013.11.006.
- [296] G. Tzanidou, P. Climent-Perez, G. Hummel, M. Schmitt, P. Stutz, D. Monekosso, P. Remagnino, Telemetry assisted frame registration and background subtraction in low-altitude UAV videos, *IEEE International Conference on in Advanced Video and Signal Based Surveillance* (2015) 1–6.
- [297] P. Torr, A. Zisserman, Feature based methods for structure and motion estimation, *International workshop on vision algorithms* (2000) 278–294doi:10.1007/3-540-44480-7\_19.
- [298] M. Irani, P. Anandan, About direct methods, *Vision Algorithms: Theory and Practice* (1999) 267–277.

- [299] B. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, *Imaging* 130 (1981) 674–679.
- [300] C. Micheloni, G. Foresti, Real-time image processing for active monitoring of wide areas, *Journal of Visual Communication and Image Representation* 17 (3) (2006) 589–604. doi:10.1016/j.jvcir.2005.08.002.
- 2490 [301] T. Kryjak, M. Komorkiewicz, M. Gorgon, Real-time implementation of foreground object detection from a moving camera using the vibe algorithm, *Computer Science and Information Systems* 11 (4) (2014) 1617–1637.
- 2495 [302] G. Allebosch, D. V. Hamme, P. Veelaert, W. Philips, Robust pan/tilt compensation for foreground–background segmentation, *Sensors* 19 (12) (2019) 2668.
- [303] L. Kurnianggoro, A. Shahbaz, K. Jo, Dense optical flow in stabilized scenes for moving object detection from a moving camera, 2016 16th International Conference on Control, Automation and Systems (ICCAS) (2016) 704–708doi:10.1109/ICCAS.2016.7832395.
- 2500 [304] T. Minematsu, H. Uchiyama, A. Shimada, H. Nagahara, R. Taniguchi, Adaptive search of background models for object detection in images, *International Conference on Image Processing (ICIP)* (2015) 3–7.
- 2505 [305] T. Minematsu, H. Uchiyama, A. Shimada, H. Nagahara, R. Taniguchi, Adaptive background model registration for moving cameras, *Pattern Recognition Letters*.
- [306] C. Guillot, M. Taron, P. Sayd, Q. Pham, C. Tilmant, J. Lavest, Background subtraction adapted to PTZ cameras by keypoint density estimation, *Proceedings of the British Machine Vision Conference 2010* (2010) 34.1–34.10doi:10.5244/C.24.34.
- 2510 [307] C. Guillot, M. Taron, P. Sayd, Q. Pham, C. Tilmant, J. Lavest, Background subtraction for PTZ cameras performing a guard tour and ap-

- 2515 application to cameras with very low frame rate, ACCV Workshops (2010)  
33–42.
- [308] N. Paragios, G. Tziritas, Adaptive detection and localization of moving objects in image sequences, *Signal Processing: Image Communication* 14 (4) (1999) 277–296. doi:10.1016/S0923-5965(98)00011-3.
- [309] Y. Ren, C. Chua, Y. Ho, Statistical background modeling for non-stationary camera, *Pattern Recognition Letters* 24 (1-3) (2003) 183–196.  
2520
- [310] S. Kim, K. Yun, K. Yi, S. Kim, J. Choi, Detection of moving objects with a moving camera using non-panoramic background model, *Machine Vision and Applications* 24 (5) (2013) 1015–1028.
- [311] A. Viswanath, R. Behera, V. Senthamilarasu, K. Kutty, Background modelling from a moving camera, *Procedia Computer Science* 58 (2015) 289–296.  
2525
- [312] F. Setyawan, J. Tan, H. Kim, S. Ishikawa, Detection of moving objects in a video captured by a moving camera using error reduction, *SICE Annual Conference* (2014) 347–352.
- [313] E. Durucan, T. Ebrahimi, Change detection and background extraction by linear algebra, *Proceedings of the IEEE* 89 (10) (2001) 1368–1381.  
2530 doi:10.1109/5.959336.
- [314] L. Gong, M. Yu, T. Gordon, Online codebook modeling based background subtraction with a moving camera, *International Conference on Frontiers of Signal Processing* (2017) 136–140doi:10.1109/ICFSP.2017.8097157.  
2535
- [315] D. Andreev, Moving objects segmentation method for flight vision systems, *IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering* (2020) 1356–1359.
- [316] O. Barnich, M. Van Droogenbroeck, ViBE: A powerful random technique to estimate the background in video sequences, *IEEE Interna-*  
2540

- tional Conference on Acoustics, Speech and Signal Processing (2009) 945–948doi:10.1109/ICASSP.2009.4959741.
- [317] O. Barnich, M. Van Droogenbroeck, ViBe: A universal background subtraction algorithm for video sequences, *IEEE Transactions on Image Processing* 20 (6) (2011) 1709–1724. doi:10.1109/TIP.2010.2101613.
- [318] A. Perera, G. Brooksby, A. Hoogs, G. Doretto, Moving object segmentation using scene understanding, 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06) (2006) 201doi:10.1109/CVPRW.2006.132.
- [319] C. Huang, Y. Wu, J. Kao, M. Shih, C. Chou, A hybrid moving object detection method for aerial images, *Advances in Multimedia Information Processing - PCM 2010* (2010) 357–368doi:10.1007/978-3-642-15702-8\_33.
- [320] S. Solehah, S. Yaakob, Z. Kadim, H. Woon, Moving object extraction in PTZ camera using the integration of background subtraction and local histogram processing, 2012 International Symposium on Computer Applications and Industrial Electronics (ISCAIE) (2012) 167–172doi:10.1109/ISCAIE.2012.6482090.
- [321] A. Elqursh, A. Elgammal, Online moving camera background subtraction, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7577 LNCS (PART 6) (2012) 228–241.
- [322] Y. Sheikh, O. Javed, T. Kanade, Background subtraction for freely moving cameras, *Proceedings of the IEEE International Conference on Computer Vision* (2009) 1219–1225.
- [323] Y. Nonaka, A. Shimada, H. Nagahara, R. Taniguchi, Real-time foreground segmentation from moving camera based on case-based trajectory classifi-

- ation, Proceedings - 2nd IAPR Asian Conference on Pattern Recognition, ACPR 2013 (2013) 808–812.
- 2570 [324] M. Berger, L. Seversky, Subspace tracking under dynamic dimensionality for online background subtraction, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2014) 1274–1281.
- 2575 [325] H. Sajid, S. Cheung, N. Jacobs, Motion and appearance based background subtraction for freely moving cameras, Signal Processing: Image Communication 75 (2019) 11–21. doi:10.1016/j.image.2019.03.003.
- [326] Y. Zhu, A. Elgammal, A multilayer-based framework for online background subtraction with freely moving cameras, Proceedings of the IEEE International Conference on Computer Vision 2017-October (2017) 5142–5151.
- 2580 [327] M. Narayana, A. Hanson, E. Learned-Miller, Coherent motion segmentation in moving camera videos using optical flow orientations, Proceedings of the IEEE International Conference on Computer Vision (2013) 1577–1584.
- 2585 [328] X. Yin, B. Wang, W. Li, Y. Liu, M. Zhang, Background subtraction for moving cameras based on trajectory-controlled segmentation and label inference, KSII Transactions on Internet and Information Systems 9 (10).
- [329] P. Bideau, E. Learned-Miller, It’s moving! a probabilistic model for causal motion segmentation in moving camera videos, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9912 LNCS (2016) 433–449.
- 2590 [330] J. Kao, D. Tian, H. Mansour, A. Vetro, A. Ortega, Moving object segmentation using depth and optical flow in car driving sequences, 2016 IEEE International Conference on Image Processing (ICIP) (2016) 11–15doi:10.1109/ICIP.2016.7532309.
- 2595

- [331] D. Sugimura, F. Teshima, T. Hamamoto, Online background subtraction with freely moving cameras using different motion boundaries, *Image and Vision Computing* doi:10.1016/j.imavis.2018.06.003.
- [332] J. Huang, W. Zou, Z. Zhu, J. Zhu, An efficient optical flow based motion  
2600 detection method for non-stationary scenes, 2019 Chinese Control And Decision Conference (CCDC) (2019) 5272–5277 doi:10.1109/CCDC.2019.8833206.
- [333] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, T. Brox, FlowNet  
2605 2.0: Evolution of optical flow estimation with deep networks, IEEE Conference on Computer Vision and Pattern Recognition (2016) 1647–1655.
- [334] W. Zhang, X. Sun, Q. Yu, Moving object detection under a moving camera via background orientation reconstruction, *Sensors* 20 (11) (2020) 3103. doi:10.3390/s20113103.
- [335] P. Tokmakov, K. Alahari, C. Schmid, Learning video object segmentation  
2610 with visual memory, ICCV doi:10.1109/ICCV.2017.480.
- [336] S. Jain, B. Xiong, K. Grauman, Fusionseg: Learning to combine motion and appearance for fully automatic segmentation of generic objects in videos, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017) 2117–2126 doi:10.1109/CVPR.2017.228.
- [337] M. Irani, P. Anandan, M. Cohen, Direct recovery of planar-parallax from  
2615 multiple frames, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (11) (2002) 1528–1534. doi:10.1109/TPAMI.2002.1046174.
- [338] M. Irani, B. Rousso, S. Peleg, Recovery of ego-motion using region alignment, *IEEE Transactions on Pattern Analysis and Machine Intelligence*  
2620 19 (3) (1997) 268–272. doi:10.1109/34.584105.
- [339] M. Irani, P. Anandan, A unified approach to moving object detection in 2D and 3D scenes, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (6) (1998) 577–589.

- [340] H. Sawhney, Y. Guo, R. Kumar, Independent motion detection in 3D scenes, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (10) (2000) 1191–1199. 2625
- [341] J. Kang, I. Cohen, Detection and tracking of moving objects from a moving platform in presence, *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1* 1 (2005) 10–17. doi:10.1109/ICCV.2005.72. 2630
- [342] T. Darrell, A. Pentland, Robust estimation of a multi-layered motion representation, *Proceedings of the IEEE Workshop on Visual Motion* (1991) 173–178doi:10.1109/WVM.1991.212810.
- [343] J. Wang, E. Adelson, Representing moving images with layers, *IEEE Transactions on Image Processing* 3 (5) (1994) 625–638. 2635
- [344] S. Ayer, H. Sawhney, Layered representation of motion video using robust maximum-likelihood estimation of mixture models and MDL encoding, *Proceedings of IEEE International Conference on Computer Vision* (1995) 777–784doi:10.1109/ICCV.1995.466859.
- [345] Y. Jin, L. Tao, H. Di, N. Rao, G. Xu, Background modeling from a free-moving camera by multi-layer homography algorithm, *Proceedings - International Conference on Image Processing, ICIP* (2008) 1572–1575. 2640
- [346] K. Patwardhan, G. Sapiro, V. Morellas, Robust foreground detection in video using pixel layers, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30 (4) (2008) 746–751. doi:10.1109/TPAMI.2007.70843. 2645
- [347] X. Zhang, S. Wang, X. Ding, Beyond dominant plane assumption: Moving objects detection in severe dynamic scenes with multi-classes RANSAC, *2012 International Conference on Audio, Language and Image Processing* (2012) 822–827doi:10.1109/ICALIP.2012.6376727. 2650

- [348] D. Zamalieva, A. Yilmaz, J. Davis, Exploiting temporal geometry for moving camera background subtraction, *International Conference on Pattern Recognition* (2014) 1200–1205.
- [349] D. Zamalieva, A. Yilmaz, J. Davis, A multi-transformational model for background subtraction with moving cameras, *Lecture Notes in Computer Science* (including subseries *Lecture Notes in Artificial Intelligence* and *Lecture Notes in Bioinformatics*) 8689 LNCS (PART 1) (2014) 803–817.
- [350] W. Hu, C. Chen, T. Chen, D. Huang, Z. Wu, Moving object detection and tracking from video captured by moving camera, *Journal of Visual Communication and Image Representation* 30 (2015) 164–180.
- [351] Y. Zhou, S. Maskell, Moving object detection using background subtraction for a moving camera with pronounced parallax, *2017 Sensor Data Fusion: Trends, Solutions, Applications (SDF)* (2017) 1–6.
- [352] S. Kim, D. Yang, H. Park, A disparity-based adaptive multihomography method for moving target detection based on global motion compensation, *IEEE Transactions on Circuits and Systems for Video Technology* 26 (8) (2016) 1407–1420.
- [353] D. Zamalieva, A. Yilmaz, Background subtraction for the moving camera: A geometric approach, *Computer Vision and Image Understanding* 127 (2014) 73–85.
- [354] T. Lim, B. Han, J. Han, Modeling and segmentation of floating foreground and background in videos, *Pattern Recognition* 45 (4) (2012) 1696–1706.
- [355] K. Yi, K. Yun, S. Kim, H. Chang, H. Jeong, J. Choi, Detection of moving objects with non-stationary cameras in 5.8ms: Bringing motion detection to your mobile device, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (2013) 27–34.

- [356] K. Yun, J. Choi, Robust and fast moving object detection in a non-stationary camera via foreground probability based sampling, 2015 IEEE International Conference on Image Processing (ICIP) (2015) 4897–4901.
- 2680 [357] W. Chung, Y. Kim, Y. Kim, D. Kim, A two-stage foreground propagation for moving object detection in a non-stationary, 2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2016 (2016) 187–193.
- 2685 [358] A. Zheng, L. Zhang, W. Zhang, C. Li, J. Tang, B. Luo, Local-to-global background modeling for moving object detection from non-static cameras, *Multimedia Tools and Applications* 76 (8) (2017) 1100311019. doi:10.1007/s11042-016-3565-1.
- 2690 [359] K. Yun, J. Lim, J. Choi, Scene conditional background update for moving object detection in a moving camera, *Pattern Recognition Letters* 88 (2017) 57–63. doi:10.1016/j.patrec.2017.01.017.
- [360] F. Sun, K. Qin, W. Sun, H. Guo, Fast background subtraction for moving cameras based on nonparametric models, *Journal of Electronic Imaging*.
- [361] Y. Wu, X. He, T. Nguyen, Moving objects detection with freely moving camera via background motion subtraction, *IEEE Transactions on Circuits and Systems for Video Technology* 27 (2) (2017) 236–248.
- 2695 [362] S. Kwak, T. Lim, W. Nam, B. Han, J. Han, Generalized background subtraction based on hybrid inference by belief propagation and Bayesian filtering, *Proceedings of the IEEE International Conference on Computer Vision* (2011) 2174–2181.
- 2700 [363] J. Lim, B. Han, Generalized background subtraction using superpixels with label integrated motion estimation, *European Conference on Computer Vision* (2014) 173–187doi:10.1007/978-3-319-10602-1\_12.

- [364] J. Kim, X. Wang, H. Wang, C. Zhu, D. Kim, Fast moving object detection with non-stationary background, *Multimedia Tools and Applications* 67 (1) (2013) 311–335. doi:10.1007/s11042-012-1075-3.
- 2705 [365] A. B. Khalifa, I. Alouani, M. Mahjoub, N. B. Amara, Pedestrian detection using a moving camera: A novel framework for foreground detection, *Cognitive Systems Research* 60 (2020) 77–96. doi:10.1016/j.cogsys.2019.12.003.
- 2710 [366] M. Unger, M. Asbach, P. Hosten, Enhanced background subtraction using global motion compensation and mosaicing, *International Conference on Image Processing* (2008) 2708–2711.
- [367] I. Kazantzidis, F. Florez-Revuelta, M. Dequidt, N. Hill, J. Nebel, Videomics: A genomics-inspired paradigm for video analysis, *Computer Vision and Image Understanding* 166 (2018) 28–40. doi:10.1016/j.cviu.2017.10.003.
- 2715 [368] K. Makino, T. Shibata, S. Yachida, T. Ogawa, K. Takahashi, Moving-object detection method for moving cameras by merging background subtraction and optical flow methods, *2017 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2017* (2017) 383–387.
- 2720 [369] D. Szolgay, J. Benois-Pineau, R. Megret, Y. Gaestel, J. Dartigues, Detection of moving foreground objects in videos with strong camera motion, *Pattern Analysis and Applications* 14 (3) (2011) 311–328. doi:10.1007/s10044-011-0221-2.
- 2725 [370] E. Komagal, B. Yogameena, Region MoG and texture descriptor-based motion segmentation under sudden illumination in continuous pan and excess zoom, *Multimedia Tools and Applications* 77 (8) (2018) 9621–9649. doi:10.1007/s11042-017-5338-x.
- [371] T. Minematsu, A. Shimada, H. Uchiyama, V. Charvillat, R. Taniguchi,

- 2730 Reconstruction-based change detection with image completion for a free-moving camera, *MDPI Sensors* 18 (4). doi:10.3390/s18041232.
- [372] V. Vezhnevets, V. Konouchine, Growcut: Interactive multi-label nd image segmentation by cellular automata, *proc. of Graphicon* 1 (4) (2005) 150–156.
- 2735 [373] R. Tron, R. Vidal, A Benchmark for the comparison of 3D motion segmentation algorithms, *IEEE Conference on Computer Vision and Pattern Recognition* (2007) 1–8doi:10.1109/CVPR.2007.382974.
- [374] P. Ochs, J. Malik, T. Brox, Segmentation of moving objects by long term video analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36 (6) (2014) 1187–1200.
- 2740 [375] T. Brox, J. Malik, Object segmentation by long term analysis of point trajectories, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6315 LNCS (PART 5) (2010) 282–295.
- 2745 [376] N. Goyette, P. Jodoin, F. Porikli, J. Konrad, P. Ishwar, Changedetection.net: A new change detection benchmark dataset, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (2012) 1–8doi:10.1109/CVPRW.2012.6238919.
- [377] Y. Wang, P. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, P. Ishwar, Cdnet 2014: An expanded change detection benchmark dataset, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (2014) 393–400doi:10.1109/CVPRW.2014.126.
- 2750 [378] F. Perazzi, J. Pont-Tuset, B. McWilliams, L. V. Gool, M. Gross, A. Sorkine-Hornung, A benchmark dataset and evaluation methodology for video object segmentation, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016) 724–732doi:10.1109/CVPR.2016.85.
- 2755

- [379] J. Pont-Tuset, F. Perazzi, S. Caelles, P. Arbeláez, A. Sorkine-Hornung, L. V. Gool, The 2017 davis challenge on video object segmentation, ArXiv abs/1704.00675.
- 2760
- [380] S. Caelles, J. Pont-Tuset, F. Perazzi, A. Montes, K. Maninis, L. V. Gool, The 2019 davis challenge on vos: Unsupervised multi-object segmentation, ArXiv abs/1905.00737.
- [381] F. Li, T. Kim, A. Humayun, D. Tsai, J. Rehg, Video segmentation by tracking many figure-ground segments, IEEE International Conference on Computer Vision (2013) 2192–2199doi:10.1109/ICCV.2013.273.
- 2765
- [382] D. Tsai, M. Flagg, A. Nakazawa, J. Rehg, Motion coherent tracking using multi-label MRF optimization, International journal of computer vision 100 (2012) 190–202. doi:10.1007/s11263-011-0512-5.