Survey

Traditional and recent approaches in background modeling for foreground detection: An overview

Thierry Bouwmans*
Lab. MIA, Univ. La Rochelle, France

ARTICLE INFO

Article history:
Received 5 March 2013
Received in revised form
20 April 2014
Accepted 20 April 2014

Keywords:
Background subtraction
Background modeling
Background initialization
Background maintenance
Foreground detection

ABSTRACT

Background modeling for foreground detection is often used in different applications to model the background and then detect the moving objects in the scene like in video surveillance. The last decade witnessed very significant publications in this field. Furthermore, several surveys can be found in the literature but none of them addresses an overall review in this field. So, the purpose of this paper is to provide a complete survey of the traditional and recent approaches. First, we categorize the different approaches found in the literature. We have classified them in terms of the mathematical models used and we have discussed them in terms of the critical situations that they claim to handle. Furthermore, we present the available resources, datasets and libraries. Then, we conclude with several promising directions for future research.

© 2014 Elsevier Inc. All rights reserved.

Contents

1. Introduction ................................................................................................................................................................................. 32

1.1. Related work ...................................................................................................................................................................... 32

1.2. Motivation and contributions ............................................................................................................................................ 33

2. Background subtraction steps....................................................................................................................................................... 34

2.1. Background modeling ........................................................................................................................................................ 34

2.2. Background initialization ................................................................................................................................................... 35

2.3. Background maintenance .................................................................................................................................................. 36

2.4. Foreground detection......................................................................................................................................................... 36

2.5. Choice of the picture's element ......................................................................................................................................... 36

2.6. Choice of the features ........................................................................................................................................................ 36

3. Applications of background subtraction ....................................................................................................................................... 37

4. Challenges and issues .................................................................................................................................................................. 37

5. Traditional background models .................................................................................................................................................... 39

* Tel.: +33 0546457202.
E-mail addresses: tbouwman@univ-lr.fr, bouwmans.thierry@gmail.com.

http://dx.doi.org/10.1016/j.cosrev.2014.04.001
1574-0137/© 2014 Elsevier Inc. All rights reserved.
1. Introduction

Analysis and understanding of video sequences is an active research field. Many applications in this research area (video surveillance [1–3], optical motion capture [4], multimedia application [5]) need in the first step to detect the moving objects in the scene. So, the basic operation needed is the separation of the moving objects called “foreground” from the static information called the “background”. The process mainly used is the background subtraction and recent surveys can be found in [6–8]. The simplest way to model the background is to acquire a background image which does not include any moving object. In some environments, the background is not available and can always be changed under critical situations like illumination changes, objects being introduced or removed from the scene. So, the background representation model must be more robust and adaptive.

Two closely related problems to background subtraction are change detection [9] and salient motion detection [10]. Change detection addresses the detection of the changes between two images. So, background subtraction is a particular case when (1) one image is the background image and the other one is the current image, and (2) the changes are due to moving objects. On the other hand, salient motion detection aims to find semantic regions and to filter out the unimportant areas. The idea of saliency detection is derived from the human visual system, where the first stage of human vision is a fast and simple pre-attentive process. So, salient motion detection can be viewed as a particular case of background subtraction.

1.1. Related work

Several surveys of background subtraction can be found in the literature but none of them address an overall review in this field. In 2000, Mc Ivor [11] surveyed 9 algorithms allowing a first comparison of the models but this survey is mainly limited on a description of the algorithms. In 2004, Piccardi [12] provided a review on 7 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 can effectively help them to select the most adapted method for a specific application. In 2005, Cheung and Kornath [1] classified several methods into non-recursive and recursive techniques. Following this classification, Elhabian et al. [6] provided a large survey in background modeling. However, this classification in term of non-recursive and recursive techniques is more adapted for the background maintenance scheme than for the background modeling one. In 2010, Cristani et al. [7] reviewed the well-known algorithms classifying them in single monocular sensor or multi-sensors but this classification is not optimal in the sense that some methods can be in the two categories. In 2010, Bouwmans et al. [8] provided a comprehensive survey on statistical background modeling methods for foreground detection classifying each approaches following the statistical models used. The updated and extended version [13] is made up from 360 papers but it covered only statistical models. So, the previous surveys made by Bouwmans et al. focus only on...
some categories [8,13] or sub-categories [14–16] of background models as follows:

- **Mixture of Gaussians models**: The background model based on Mixture of Gaussians developed by Stauffer and Grimson [17] is the most common approach. Bouwmans et al. [14] provided a survey and an original classification of the numerous improvements of the original MOG. This survey is made from 170 papers but it covered only background models based on MOG.

- **Subspace learning models**: Subspace learning methods have been used to model the background in the idea to represent online data content while reducing dimension significantly. The first method using Principal Component Analysis (PCA) was proposed by Oliver et al. [18]. Bouwmans [15] provided a survey and an original classification of these improvements. Furthermore, it presented a comparative evaluation of the variants and evaluate them with the state-of-art algorithms (SG, MOG, and KDE) by using the Wallflower dataset [19]. This survey is made from 60 papers but it covered only background models based on subspace learning methods.

- **Fuzzy models**: Critical situations met in video surveillance generate imprecision and uncertainties in the whole process of background subtraction. Therefore, some authors have recently introduced fuzzy concepts in the different steps of background subtraction. Bouwmans [20] provided a comprehensive survey on fuzzy concepts used in background subtraction steps. This survey is made from 50 papers but it covered only background models based on fuzzy concepts.

- **Robust PCA models**: Robust Principal Components Analysis (RPCA) models have been recently developed in the literature [21]. Recently, Bouwmans and Zahzah [16] initiated a comprehensive review of RPCA-PCP based methods for testing and ranking existing algorithms for foreground detection. This survey is made from 70 papers but it covered only background models based on RPCA.

Considering all of this, we present a first complete overview of all the background models with a classification of them following the mathematical models used. This classification is better adapted to this field as shown in the survey [13] on recent advanced statistical background models. For more details on a category or a sub-category of a model, the reader is invited to read the corresponding surveys.

### 1.2. Motivation and contributions

The last decade witnessed very significant publications on background subtraction and recently new applications in which the background is not static, such as recordings taken from mobile devices or Internet videos, generate new developments to detect robustly moving objects in challenging environments. Thus, effective methods for robustness to deal both with dynamic backgrounds, illumination changes in real scene with fixed cameras or mobile devices have been recently developed, and so different strategies are used such as automatic feature selection, model selection or hierarchical models. Another feature of background modeling methods is that the use of advanced models has to be computed in real-time and low memory requirements. Algorithms are redesigned to meet these requirements.

In this context, the aim of this survey is then to provide a first complete overview of all the background models for (1) novices who could be students or engineers beginning in the field of computer vision or biologists who need to detect animals in videos for ethology, (2) experts as we put forward the recent advances that need to be improved, and (3) reviewers to evaluate papers in journals, conferences (AVSS, etc.) and workshops (ChangeDetection.net, BMC). So, this survey is intended to be a reference for researchers and developers in industries, as well as graduate students, interested in background modeling and foreground detection applied to video surveillance and other related areas, such as optical motion capture, multimedia applications, teleconferencing, video editing, human–computer interface. It can also be suggested as a reading text to teach graduate courses in subjects such as computer vision, image processing, real-time architecture, machine learning and data mining.

Although there are numerous works on background subtraction and foreground detection and the fact that it can gives the feeling that this issue is solved or is over-studied, no traditional algorithm today seems to be able to simultaneously address all the key challenges that we can meet in videos. This is due to three main reasons:

1. The lack of common framework, that is, each method was developed in several different contexts (video-surveillance, optical motion capture, ...) under their different challenges. The different steps and key challenges were rarely well identified, except by Toyama et al. [19] in 1999 for the case of video surveillance applications.

2. The lack of scientific progress, that is, there are mainly improvements of the state-of-art methods such as Mixture of Gaussians [17], and other recent representation models are sometimes investigated insufficiently.

3. The absence of a single realistic large-scale dataset with accurate ground-truth providing a balanced coverage of the range of challenges present in the real world.

These three problems have been recently addressed by (1) some recent surveys on the field, (2) the application of other mathematical tools, and (3) the creation of large datasets such as ChangeDetection.net [22], SABS [23] and BMC [24]. The community has recently gathered together around special events such as (1) the ChangeDetection.net (CDW 2012, CDW 2014) and the BMC 2012 workshops, and two special issues in journals [25,26] to solve these key issues.

In this context, we believe that we are living in a key transition in the field of background subtraction and the aim of this survey is to review these key new points in background modeling and foreground detection. Thus, it reviews all the models since the first works in the field to the recent ones. By reviewing both existing and new ideas, this survey gives a complete overview of the concepts, theories, algorithms, and applications related to background modeling and foreground detection. First, an introduction to background modeling and foreground detection for beginners is provided. For this, the different steps and components in background subtraction
are presented as several applications in which background subtraction is required to detect moving objects. Then, we review the different challenges and which steps and issues are concerned to deal with them. Furthermore, existing and recent methods for detecting moving objects are presented. A description of recent complete datasets and codes are given. Moreover, an accompanying website called the Background Subtraction Web Site\(^1\) is provided. It allows the reader to have a quick access to the main resources, datasets and codes in the field. Finally, with this survey, we aim to bring a one-stop solution, i.e., access to a number of different models, algorithms, implementations and benchmarking techniques in a single paper.

Thus, we present the different models in two main categories to strengthen this transition:

- **Traditional models** which present the following characteristics: (1) These models are the first models used in the field and are generally basics. (2) These models allow us to handle some specific challenges, and are generally easy to implement. (3) Numerous improvements concern these models and their limitations seem to be reached.

- **Recent models** which present on the other hand the following characteristics: (1) These models are more sophisticated to robustly handle a lot of challenges. (2) Most of them need improvements to achieve incremental and real-time requirements.

Contributions of this paper can be summarized as follows:

- A first complete overview of background models over the last decade concerning more than 300 papers. The traditional and recent models are presented in Section 5 and the Section 6, respectively. Table 1 shows the complete overview of this survey. The previous surveys are indicated in bold and the reader can refer to them for more references on the corresponding category or sub-category. Furthermore, we present the available resources such as the datasets and libraries which enable easy comparison and evaluation of background subtraction algorithms. Moreover, this paper is provided with an accompanying web site: the Background Subtraction Web Site. This website contains a full list of the references in the field, links to datasets and codes. In each case, the list is regularly updated and classified following the background models as in this paper.

\(^1\)http://sites.google.com/site/backgroundsubtraction/Home.

---

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

The rest of this paper is organized as follows: In Section 2, the different steps and components in background subtraction are presented. In Section 3, we present several applications in which background subtraction is required to detect moving objects. In Section 4, we review the different challenges met in video surveillance, and which steps and issues are involved to deal with them. Then, in Section 5, we give an overview of the traditional background models classified following the mathematical tools. In Section 6, we investigate the recent background models. In Section 7, we present resources, datasets and codes publicly available to test and evaluate algorithms. Finally, we provide a conclusion and perspectives for future research.

## 2. Background subtraction steps

Fig. 1 shows an overview of the background subtraction steps: (1) Background initialization using \(N\) frames to obtain the first background image without the moving objects. (2) Then, the moving object detection is made through the foreground detection that consists in classifying pixels as foreground or background by comparing the background image and the current frame. (3) Background maintenance to update the background image over time. The steps (2) and (3) are executed repeatedly as time progresses. These different steps are developed in the following sub-sections.

### 2.1. Background modeling

Background modeling which describes the kind of model used to represent the background. It determines mainly the ability of the model to deal with uni-modal or multi-modal backgrounds.
2.2. Background initialization

Background initialization which refers to the initialization of the model. In contrast to the background model representation and the model maintenance, the initialization of the background model was only marginally investigated. The main reason is that it is often assumed that initialization can be achieved by exploiting some clean frames at the beginning of the sequence. Naturally, this assumption is rarely encountered in real-life scenarios, because of continuous clutter presence. Generally, the model is initialized using the first frame or a background model over a set of training frames, which contains or do not contain foreground objects. The main challenge is to obtain a first background model when more than half of the training contains foreground objects. Practically, some algorithms...
are: (1) batch ones using N training frames (consecutive or not) [132], (2) incremental with known N or (3) progressive ones with unknown N as the process generates partial backgrounds and continues until a complete background image is obtained [133,134]. Furthermore, initialization algorithms depend on the number of modes and the complexity of their background models [135].

### 2.3. Background maintenance

Background maintenance which relies on the mechanism used for adapting the model to the changes occurred in the scene over time. This learning process has to be achieved online and so the algorithm has to be an incremental one. The key issues of this step are the following ones:

- **Maintenance schemes**: In the literature, three maintenance schemes are present: the blind, the selective, and the fuzzy adaptive schemes [136]. The blind background maintenance updates all the pixels with the same rules which is usually an IIR filter:

\[
B_{t+1}(x, y) = (1 - \alpha) B_t(x, y) + \alpha I_t(x, y)
\]

where \( \alpha \) is the learning rate which is a constant in \([0, 1]\).

- **Maintenance mechanisms**: The learning rate determines the speed of the adaptation to the scene changes. It can be (1) fixed, or dynamically adjusted by (2) a statistical or (3) a fuzzy method [137–140].

- **Learning rate**: The learning rate determines the speed of the foreground detection such as in [136]. Furthermore, initialization of the foreground detection may result as background very quickly and a pixel classified as foreground very slowly. For this reason, \( \beta \ll \alpha \) and usually \( \beta = 0 \). So the Eq. (3) becomes:

\[
B_{t+1}(x, y) = B_t(x, y)
\]

But 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 graduateing the update rule using the result of the foreground detection such as in [136].

- **Frequency of the update**: The aim is to update the background only when 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 [143,144].

### 2.4. Foreground detection

Foreground detection which consists of labeling pixels as background or as foreground pixels. This task is a classification task.

### 2.5. Choice of the picture’s element

Choice of the picture’s element which is used in the previous steps. This element may be a pixel [17], a block [145] or a cluster [146]. The size of the element determines the robustness to noise, and the precision. A pixel-based method gives a pixel-based precision but it is less robust to noise than block-based or cluster based methods.

### 2.6. Choice of the features

In the literature, there are five features commonly used: color features, edge features, stereo features, motion features and texture features. In [147], these features are classified as spectral features (color features), spatial features (edge features, texture features) and temporal features (motion features). These features have different properties which allow us to handle the critical situations differently (illumination changes, motion changes, structure background changes). Color features are often very discriminative but they have several limitations in the presence of illumination changes, camouflage and shadows. The addition of other features allows us to be more robust in these challenges. For example, stereo features allow us to alleviate the camouflage in color but two cameras are needed. Edge features handle the local illumination changes and the ghost left when waking foreground objects begin to move. Texture is adapted to the illumination changes and shadows. If more than one feature is used, the operator of their combination has to be determined. It may be a crisp operator, a statistical operator or a fuzzy operator [148,149].

While developing a background subtraction method, researchers must design each step and choose the features according to the critical situations [19] that they want to handle: Noise image due to a poor quality image source, camera jitter, camera automatic adjustments, time of the day, light switch, bootstrapping, camouflage, foreground aperture, moved background objects, inserted background objects, multi-modal background, waking foreground object, sleeping foreground object and shadows. These critical situations...
have different spatial and temporal properties. The main difficulties come from the illumination changes and dynamic backgrounds (See Section 4). For example, Fig. 2 shows the original frame 309 of the sequence from [37], the generated background, the ground-truth and the foreground mask. We can see that several false positive detections are generated by the camera jitter.

3. Applications of background subtraction

Segmentation of moving foreground objects from video stream is the fundamental step in many computer vision applications, such as:

- **Intelligent visual surveillance**: This is the main application of background modeling and foreground detection. The goal is to detect moving objects or abandoned objects to assure the security of the concerned area, or to compute statistics on the traffic such as in road [150], airport [151] or maritime surveillance [152]. The objects of interest are very different such as vehicles, airplanes, boats, people and baggages. Surveillance can be more specific such as studying consumer behavior in stores [153–155].

- **Intelligent visual observation of animals and insects**: Surveillance can also concern animal activities in protected areas (river, ocean, etc.) or zoos for ethology. The objects of interest are then animals such as birds [156–158], fish [159], honeybees [160–162] or hinds [163,164].

- **Optical motion capture**: The aim is to obtain a full and precise capture of a human with cameras [165]. The silhouette is generally extracted in each view by background subtraction. Then, the visual hull is obtained in three dimensions.

- **Human–machine interaction**: Several applications need interactions between human and machine through a video acquired in real-time by fixed cameras such as games (Microsoft’s Kinect) and ludo-applications such as Aqu@theque [4].

- **Content based video coding**: To generate the video content, the video has to be segmented into video objects and tracked as they transverse across the video frames. The registered background and the video objects are then encoded separately. So, video coding needs effective methods for object detection from static and dynamic environments [166,167].

4. Challenges and issues

There are three main conditions which insure a good functioning of the background subtraction methods: the
Table 2 – Applications of background modeling and foreground detection: An Overview.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Sub-categories</th>
<th>Objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intelligent Visual Surveillance of Human Activities</td>
<td>Road Surveillance</td>
<td>Cars, people [150]</td>
</tr>
<tr>
<td></td>
<td>Airport Surveillance</td>
<td>Airplanes, people [151]</td>
</tr>
<tr>
<td></td>
<td>Maritime Surveillance</td>
<td>Boats, people [152]</td>
</tr>
<tr>
<td></td>
<td>Store Surveillance</td>
<td>People [153–155]</td>
</tr>
<tr>
<td>Intelligent Visual Observation of Animal and Insect Behaviors</td>
<td>Birds Surveillance</td>
<td>Birds [156–158]</td>
</tr>
<tr>
<td></td>
<td>Insect Surveillance</td>
<td>Honeybees [160–162]</td>
</tr>
<tr>
<td></td>
<td>Fish Surveillance</td>
<td>Fish [159]</td>
</tr>
<tr>
<td>Intelligent Visual Observation of Natural Environments</td>
<td>River Surveillance</td>
<td>Woods [178,179]</td>
</tr>
<tr>
<td>Optical Motion Capture</td>
<td>Visual Hull</td>
<td>People [165]</td>
</tr>
<tr>
<td>Human–Machine Interaction</td>
<td>Games</td>
<td>People</td>
</tr>
<tr>
<td></td>
<td>Ludo-applications</td>
<td>People, Animals [4]</td>
</tr>
<tr>
<td>Content based Video Coding</td>
<td>Video Content</td>
<td>Objects [166,167]</td>
</tr>
</tbody>
</table>

Fig. 4 – The first row presents original frames in animals surveillance: Birds [156], Honeybees [161] and Fish [5], the second row shows the corresponding ground truth (GT) or segmentation results.

camera is fixed, the illumination is constant and the background is static (that is pixels have a unimodal distribution and no background objects are moved or inserted in the scene). In these ideal conditions, background subtraction gives good results. In practice, some critical situations may appear and disturb this process. In 1999, Toyama et al. [19] identified 10 challenging situations in the field of video surveillance. In this paper, we extend this list to 13 which are the following:

- **Noisy image:** It is due to a poor quality image source such as images acquired by a web cam or images after compression.
- **Camera jitter:** In some conditions, the wind may cause the camera to sway back and forth. So, it causes nominal motion in the sequence. Foreground mask shows false detections due to the motion without a robust maintenance mechanism.
- **Camera automatic adjustments:** Many modern cameras have auto focus, automatic gain control, automatic white balance and auto brightness control. These adjustments modify the dynamic of the color levels between different frames in the sequence.
- **Illumination changes:** They can be gradual such as ones in a day in an outdoor scene or sudden such as a light switch in an indoor scene. Fig. 5 shows an indoor environment which presents a gradual illumination change. It causes false detections in several parts of the foreground mask as can be seen in Fig. 5(d). Fig. 6 shows the case of a sudden illumination change due to a light “on/off”. As all the pixels are affected by this change, a big amount of false detections is generated (see Fig. 6(c)).
- **Bootstrapping:** During the training period, the background is not available in some environments. Then, it is impossible to compute a representative background image.
- **Camouflage:** A foreground objects pixel characteristics may be subsumed by the modeled background. Then, the foreground and the background cannot be distinguished.
- **Foreground aperture:** When a moved object has uniform colored regions, changes inside these regions may not be detected. Thus, the entire object may not appear as foreground. Foreground masks contain false negative detections.
- **Moved background objects:** Background objects can be moved. These objects should not be considered part of the foreground. Generally, both the initial and the new position of the objects are detected without a robust maintenance mechanism.
- **Inserted background objects:** A new background object can be inserted. This object should not be considered part of the foreground. Generally, the inserted background object is detected without a robust maintenance mechanism.
- **Dynamic backgrounds:** Backgrounds can vacillate and this requires models which can represent disjoint sets of pixel values. Fig. 7 shows three main types of dynamic backgrounds: waving trees, water rippling and water surface. In each case, there is a big amount of false detections.
- **Beginning moving object:** When an object initially in the background moves, both it and the newly revealed parts of the background called “ghost” are detected.
- **Sleeping foreground object:** Foreground object that becomes motionless cannot be distinguished from a background object and then it will be incorporated in the background. How to manage this situation depends
Fig. 5 – From the left to the right: (a) The first image presents an indoor scene with low illumination. (b) The second image presents the same scene with moderate illumination while the third image (c) shows the scene with high illumination. (d) The fourth image shows the foreground mask obtained with MOG [17]. This sequence called “Time of Day” comes from the Wallflower dataset [19].

(a) Low Illum.  (b) Moderate Illum.  (c) High Illum.  (d) Foreground mask.

Fig. 6 – From the left to the right: (a) The first image presents an indoor scene with light-on. (b) The second image shows the same scene with light-off. (c) The third image shows the foreground mask obtained with MOG [17]. This sequence called “Light Switch” comes from the Wallflower dataset [19].

(a) Light-on.  (b) Light-off.  (c) Foreground mask.

Fig. 7 – The first row presents original scenes containing dynamic backgrounds: (a) Boats Sequence, (b) Fountain Sequence and (c) Overpass Sequence. These three sequences come from the ChangeDetection.net dataset [22]. The second row shows the corresponding foreground masks obtained by the MOG [17].

(a) Water Surface.  (b) Fountain.  (c) Waving trees.

These critical situations have different spatial and temporal properties. Table 3 gives an overview of which steps and issues are involved to deal with them. The first column indicates the challenges and the second column shows the involved steps or issues with the corresponding solutions. The reader is invited to read the following sections for the meaning of each acronym. Furthermore, a similar table is available for the MOG improvements in [14] and for the KDE improvements in [13].

5. Traditional background models

The different background representation models can be classified in the following categories: basic models, statistical models, cluster models, neural network models and estimation models.

5.1. Basic models

In this case, the background is modeled using an average [27], a median [28] or an histogram analysis over time [29]. Once the model is computed, pixels of the current image are classified as foreground by thresholding the difference between the background image and the current frame as follows:

\[ d \left( I(x, y), B_{t-1}(x, y) \right) > T \]  

(5)

on the context. Indeed, in some applications, motionless foreground objects must be incorporated [180] and in others it is not the case [181].

- Shadows: Shadows can be detected as foreground and can come from background objects or moving objects (See Fig. 8). Shadow detection is a research field in itself. Complete studies and surveys can be found in [182–186].
Table 3 – Challenges and Solutions. The first column indicates the challenges and the second column shows the concerned steps or issues with the corresponding solutions. The reader is invited to read the following sections for the meaning of each acronym.

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noisy image</td>
<td>Clusters Models: K-means [47], Codebook [48,187], BSC [49]</td>
</tr>
<tr>
<td></td>
<td>Features: Blocks [145], Clusters [146]</td>
</tr>
<tr>
<td>Camera jitter</td>
<td>Statistical Models: MOG [17,188–190], KDE [191–193], SVR [39], SVDD [40]</td>
</tr>
<tr>
<td></td>
<td>Advanced Statistical Models: DMM [64], KGHMM [67], Vibe [69], PBAS [70]</td>
</tr>
<tr>
<td></td>
<td>Fuzzy Models: T2-FMOG [75]</td>
</tr>
<tr>
<td>Camera automatic adjustments</td>
<td>Background Maintenance: MOG [195]</td>
</tr>
<tr>
<td></td>
<td>Features: Edges [196]</td>
</tr>
<tr>
<td>Illumination changes</td>
<td>Subspace Learning Models: PCA [18], ICA [197], INMF [42], IRT [44]</td>
</tr>
<tr>
<td></td>
<td>Filter Models: Wiener Filter [19], Kalman Filter [198], Chebychev Filter [62]</td>
</tr>
<tr>
<td></td>
<td>RPCA Models: RPCA-PCP [21], RPCA-IRLS [94], BRPCA [96], GoDec [98], GRSTA [104]</td>
</tr>
<tr>
<td></td>
<td>Discriminative and Mixed Subspace Models: IMMC [76], PCA–ILDA [77]</td>
</tr>
<tr>
<td></td>
<td>Sparse Models: SS [114], DGS [115], DRL [199], SEE [200]</td>
</tr>
<tr>
<td></td>
<td>Features: Textures (SURF [201])</td>
</tr>
<tr>
<td>Bootstrapping</td>
<td>Background Initialization: SVM [38], Consecutive Frames [202]</td>
</tr>
<tr>
<td></td>
<td>Cluster Models: K-means [47], Codebook [48,187], BSC [49]</td>
</tr>
<tr>
<td></td>
<td>Features: Blocks [202,203]</td>
</tr>
<tr>
<td></td>
<td>Strategies: Markov Random Fields (MRF) [204]</td>
</tr>
<tr>
<td>Camouflage</td>
<td>Features: Disparity [205], Depth [206]</td>
</tr>
<tr>
<td>Foreground aperture</td>
<td>Background Maintenance: MOG [207]</td>
</tr>
<tr>
<td>Moved background objects</td>
<td>Background Maintenance: MOG [208–210]</td>
</tr>
<tr>
<td>Inserted background objects</td>
<td>Background Maintenance: MOG [208–210]</td>
</tr>
<tr>
<td>Dynamic backgrounds</td>
<td>Statistical Models: MOG [17,188–190], KDE [37,211,212,193], SVR [39], SVDD [40,213]</td>
</tr>
<tr>
<td></td>
<td>Advanced Statistical Models: DMM [64], KGHMM [67], Vibe [69], PBAS [70]</td>
</tr>
<tr>
<td></td>
<td>Fuzzy Models: T2-FMOG [75]</td>
</tr>
<tr>
<td></td>
<td>Domain Transform Models: Waviz [127], Wave-Back [128]</td>
</tr>
<tr>
<td></td>
<td>Features: Texture (LBP [214], STLBP [215], ELBP [216], SALBP [217], SILTP [218])</td>
</tr>
<tr>
<td></td>
<td>Features: Histogram (LDH [219]), Pattern (LDP [220], SDLP [221])</td>
</tr>
<tr>
<td>Beginning moving object</td>
<td>Background Maintenance: MOG [222,223]</td>
</tr>
<tr>
<td>Sleeping foreground object</td>
<td>Background Maintenance: MOG [222,223]</td>
</tr>
<tr>
<td>Shadows</td>
<td>Features: Colors, edges, textures [182–186,224,225]</td>
</tr>
<tr>
<td></td>
<td>Strategies: Markov Random Fields (MRF) [225]</td>
</tr>
</tbody>
</table>

Otherwise, pixels are classified as background. $T$ is a constant threshold, $I_t(x, y)$, $B_{t-1}(x, y)$ are respectively the current image at time $t$ and the background image at time $t - 1$. $d(\cdot, \cdot)$ is a distance measure which is usually the absolute difference between the current and the background images.

5.2. Statistical models

The statistical models offer more robustness to illumination changes and dynamic backgrounds [8,13]. The statistical background models can be classified in the following categories as in the surveys [8,13]: Gaussian models, support vector models and subspace learning models.

- **Gaussian models**: The simplest way to represent the background is to assume that the history over time of pixel’s intensity values can be modeled by a Gaussian. Following this idea, Wren et al. [34] have proposed to use a single Gaussian (SG). Kim et al. [35] generalized the SG using single general Gaussian (SGG) to alleviate the constraint of a strict Gaussian. However, a unimodal model cannot handle dynamic backgrounds when there are waving trees, water rippling or moving algae. To solve this problem, the Mixture of Gaussians (MOG) or Gaussian Mixture Model (GMM) [17] has been used to model dynamic backgrounds. Many improvements of the MOG were developed to be more robust and adaptive to the critical situations and a good survey can be found in the one made by Bouwmans et al. [14]. For example, Porikli and Tuzel [226] defined each pixel as layers of 3D multivariate Gaussians. Each layer corresponds to a different appearance of the pixel. Using a Bayesian
Fig. 8 – The first row presents original scenes containing shadows: (a) Backdoor, (b) Bungalows and (c) PeopleInShade. These sequences come from the ChangeDetection.net dataset [22]. The second row shows the corresponding foreground masks obtained by the MOG [17].

approach, the probability distributions of mean and variance are estimated instead of the mean and variance themselves. Another improvement developed by Alvar et al. [227] used a Real-Time Dynamic Ellipsoidal Neural Networks (RTDENN) to improve the MOG maintenance. In another way, Allili et al. [36] proposed the mixture of general Gaussians (MOGG) to alleviate the constraint of a strict Gaussian. However, the MOG and MOGG present several disadvantages. For example, a background having fast variations cannot be accurately modeled with just a few Gaussians (usually 3 to 5), which causes problems for sensitive detection. So, a non-parametric technique [37] was developed to estimate background probabilities at each pixel from many recent samples over time using Kernel density estimation (KDE) but it is time consuming. Several improvements of the KDE can be found in the surveys [8,13]. For example, Sheikh and Shah [191,192] modeled the background using a KDE method over a joint domain-range representation of image pixels. So, multi-modal spatial uncertainties and complex dependencies between the domain and range are directly modeled. Furthermore, the background and foreground models are used competitively in a MAP-MRF (Maximum A Posteriori Markov Random Field) [192] decision framework. A sequential KDE algorithm is presented in [228].

- **Support vector models:** The second category uses more sophisticated statistical models such as support vector machine (SVM) [38], support vector regression (SVR) [39] and support vector data description (SVDD) [40]. First, Lin et al. [38,229] proposed to initialize the background using a probabilistic Support Vector Machine (SVM). SVM classification is applied for all pixels of each training frame by computing the output probabilities. Newly found pixels are evaluated and determined if they should be added to the background model. The background initialization continues until there are no more new background pixels to be considered. The features used are the optical flow value and inter-frame difference. In a similar way, Wang et al. [39,230] used a separate SVR to model each background pixel as a function of the intensity. The background initialization is made using a batch algorithm and the background maintenance is achieved with an online SVR algorithm [231]. For the foreground detection, they fed its intensity value to the SVR model associated with the pixel and they thresholded the output of the SVR. In another way, Tavakkoli et al. [40,232] proposed to label pixels in video sequences into foreground and background classes using Support Vector Data Description (SVDD). Unlike parametric and non-parametric density estimation techniques, the background model is not based on the probability function of the background or foreground. Indeed, it is an analytical description of the decision boundary between background and foreground classes. So, the model accuracy is not bounded to the accuracy of the estimated probability density functions. Therefore, the memory requirements are less than those of non-parametric techniques as, in non-parametric density estimation methods, pixel feature vectors for all background training frames need to be stored to regenerate the probability of pixels in new frames. For the background initialization, the process is made off-line and the background maintenance is made using an on-line algorithm. For the foreground detection, pixels are only compared with the support vectors, which are practically much fewer than the number of frames in the temporal window. Furthermore, SVDD explicitly models the decision boundary of the known class. So, this results in less parameter tuning and automatic classification. The background initialization was improved by a genetic approach [233] and the background maintenance is made by an incremental SVDD algorithm [234,235].

- **Subspace learning models:** The third category employs subspace learning methods. Subspace learning using Principal Component Analysis (SL-PCA) [18] is applied on N images to construct a background model, which is represented by the mean image and the projection matrix comprising the first p significant eigenvectors of PCA. In this way, foreground segmentation is accomplished by computing the difference between the input image and its reconstruction. However, this model presents several limitations as developed in the survey made by Bouwmans [15]: (1) The size of the foreground objects must be small and they should not appear in the same location during a long period in the training sequence. Some authors [236–238] partially alleviate these constraints; (2) For the background maintenance, it is computationally intensive to perform model updating using the batch mode PCA. Moreover without a mechanism of robust analysis, the outliers or foreground objects may be absorbed into the background model. Some incremental mechanisms robust to outliers have been developed in [239–241]; (3) The application of this model is mostly limited to gray-scale images and pixel-wise aspects since the integration of multi-channel data is not straightforward. It involves a much higher dimensional space and generally causes additional difficulty to manage data. Recently, Han and Jain [242] proposed an efficient algorithm using a weighted incremental 2-Dimensional Principal Component Analysis. The proposed algorithm was applied to 3-channel (RGB) and 4-channel (RGB+IR) data. Results...
Table 4 – Statistical Models: An Overview. The first column indicates the category model and the second column the name of each method. Their corresponding acronym is indicated in the first parenthesis and the number of papers counted for each method in the second parenthesis. The third column gives the name of the authors and the date of the related publication.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Methods</th>
<th>Authors - Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian Models</td>
<td>Single Gaussian (SG) (34)</td>
<td>Wren et al. (1997) [34]</td>
</tr>
<tr>
<td></td>
<td>Single General Gaussian (SGG) (3)</td>
<td>Kim et al. (2007) [35]</td>
</tr>
<tr>
<td></td>
<td>Mixture of Gaussians (MOG) (250)</td>
<td>Stauffer and Grimson (1999) [17]</td>
</tr>
<tr>
<td></td>
<td>Mixture of General Gaussians (MOGG) (3)</td>
<td>Allili et al. (2007) [36]</td>
</tr>
<tr>
<td></td>
<td>Kernel Density Estimation (KDE) (70)</td>
<td>Elgammal et al. (2000) [37]</td>
</tr>
<tr>
<td>Support Vector Models</td>
<td>Support Vector Machine (SVM) (11)</td>
<td>Lin et al. (2002) [38]</td>
</tr>
<tr>
<td></td>
<td>Support Vector Regression (SVR) (3)</td>
<td>Wang et al. (2006) [39]</td>
</tr>
<tr>
<td></td>
<td>Support Vector Data Description (SVDD) (6)</td>
<td>Tavakkoli et al. (2006) [40]</td>
</tr>
<tr>
<td>Subspace Learning Models</td>
<td>1) Matrices</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Principal Components Analysis (PCA) (30)</td>
<td>Oliver et al. (1999) [18]</td>
</tr>
<tr>
<td></td>
<td>Independent Component Analysis (ICA) (6)</td>
<td>Yamazaki et al. (2006) [41]</td>
</tr>
<tr>
<td></td>
<td>Incremental Non Negative Matrix Factorization (INMF) (3)</td>
<td>Bucak et al. (2007) [42]</td>
</tr>
<tr>
<td></td>
<td>Locally Preserving Projections (LoPP) (1)</td>
<td>Krishna et al. (2012) [43]</td>
</tr>
<tr>
<td></td>
<td>2) Tensors</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Incremental Rank-(R1,R2,R3) Tensor (IRT) (2)</td>
<td>Li et al. (2008) [44]</td>
</tr>
<tr>
<td></td>
<td>Tensor Locally Preserving Projections (1)</td>
<td>Krishna et al. (2013) [45]</td>
</tr>
<tr>
<td></td>
<td>Diffusion Bases (1)</td>
<td>Dushnik et al. (2013) [46]</td>
</tr>
</tbody>
</table>

show noticeable improvements in the presence of multi-modal background and shadows; (4) The representation is not multi-modal so various illumination changes cannot be handled correctly. Recently, Dong et al. [243] proposed to use a multi-subspace learning to handle different illumination changes. The feature space is organized into clusters which represent the different lighting conditions. A Local Principle Component Analysis (LPCA)\(^2\) transformation is used to separately learn an eigen-subspace for each cluster. When a current image arrives, the algorithm selects the learned subspace which shares the nearest lighting condition. The results [243] show that the LPCA algorithm outperforms the original PCA especially under sudden illumination changes.

Recently, other reconstructive subspace models were used. Yamazaki et al. [41] and Tsai et al. [197] used an Independent Component Analysis (ICA). In another way, Bucak et al. [42] proposed an Incremental Non-negative Matrix Factorization (INMF) to reduce the dimension. In order to take into account the spatial information, Li et al. [44] used an Incremental Rank-(R1,R2,R3) Tensor (IRT). Recently, Krishna et al. [43] used the Locality Preserving Projections (LoPP) also known as Laplacian eigenmaps. LoPP is a classical linear technique that projects the data along the directions of maximal variance and optimally preserves the neighborhood structure of the observation matrix. Furthermore, LoPP shares many of the data representation properties of nonlinear techniques, which are interesting to separate background and foreground. A complete survey on reconstructive subspace learning models used in background modeling and foreground detection can be found [15].

Table 4 shows an overview of the statistical background modeling methods. The Gaussian models and support vector models are greatly designed for dynamics backgrounds and subspace learning models for illumination changes. The statistical models are the most used due to a good compromise between their performance and their computation cost.

5.3. Cluster models

Cluster models suppose that each pixel in the frame can be temporally represented by clusters. The clustering approaches consist of K-mean algorithms [47], Codebooks [48] or basic sequential clustering methods [49].

- **K-means models**: Butler et al. [47] proposed an algorithm that assigns a group of clusters to each pixel in the frame. The background initialization is achieved off-line. The clusters are ordered according to the likelihood that they model the background and are adapted to deal with background and lighting variations. Incoming pixels are matched against the corresponding cluster group and are classified according to whether the matching cluster is considered part of the background or not. To improve the robustness, Duan et al. [244] proposed to use a genetic K-means algorithm. The idea is to alleviate the disadvantages of the traditional K-means algorithm which has random and locality aspects causing lack of global optimization.

- **Codebook models**: Kim et al. [48] proposed to model the background using a codebook model. For each pixel, a codebook is constructed and consists of one or more codewords. Samples at each pixel are clustered into a set of codewords based on a color distortion metric together with brightness bounds. The number of codewords is different following the pixel’s activities. The clusters represented by codewords do not necessarily correspond to single Gaussian or other parametric distributions. The background is encoded on a pixel-by-pixel basis.

\(^2\) [http://vigir.missouri.edu/BackgroundSub](http://vigir.missouri.edu/BackgroundSub).
Detection involves testing the difference of the current image from the background model with respect to color and brightness differences. If an incoming pixel verifies (1) the color distortion to some codewords is less than the detection threshold, and (2) its brightness lies within the brightness range of that codeword, it is classified as background. Otherwise, it is classified as foreground. This original algorithm has been improved in several ways. For example, Kim et al. [245] presented a modified algorithm which a layered modeling/detection and adaptive codebook updating. Sigari and Fatih [246] proposed a two-layer codebook model. The first layer is the main codebook, the second one is the cache codebook, and both contain some codewords relative to a pixel. Main codebook models the current background images and cache codebook is used to model new background images during input sequence. To be robust to illumination changes, other improvements concerned the color models which can be used instead of the cone cylinder model such as the hybrid-cone cylinder model in [247], and the spherical model in [248]. Other modifications concerned block based approach [249,250], hierarchical approach [251] or multi-scale approach [252] to reach real-time requirements.

- **Basic sequential clustering:** Another clustering approach was developed in [49] based on the assumption that the background would not be the parts which appear in the sequence for a short time. First, pixels intensities are classified based with an on-line clustering model. Then, cluster centers and appearance probabilities of each cluster are calculated. Finally, a single or multi intensities clusters with the appearance probability greater than a threshold are selected as the background pixel intensity value. A modified version proposed in [253] adds in the second step a merging procedure to classify classes. When a cluster deviates to another and becomes very close, the two clusters are fused in one cluster. Xiao and Zhang [254] improved this approach with a two-threshold sequential clustering algorithm. The avoidance of quick deviation of clusters to themselves is improved by the second threshold on the creation of a new cluster. Recently, Benalia and Ait-Aoudia [255] proposed to address the problem of cluster deviation without the use of a margin procedure or the two thresholds. The algorithm consists in saving the first value of the cluster, when it is created, in another cluster center. Then, the current cluster value is compared to its past value after every updating operation of the cluster value for deviation control. If the deviation is important and bigger than a threshold, a new cluster is created from the past one and takes the weight of the current one. To optimize the used memory, the old clusters with no updating are deleted based on the assumption that a background cluster is updated with a large time frequency.

The cluster models seem well adapted to deal with dynamic backgrounds and noise from video compression. Table 5 shows an overview of the cluster background models.

### Table 5 – Cluster Models: An Overview

The first column indicates the category model and the second column the name of each method. Their corresponding acronym is indicated in the first parenthesis and the number of papers counted for each method in the second parenthesis. The third column gives the name of the authors and the date of the related publication.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Methods</th>
<th>Authors - Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-Means</td>
<td>K-means (KM) (9)</td>
<td>Butler et al. (2005) [47]</td>
</tr>
<tr>
<td></td>
<td>Genetic K-Means (GKM) (1)</td>
<td>Duan et al. (2011) [244]</td>
</tr>
<tr>
<td>Codebook</td>
<td>Original codebook model (CB) (1)</td>
<td>Kim et al. (2003) [48]</td>
</tr>
<tr>
<td></td>
<td>Layered codebook model (LCB) (2)</td>
<td>Kim et al. (2005) [245]</td>
</tr>
<tr>
<td></td>
<td>Hybrid cone cylinder codebook model (HCB) (2)</td>
<td>Doshi and Trivedi (2006) [247]</td>
</tr>
<tr>
<td></td>
<td>Spherical codebook model (SCB) (1)</td>
<td>Hu et al. (2012) [248]</td>
</tr>
<tr>
<td></td>
<td>Block based codebook model (BCB) (2)</td>
<td>Deng et al. (2008) [249]</td>
</tr>
<tr>
<td></td>
<td>Hierarchical codebook model (HCB) (3)</td>
<td>Guo and Heu (2010) [251]</td>
</tr>
<tr>
<td></td>
<td>Multi-scale codebook model (MCB) (1)</td>
<td>Zaharescu and Jamieson (2011) [252]</td>
</tr>
<tr>
<td>Basic Sequential Clustering</td>
<td>Basic Sequential Clustering (BSC) (2)</td>
<td>Xiao et al. (2006) [49]</td>
</tr>
<tr>
<td></td>
<td>Modified BSC (MBSC) (2)</td>
<td>Xiao and Zhang (2008) [253]</td>
</tr>
<tr>
<td></td>
<td>Two-Threshold Sequential Clustering (TTSC) (1)</td>
<td>Xiao and Zhang (2008) [254]</td>
</tr>
<tr>
<td></td>
<td>Improved MBSC (IMBSC) (1)</td>
<td>Benalia and Ait-Aoudia (2012) [255]</td>
</tr>
</tbody>
</table>

5.4. **Neural network models**

In this case, the background is represented by means of the weights of a neural network suitably trained on N clean frames. The network learns how to classify each pixel as background or foreground. The main approaches can be classified in the following way:

- **General Regression Neural Network:** Culbrik et al. [50,51] proposed to use a neural network (NN) architecture to form an unsupervised Bayesian classifier for background modeling and foreground detection. The constructed classifier efficiently handles the segmentation in natural-scene sequences with complex background motion and changes in illumination. The weights allow us to model the background and are updated to reflect the change statistics of the background. Furthermore, this algorithm is parallelized on a sub-pixel level and designed to enable efficient hardware implementation.

- **Multivalued Neural Network:** Luque et al. [52] used a foreground detection method based on the use of a multivalued discrete neural network. The multivalued neural network model is used to detect and correct some of the deficiencies and errors of the MOG algorithm. One
of the advantages of using a multivalued neural network for foreground detection is that all process units (neurons) compute the solution to the problem in parallel. Another advantage of the multivalued neural model is its ability to represent nonnumerical classes or states, which can be very useful when dealing with foreground detection problems, in which pixel states are usually defined with qualitative labels: foreground, background and shadow.

- **Competitive Neural Network**: Luque et al. [53] used an unsupervised competitive neural network (CNN) based on adaptive neighborhoods to construct the background model. The weights and the adaptive neighborhood of the neurons models the background and are updated. This algorithm is parallelized on the pixel level and designed for hardware implementation to achieve real-time processing.

- **Dipolar Competitive Neural Network**: Luque et al. [54] improved the CNN approach by using a dipolar CNN which is used to classify the pixels as background or foreground. The dipolar representation is designed to deal with the problem of estimating the directionality of data at a low computational cost. The dipolar CNN achieved better results in terms of precision rate and false positive rate, whereas the false negative rate is, at least, comparable to the CNN.

- **Self Organizing Neural Network**: Maddalena and Petrosino [55] adopted a self-organizing neural network for learning motion patterns in the HSV color space. So, the background model is constructed. This algorithm named Self-Organizing Background Subtraction (SOBS) detects the moving object by using the background model through a map of motion and stationary patterns. Furthermore, an update neural network mapping method is used to make the neural network structure much simpler and the training step much more efficient. Recently, Maddalena and Petrosino [56] improved the SOBS by introducing spatial coherence into the background update procedure. This lead to the so-called SC-SOBS algorithm, that provides further robustness against false detections.

- **Growing Hierarchical Self Organizing Neural Network**: The Self Organizing Neural Network (SONN) presents some limitations related to their fixed network structure in terms of number and neuron arrangements, which has to be defined in advance, and their lack of representation of hierarchical relations among input. To address both limitations, Palomo et al. [58] proposed a growing hierarchical neural network. This neural network model has a hierarchical structure divided into layers, where each layer is composed of different single SONNs with adaptive structures that are determined during the unsupervised learning process according to input data. Experimental results show good performances in the case of illumination changes.

The SOBS [55] and the SOBS-SC [56] are the leading methods in the baseline category of the ChangeDetection.net dataset [22]. Table 6 shows an overview of the neural networks models.

### 5.5. Estimation models

The background is estimated using a filter. Any pixel of the current image that deviates significantly from its predicted value is declared foreground. This filter may be a Wiener filter [19], a Kalman filter [60] or a Chebychev filter [62].

- **Wiener filter**: Toyama et al. [19] 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. The Wiener filter is a linear predictor based on a recent history of values. Any pixel that deviates significantly from its predicted value is declared foreground. In their implementation, they used the past 50 values to compute 30 prediction coefficients. The Wiener filter works well for periodically changing pixels, and for random changes it produces a larger value of the threshold used in the foreground detection. The main advantage of the Wiener filter is that it reduces the uncertainty of a pixels value by accounting for how it varies with time. A disadvantage occurs when a moving object corrupts the history values. To solve this, Toyama et al. [19] kept a history of predicted values for each pixel as well as the history of actual values. For each new pixel, they computed two predictions based on the actual history and the predicted history. If either prediction is within the tolerance, the pixel is considered as background. To handle adaptation, the prediction coefficients are recomputed for every new frame. Furthermore, they added a frame-level algorithm to deal with global illumination changes.

---


---

Table 6 – Neural Networks Models: An Overview. The first column indicates the category model and the second column the name of each method. Their corresponding acronym is indicated in the first parenthesis and the number of papers counted for each method in the second parenthesis. The third column gives the name of the authors and the date of the related publication.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Authors - Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Regression Neural Network (GNR) [4]</td>
<td>Culbrik et al. (2006) [50]</td>
</tr>
<tr>
<td>Multivalued Neural Network (MNN) [1]</td>
<td>Luque et al. (2008) [52]</td>
</tr>
<tr>
<td>Competitive Neural Network (CNN) [2]</td>
<td>Luque et al. (2008) [53]</td>
</tr>
<tr>
<td>Dipolar Competitive Neural Network (DCNN) [1]</td>
<td>Luque et al. (2008) [54]</td>
</tr>
<tr>
<td>Growing Hierarchical Self Organizing Neural Network (GHISONN) [1]</td>
<td>Palomo et al. (2009) [58]</td>
</tr>
</tbody>
</table>
Kalman filter: Karmann et al. [60] proposed a background estimation algorithm based on the Kalman filter which is an optimal estimator of the state of processes which verifies: (1) they can be modeled by a linear system, (2) the measurement and the process noise are white, and have zero mean Gaussian distributions. Under these conditions, the Kalman filter provides an optimal estimate of the state of the process. In the algorithm developed by [60], the state corresponds to the background image \( B_t \) and the measurements to the input gray levels \( I_t \). So, the method assumes that the evolution of the background pixel intensity can be described by a finite-dimensional dynamic system. The system input term is set to zero, and the temporal distributions of the background intensities are considered constant. All unexpected changes are described by random noise, which by hypothesis is a zero mean Gaussian variable. In order to prevent foreground pixels modifying the background image, a different gain factor is introduced if the innovation overcomes a given threshold. In this approach, gradual illumination changes can be captured only by the random noise term which has to vary in time according to them. In this way the zero-mean hypothesis could be not respected. Moreover, sudden illumination changes cause intensity variations that are considered as foreground pixels and cannot be correctly managed. To address these disadvantages, Boninsegna and Bozzoli [256] introduced a new term to model the intensity variations caused by gradual illumination changes. In addition, Messelodi et al. [198] developed a module that measures such global changes, and used this information as an external input to the system. Other improvements concern the texture features instead of the intensity features [257–259], and a local-region feature rather than a pixel feature [260]. Wang et al. [261] proposed to use an extension of the Kalman filter to nonlinear systems which is called Unscented Kalman Filter (UKF) to address more robustly dynamic backgrounds, abrupt illumination changes and camera jitter. Fan et al. [61] used both static background and dynamic background information to renew a Self-Adaptive Kalman filter (SAKF). The two renewal rates in SAKF are automatically obtained by the cumulants of the foreground detection in object region and background region. This updating method is simple, efficient and can be used in real-time detection systems.

Correntropy filter: The Kalman filter gives the optimal solution to the estimation problem when all the processes are Gaussian random processes but it offers a sub-optimal behavior in non-Gaussian settings. That is the case in some challenging situations met in video-surveillance. So, Cinar and Principe [67] proposed a Correntropy filter (CF) that extracts higher order information from the sequence. The information theoretic cost function is based on the similarity measure Correntropy. The Correntropy filter copes with salt and pepper noise which is not Gaussian.

Chebychev filter: Chang et al. [62] used a Chebychev filter to model the background. The idea is to slowly update the background for changes in lighting and scenery while utilizing a small memory footprint as well as having a low computational complexity. The Chebychev filter used has a pass-band frequency of 0.06 Hz and ripple of 10 for a sampling rate of 30 frames per second. The correct estimation is reached after 1250 frames. The drastic changes can be detected if there is a large discrepancy between the estimated background and the current frame that persists for several frames throughout the entire image. Only 2 frames are kept in memory to represent the filter. Furthermore, only the intensity channel is utilized to generate the background image.

The estimation models seem well adapted for gradual illumination changes. Table 7 shows an overview of the estimation background models.

### Table 7 - Estimation Models: An Overview

<table>
<thead>
<tr>
<th>Estimation Models</th>
<th>Algorithm</th>
<th>Authors - Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wiener filter</td>
<td>Wiener filter (WF) (1)</td>
<td>Toyama et al. (1999) [19]</td>
</tr>
<tr>
<td>Kalman filter</td>
<td>Kalman filter (KF) (17)</td>
<td>Karmann et al. (1994) [60]</td>
</tr>
<tr>
<td>Unscented Kalman Filter (UKF) (1)</td>
<td>Wang et al. (2006) [261]</td>
<td></td>
</tr>
<tr>
<td>Self-adaptive Kalman Filter (SAKF) (2)</td>
<td>Fan et al. (2008) [61]</td>
<td></td>
</tr>
<tr>
<td>Correntropy filter</td>
<td>Correntropy (CF) (1)</td>
<td>Cinar and Principe (2011) [262]</td>
</tr>
<tr>
<td>Chebychev filter</td>
<td>Chebychev filter (CF) (2)</td>
<td>Chang et al. (2004) [62]</td>
</tr>
</tbody>
</table>

6. Recent background models

The recent background representation models can be classified in the following categories: advanced statistical background models, fuzzy background models, discriminative subspace learning models, RPCA models, sparse models and transform domain models.

6.1. Advanced statistical models

Advanced statistical background models have been recently developed and can be classified as follows:

- **Mixture models:** In this category, the authors used an other distribution than the Gaussian one used in the GMM listed in Section 5.2. Some authors used the Student-t Mixture Model [63,263] or the Dirichlet Mixture Model [264,64]. Student’s t-mixture model (STMM) has
proven to be very robust against noises due to its more heavily-tailed nature compared to Gaussian mixture model [63] but STMM has not been previously applied to video processing, because EM algorithm cannot be directly applied to the process. The huge increase in complexity would prevent a real-time implementation. Thus, Mukherjee et al. [63] proposed a new real-time recursive filter approach to update the parameters of the distribution effectively. So, the method allows us to model the background and to separate the foreground with high accuracy in the case of slow foreground objects and dynamic backgrounds. In another way, He et al. [264] used a Dirichlet mixture model, which constantly adapts both the parameters and the number of components of the mixture to the scene in a block-based method but the potential advantages were not completely exploited and the algorithm was computationally expensive. Recently, Haines and Xiang [64] improved the DMM to estimate a per-pixel background distribution by a probabilistic regularization. The key improvements consist of the inference for the per-pixel mode count, such that it accurately models dynamic backgrounds, and of the maintenance scheme which updates continuously the background. In a similar way, Fan and Bouguila [65] model and update the background using Dirichlet Mixture Model (DMM) which allows us to deal with non Gaussian processes that are met in real videos. They proposed a batch algorithm and adopted a variational Bayes approach based algorithm over the attained feature representation space. A group-based method (WGKVS) which employs multiple kernel representations incorporates different color representations to conform an enhanced feature representation space. A grouping-based algorithm is employed over the attained feature space to group pixels into foreground or background.

- **Hybrid models:** Ding et al. [67] used a mixture of nonparametric regional model (KDE) and parametric pixel-wise model (GMM) to approximate the background color distribution. The foreground color distribution is learned from neighboring pixels of the previous frame. The locality distributions of background and foreground are approximated with the nonparametric model (KDE). The temporal coherence is modeled with a Markov chain. So, color, locality, temporal coherence and spatial consistency are fused together in the same framework. The models of color, locality and temporal coherence are learned online from complex dynamic backgrounds. In the same idea, Liu et al. [68,194] used a KDE–GMM hybrid model. Under this probabilistic framework, this method deal with foreground detection and shadow removal simultaneously by constructing probability density functions (PDFs) of moving objects and non-moving objects. Here, these PDFs are constructed based on KDE–GMM hybrid model (KGHM) which has advantages of both KDE and GMM. This KGHM models the spatial dependencies of neighboring pixel colors in order to deal with highly dynamic backgrounds.

- **Nonparametric models:** This group of models follows a nonparametric background modeling paradigm. Barnich et al. [69] proposed a sample-based algorithm called Visual Background Extractor (ViBe) [69] that builds the background model by aggregating previously observed values for each pixel location. The key innovation of ViBe is a random selection policy that ensures a smooth exponentially decaying lifespan for the sample values that constitute the pixel models. The second innovation concerns the post-processing to give spatial consistency by using a fast and spatial information propagation method that randomly diffuses pixel values across neighboring pixels. The third innovation relates to the background initialization which is instantaneous and allows the algorithm to start from the second frame of the sequence. Although ViBe gives acceptable detection results in many scenarios, it is problematic with challenging scenarios such as darker background, shadows, and frequent background change. Another approach developed by Hofmann et al. [70] and called Pixel-Based Adaptive Segmenter (PBAS) models the background by a history of recently observed pixel values. PBAS is made of several components. As a central component, the decision block decides for or against foreground based on the current image and a background. This decision is based on the per-pixel threshold. Furthermore, the background model is updated over time in order to deal with gradual background changes. This update depends on a per-pixel learning parameter. The key innovation in PBAS approach is that both of these two per-pixel thresholds change the estimate of the background dynamics. The foreground decision depends on a decision threshold. PBAS outperforms most state-of-the-art methods.

- **Multi-Kernels Models:** Recently, Molina-Giraldo et al. [73, 265] proposed a Weighted Gaussian Video Segmentation (WGKVS) which employs multiple kernel representations and incorporates different color representations to conform an enhanced feature representation space. A grouping-based algorithm is employed over the attained feature space to group pixels into foreground or background.

ViBe [69] and PBAS [70] are the leading methods in the ChangeDetection.net dataset [22]. Table 8 shows an overview of the advanced statistical models.

### 6.2. Fuzzy models

All the critical situations generate imprecisions and uncertainties in the whole process of background subtraction. Therefore, some authors have recently introduced fuzzy concepts in the different steps of background subtraction to take into account these imprecisions and uncertainties. Different fuzzy methods have been developed and are classified in the recent survey made by Bouwmans [20]:

- **Fuzzy background modeling:** The main challenge addressed here consists in modeling multi-modal backgrounds. The algorithm usually used is the Gaussian Mixture Model [17] to deal with this challenge but the parameters are determined using a training sequence which contains insufficient or noisy data. So, the parameters are not well determined. In this context, Type-2 Fuzzy Mixtures of Gaussians (T2F-MOG) [75,267,268] are used to model uncertainties when dynamic backgrounds occur. El Baf et al. proposed two algorithms, that is, one for the uncertainty over the mean and one for the uncertainty over the variance, called T2-FMOG-UM and T2-FMOG-UV, respectively.
The T2-FMOG-UM and T2-FMOG-UV are more robust than the crisp MOG [17]. Practically, T2-MOG-UM is more robust than T2-FMOG-UV. Indeed, only the means are estimated and tracked correctly over time in the MOG maintenance. The variance and the weights are unstable and unreliable, so generating less robustness for the T2-FMOG-UV. In another way, Kim and Kim [74] adopted a fuzzy c-means clustering model that uses fuzzy color histogram as feature. This model allows us to attenuate color variations generated by background motions while still highlighting foreground objects, and receives better results with dynamic backgrounds than the MOG [17].

**Fuzzy foreground detection:** In this case, a saturating linear function can be used to avoid a crisp decision in the classification of the pixels as background or foreground. The background model can be unimodal, such as the running average in [137,269], or multimodal, such as the background modeling with confidence measure proposed in [270]. Another approach consists of aggregating different features such as color and texture features. As seen in Section 2, the choice of the feature is essential and using more than one feature allows it to be more robust to illumination changes and shadows. In this context, Zhang and Xu [148] used texture and color features to compute similarity measures between current and background pixels. Then, these similarity measures are aggregated by applying the Sugeno integral. The assumption made by the authors reflects that the scale is ordinal. The moving objects are detected by thresholding the results of the Sugeno integral. Recently, El Baf et al. [149] have shown that the scale is continuum in the foreground detection. Therefore, they used the same features but with the Choquet integral instead of the Sugeno integral. Ding et al. [271] used the Choquet integral too but they change the similarity measures. Recently, Azab et al. [272] have aggregated three features, that are color, edge and texture. Fuzzy foreground detection is more robust to illumination changes and shadows than crisp foreground detection.

**Fuzzy background maintenance:** The idea is to update the background following the membership of the pixel at the class background or foreground. The membership comes from the fuzzy foreground detection and can be introduced in the maintenance scheme in two ways. The first way [273] consists of adapting the learning rate following the classification of the pixel in a fuzzy manner. For the second way, the maintenance rule becomes a fuzzy combination of two crisp rules [136,274]. The fuzzy adaptive background maintenance allows one to deal robustly with illumination changes and shadows.

- **Fuzzy features:** First, Chiranjeevi and Sengupta [275] proposed a fuzzy feature, called fuzzy correlogram, which is obtained by applying fuzzy c-means algorithm on correlogram. The fuzzy correlogram is computed at each pixel in the background image and the current image. The distance between the two correlograms is obtained by a modified version of K-L divergence distance. If the distance is less than the threshold, then it implies that the current fuzzy correlogram is matched with the background model and the pixel is labeled as background otherwise it is labeled as foreground. Then, the background image is updated with the current fuzzy correlogram by simple adaptive filtering. In a similar way, Chiranjeevi and Sengupta [276] applied a membership transformation on a co-occurrence vector to derive a fuzzy transformed co-occurrence vector with shared membership values in a reduced dimensionality vector space. Fuzzy statistical texture features (FST), derived from this fuzzy transformed co-occurrence vector, are combined with the intensity feature using the Choquet integral. The FST features allow us to deal with dynamic backgrounds better than the crisp statistical texture features. Recently, Chiranjeevi and Sengupta [277] adopted a fuzzy rough-set theoretic measures to embed the spatial similarity around a neighborhood as a model for the pixel. First, they extended the basic histon concept to a 3D histon one, which considers the intensities across the color planes in a combined manner, instead of considering independent color planes. Then, they incorporated fuzziness into the 3D HRI measure. The foreground detection is based on the Bhattacharyya distance between the 3D fuzzy histon model and the corresponding measure in the current image. The background maintenance is made using a selective update scheme.

- **Fuzzy post-processing:** It can be applied on the results. For example, fuzzy inference can be used between the

---

The table below provides an overview of different statistical models used for video background extraction.

**Table 8 - Advanced Statistical Models: An Overview.**

<table>
<thead>
<tr>
<th>Categories</th>
<th>Methods</th>
<th>Authors - Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixture Models</td>
<td>Student-t Mixture Models (STMM) (2)</td>
<td>Mukherjee et al. (2012) [63]</td>
</tr>
<tr>
<td></td>
<td>Dirichlet Process Gaussian Mixture Model (DP-GMM) (2)</td>
<td>He et al. (2012) [264]</td>
</tr>
<tr>
<td></td>
<td>Varational Dirichlet Mixture Model (varDMM) (1)</td>
<td>Fan and Bougula (2012) [65]</td>
</tr>
<tr>
<td></td>
<td>Finite Mixtures of Asymmetric Gaussian Distributions (FMAGD) (1)</td>
<td>Fan and Bougula (2014) [266]</td>
</tr>
<tr>
<td>Hybrid Models</td>
<td>KDE–GMM (KGMM) (1)</td>
<td>Ding et al. (2010) [67]</td>
</tr>
<tr>
<td></td>
<td>KDE–GMM Hybrid Model (KGHM) (2)</td>
<td>Liu et al. (2008) [68]</td>
</tr>
<tr>
<td>Non Parametric Models</td>
<td>Video Background Extractor (ViBe) (3)</td>
<td>Barnich et al. (2009) [69]</td>
</tr>
<tr>
<td></td>
<td>Pixel-Based Adaptive Segmenter (PBASP) (1)</td>
<td>Hofmann et al. (2012) [70]</td>
</tr>
<tr>
<td>Multi-Kernels Models</td>
<td>Weighted Gaussian Video Segmentation (WGVS) (2)</td>
<td>Molina-Giraldo et al. (2013) [73]</td>
</tr>
</tbody>
</table>

*ahttp://sites.google.com/site/pbassegmenter/home.*
Discriminative and mixed subspace learning models

In the literature, only reconstructive subspace learning models such as PCA and NMF have attracted attention in the context of background modeling and foreground detection. However, subspace learning methods can be classified into two main categories: reconstructive or discriminative methods [279]. With the reconstructive representations, the model strives to be as informative as possible in terms of well approximating the original data [280]. Their main goal is encompassing the variability of the training data gathered, meaning these representations are not task-dependent. On the other hand, the discriminative methods provide a supervised reconstruction of the data. These methods are task-dependent, but are also spatially and computationally far more efficient and they will often give better classification results when compared to the reconstructive methods [280]. Practically, reconstructive subspace learning models give more effort to construct a robust background model in an unsupervised manner rather than providing a good classification in the foreground detection. Furthermore, they assume that the foreground has a low contribution in the training step, even though this assumption can only be verified when the moving objects in question are either small or far enough away from the camera. In the end, the only advantage in modeling the background with a reconstructive subspace learning is the lack of supervision required. On the other hand, discriminative subspace learning models allow us a robust supervised initialization of the background and a robust classification of pixels as background or foreground. So, some authors achieved background and foreground separation using discriminative or mixed subspace models:

- **Discriminative subspace models**: Farcas et al. [76] proposed to use a discriminative and supervised approach. This approach is based on an incremental discriminative subspace learning algorithm, called Incremental Maximum Margin Criterion (IMMC) [281]. It derives the online adaptive supervised subspace from sequential data samples and incrementally updates the eigenvectors of the criterion matrix. IMMC does not need to reconstruct the criterion matrix when it receives a new sample, thus the computation is very fast. This method outperforms the reconstructive ones for the foreground detection but the main drawback is that the ground truth images are needed for the background initialization.

- **Mixed subspace models**: Recently, Marghes et al. [77] used a mixed method that combines a reconstructive method (PCA) with a discriminative one (LDA) [282] to robustly model the background. The objective is firstly to enable a robust model of the background and secondly a robust classification of pixels as background or foreground. So, Marghes et al. [77] used the PCA to model the primary distribution of the pixel values among multiple images, and regards the primary distribution as the knowledge of the background. Then, this method assumes that the low-rank principal vectors of the image space contain discriminative information. Then, Marghes et al. [77] applied on them the LDA for background/foreground classification. Experiments show that the mixed model is more robust than the reconstructive methods (PCA, ICA, INMF and IRT) and the discriminative one (IMMC).

Discriminative and mixed subspace models offer a nice framework for background modeling and foreground detection. Furthermore, there are several algorithms that can be evaluated for this field such as Linear Discriminant Analysis (LDA) [283] and Canonical Correlation Analysis (CCA) [284]. For example, LDA exists in several incremental versions as ILDA using fixed point method [285] or sufficient spanning set approximations [286]. In the same way, Partial Least Squares (PLS) methods [287] give a good perspective to model and robustly update the background.

---

Table 9 – Fuzzy models: An overview. The first column indicates the background subtraction steps and the second column the name of each method. Their corresponding acronym is indicated in the first parenthesis and the number of papers counted for each method in the second parenthesis. The third column gives the name of the authors and the date of their first publication that use the corresponding fuzzy concept.

<table>
<thead>
<tr>
<th>Background subtraction</th>
<th>Algorithm</th>
<th>Authors - Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background modeling</td>
<td>Fuzzy C-means Clustering (FCM) (1)</td>
<td>Kim and Kim (2012) [74]</td>
</tr>
<tr>
<td></td>
<td>Type-2 Fuzzy MOG (T2-FMOG) (3)</td>
<td>El Baf et al. (2008) [75]</td>
</tr>
<tr>
<td></td>
<td>Saturation Linear Function (SLF) (5)</td>
<td>Sigari et al. (2008) [137]</td>
</tr>
<tr>
<td>Foreground detection</td>
<td>Sugeno Integral (SI) (1)</td>
<td>Zhang and Xu (2006) [148]</td>
</tr>
<tr>
<td></td>
<td>Choquet Integral (CI) (5)</td>
<td>El Baf et al. (2008) [149]</td>
</tr>
<tr>
<td>Background maintenance</td>
<td>Fuzzy Learning Rate (FLR) (7)</td>
<td>Maddalena et al. (2009) [273]</td>
</tr>
<tr>
<td></td>
<td>Fuzzy Maintenance Rule (FMR) (2)</td>
<td>El Baf et al. (2008) [136]</td>
</tr>
<tr>
<td>Features</td>
<td>Fuzzy Correlogram (FC) (1)</td>
<td>Chiranjeevi and Sengupta (2013) [275]</td>
</tr>
<tr>
<td></td>
<td>Fuzzy statistical texture features (FST) (1)</td>
<td>Chiranjeevi and Sengupta (2012) [276]</td>
</tr>
<tr>
<td></td>
<td>Fuzzy 3D Histon (F3DH) (1)</td>
<td>Chiranjeevi and Sengupta (2012) [277]</td>
</tr>
<tr>
<td>Post-processing</td>
<td>Fuzzy Inference (FI) (5)</td>
<td>Sivabalakrishnan and Manjula (2010) [278]</td>
</tr>
</tbody>
</table>

*http://sites.google.com/site/t2fmog/.*
6.4. Robust subspace models

The background and the foreground are separated via a robust subspace model which is based on a low-rank and sparse decomposition. This decomposition may be done in a Robust Principal Components Analysis [21], or Robust Non-negative Matrix Factorization [101,102], or Robust Orthogonal Subspace Learning [103] framework. Here, we focus on the RPCA models as the other ones attracted less attention for the moment.

Recent research on subspace estimation by sparse representation and rank minimization shows a nice framework to separate moving objects from the background. Robust Principal Component Analysis (RPCA) solved via Principal Component Pursuit (PCP) [21] decomposes a data matrix $A$ in two components such that $A = L + S$, where $L$ is a low-rank matrix and $S$ is a sparse noise matrix. Practically, $A$ contains the training sequence. So, the background sequence is then modeled by the low-rank subspace $L$ that can gradually change over time, while the moving foreground objects constitute the correlated sparse outliers $S$. Robust PCA models can be classified in the following categories:

- **RPCA via Principal Component Pursuit**: The first work on RPCA-PCP developed by Candès et al. [21] proposed a convex optimization to address the robust PCA problem. Under minimal assumptions, this approach called Principal Component Pursuit (PCP) perfectly recovers the low-rank and the sparse matrices. The background sequence is then modeled by a low-rank subspace that can gradually change over time, while the moving foreground objects constitute the correlated sparse outliers. So, Candès et al. [21] showed visual results on foreground detection that demonstrated encouraging performances but PCP presents several limitations for foreground detection. The first limitation is that it requires algorithms to be solved that are computationally expensive. The second limitation is that PCP is a batch method that stacked a number of training frames in the input observation matrix. In real-time application such as foreground detection, it would be more useful to estimate the low-rank matrix and the sparse matrix in an incremental way quickly when a new frame comes rather than in a batch way. The third limitation is that the spatial and temporal features are lost as each frame is considered as a column vector. The fourth limitation is that PCP imposed the low-rank component being exactly low-rank and the sparse component being exactly sparse but the observations such as in video surveillance are often corrupted by noise affecting every entry of the data matrix. The fifth limitation is that PCP assumed that all entries of the matrix to be recovered are exactly known via the observation and that the distribution of corruption should be sparse and random enough without noise. These assumptions are rarely verified in the case of real applications because (1) only a fraction of entries of the matrix can be observed in some environments, (2) the observation can be corrupted by both impulsive and Gaussian noise (3) the outliers i.e moving objects are spatially localized. Many efforts have been recently concentrated to develop low-computational algorithms for solving PCP (Accelerated Proximal Gradient (APG) [288], Augmented Lagrange Multiplier (ALM) [289], Alternating Direction Method (ADM) [290]), to develop incremental algorithms of PCP to update the low-rank and sparse matrix when new data arrive [291] and real-time implementations [292]. Moreover, other efforts have addressed problems that appear specifically in real application such as: (1) Presence of noise, (2) Quantization of the pixels, (3) Spatial constraints of the foreground pixels and (4) Local variation in the background. To address (1), Zhou et al. [86] proposed a stable PCP (SPCP) that guarantees stable and accurate recovery in the presence of entry-wise noise. Becker et al. [87] proposed an inequality constrained version of PCP to take into account the quantization error of the pixel’s value (2). To address (3), Tang and Nehorai [88] proposed a Block-based PCP (BPCP) method via a decomposition that enforces the low-rankness of one part and the block sparsity of the other part. Wohlerig et al. [89] used a decomposition corresponding to a more general underlying model consisting of a union of low-dimensional subspaces for local variation in the background (4). Furthermore, RPCA-PCP can be extended to the measurement domain, rather than the pixel domain, for use in conjunction with compressive sensing [293].

- **RPCA via Outlier Pursuit**: Xu et al. [90] proposed a robust PCA via Outlier Pursuit to obtain a robust decomposition when the outliers corrupted entire columns, that is every entry is corrupted in some columns. Moreover, Xu et al. [90] proposed a stable OP (SOP) that guarantee stable and accurate recovery in the presence of entry-wise noise. **RPCA via Sparsity Control**: Mateos and Giannakis [91] proposed a robust PCA where a tunable parameter controls the sparsity of the estimated matrix, and the number of outliers as a by product.

- **RPCA via Sparse Corruptions**: Even if the matrix $A$ is exactly the sum of a sparse matrix $S$ and a low-rank matrix $L$, it may be impossible to identify these components from the sum. For example, the sparse matrix $S$ may be low-rank, or the low-rank matrix $L$ may be sparse. To address this issue, Hsu et al. [92] imposed conditions on the sparse and low-rank components in order to guarantee their identifiability from $A$.

- **RPCA via Log-sum heuristic Recovery**: When the matrix has high intrinsic rank structure or the corrupted errors become dense, the convex approaches may not achieve good performances. Then, Deng et al. [93] used the log-sum heuristic recovery to learn the low-rank structure.

- **RPCA via Iteratively Reweighted Least Squares**: Guyon et al. [94,294] proposed to solve the RPCA problem by using an Iteratively Reweighted Least Squares (IRLS) alternating scheme for matrix low rank decomposition. Furthermore, spatial constraint are added in the minimization process to take into account the spatial connectivity of pixels. The advantage of IRLS over the classical solvers is its fast convergence and its low computational cost. Recently, Guyon et al. [95] improved this scheme by addressing in the minimization the temporal sparseness of moving objects.

- **Bayesian RPCA**: Ding et al. [96] proposed a Bayesian framework which infers an approximate representation for the noise statistics while simultaneously inferring the...
low-rank and sparse components. Furthermore, Markov dependency is introduced spatially and temporarily between consecutive rows or columns corresponding to image frames. This method has been improved in a variational Bayesian framework [97].

- **Approximated RPCA**: Zhou and Tao [98] proposed an approximated low-rank and sparse matrix decomposition. This method called Go Decomposition (GoDec) produces an approximated decomposition of the data matrix whose RPCA exact decomposition does not exist due to the additive noise, the predefined rank on the low-rank matrix and the predefined cardinality of the sparse matrix. GoDec is significantly accelerated by using bilateral random projection. Furthermore, Zhou and Tao [98] proposed a Semi-Soft GoDec which adopts soft thresholding to the entries of $S$, instead of GoDec which imposes hard thresholding to both the singular values of the low-rank part $L$ and the entries of the sparse part $S$.

These recent advances in RPCA are fundamental for background modeling and foreground detection. However, no RPCA algorithm today seems to emerge and to be able to simultaneously address all the key challenges that accompany real-world videos. This is due, in part, to the absence of a rigorous quantitative evaluation as the authors mainly present visual results.

Some recent quantitative evaluations using the performance metrics have been made but they are limited to one algorithm. For example, Wang et al. [295] studied only RPCA-PCP solved by APG [288]. Xue et al. [296] evaluated the RPCA-PCP solved by Inexact ALM [289] and Guyon et al. [297] adapted the RPCA-PCP with Linearized Alternating Direction Method with Adaptive Penalty [289]. Guyon et al. [299] showed that BPICP [88] gives a better performance than PCP.

In a recent comparative analysis and evaluation, Guyon et al. [85] compared RPCA-PCP solved via Exact ALM [289], RPCA-PCP solved via Inexact ALM [289], QPCP [87] and BRPCA [96] with the Wallflower dataset [19], the I2R dataset [147] and Shah dataset [300]. Experimental results show that BRPCA that address spatial and temporal constraints outperforms the other methods. In the same way, Rueda et al. [301] compare RPCA-PCP solved Exact ALM [289], BRPCA [96] and GoDec [98]. The authors also concluded that the BRPCA offers the best results in dynamic and static scenes by exploiting the existing correlation between frames of the video sequence using Markov dependencies. Table 10 shows an overview of the RPCA models.

<table>
<thead>
<tr>
<th>RPCA models</th>
<th>Algorithm</th>
<th>Authors - Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RPCA</strong></td>
<td>Principal Components Pursuit (PCP) (19)</td>
<td>Candes et al. (2011) [21]</td>
</tr>
<tr>
<td></td>
<td>Stable PCP (SPC) (3)</td>
<td>Zhou et al. (2010) [86]</td>
</tr>
<tr>
<td></td>
<td>Quantized PCP (QPCP) (1)</td>
<td>Becker et al. (2011) [87]</td>
</tr>
<tr>
<td></td>
<td>Block-based PCP (BPCP) (1)</td>
<td>Tang and Nehorai (2011) [88]</td>
</tr>
<tr>
<td></td>
<td>Local PCP (LPCP) (1)</td>
<td>Wohlgem et al. (2012) [89]</td>
</tr>
<tr>
<td><strong>RPCA-OP</strong></td>
<td>Outlier Pursuit (OP) (1)</td>
<td>Xu et al. (2010) [90]</td>
</tr>
<tr>
<td></td>
<td>Stable OP (SOP) (1)</td>
<td>Xu et al. (2010) [90]</td>
</tr>
<tr>
<td><strong>RPCA-SpaCtrl</strong></td>
<td>Sparsity Control (SpaCtrl) (2)</td>
<td>Mateos and Giannakis (2010) [91]</td>
</tr>
<tr>
<td><strong>RPCA-SpaCorr</strong></td>
<td>Sparse Corruptions (SpaCorr) (1)</td>
<td>Hsu et al. (2011) [92]</td>
</tr>
<tr>
<td><strong>RPCA-LHR</strong></td>
<td>Log-sum heuristic Recovery (LHR) (1)</td>
<td>Deng et al. (2012) [93]</td>
</tr>
<tr>
<td><strong>RPCA-IRLS</strong></td>
<td>Iteratively Reweighted Least Squares (IRLS) (3)</td>
<td>Guyon et al. (2012) [94]</td>
</tr>
<tr>
<td><strong>Bayesian RPCA</strong></td>
<td>Bayesian RPCA (BRPCA) (1)</td>
<td>Ding et al. (2011) [96]</td>
</tr>
<tr>
<td></td>
<td>Variational Bayesian RPCA (VBRPCA) (1)</td>
<td>Babacan et al. (2011) [97]</td>
</tr>
<tr>
<td><strong>Approximated RPCA</strong></td>
<td>GoDec (1)</td>
<td>Zhou and Tao (2011) [98]</td>
</tr>
<tr>
<td></td>
<td>Semi-soft GoDec (1)</td>
<td>Zhou and Tao (2011) [98]</td>
</tr>
</tbody>
</table>

*http://www2.ie.psu.edu/sybat/codes.html.
*http://sites.google.com/site/godecomposition/code.

### 6.5. Subspace tracking

He et al. [104] proposed an incremental gradient descent on the Grassmannian, the manifold of all $d$-dimensional subspaces for fixed $d$. This algorithm called Grassmannian Robust Adaptive Subspace Tracking Algorithm (GRASTA) uses a robust $l_2$-norm cost function in order to estimate and track non-stationary subspaces when the streaming data vectors (that are image frames in foreground detection) are corrupted with outliers, that are foreground objects. This algorithm allows to separate background and foreground on-line. GRASTA shows high-quality visual separation of foreground from background. Recently, He et al. [105] proposed t-GRASTA (transformed-GRASTA) which iteratively performs incremental gradient descent constrained to
Detecting Contiguous Outlier detection in the Low-rank Representation (DECOLOR). The decomposition involves the same model than SPCP. The objective function is non-convex and it includes both continuous and discrete variables. Zhou et al. [109] adopted an alternating algorithm that separates the energy minimization into two steps. B-step is a convex optimization problem and F-step is a combinatorial optimization problem.

- **Direct Robust Matrix Factorization**: Xiong et al. [110] proposed a direct robust matrix factorization (DRMF) assuming that a small portion of the matrix A has been corrupted by some arbitrary outliers. The aim is to get a reliable estimation of the true low-rank structure of this matrix and to identify the outliers. To achieve this, the outliers are excluded from the model estimation. The decomposition involves the same model than FCP. Comparing DRMF to the conventional LRM, the difference is that the outliers S can be excluded from the low-rank approximation, as long as the number of outliers is not too large, that is, S is sufficiently sparse. By excluding the outliers from the low-rank approximation, Xiong et al. [110] ensured the reliability of the estimated low-rank structure. Computation is accelerated using a partial SVD algorithm.

- **Direct Robust Matrix Factorization-Row**: Xiong et al. [110] proposed an extension of DRMF to deal with the presence of outliers in entire columns. This method is called DRMF-Row (DRMF-R). Instead of counting the number of outlier entries, the number of outliers patterns is counted using the structured $l_2,0$-norm.

- **Probabilistic Robust Matrix Factorization**: Wang et al. [111] proposed a Probabilistic Robust Matrix Factorization (PRMF) which is formulated with a Laplace error and a Gaussian prior which correspond to an $l_1$ loss and an $l_2$ regularizer, respectively. For model learning, a parallelizable expectation–maximization (EM) algorithm is developed. Furthermore, an online extension of the algorithm for sequential data is provided to offer further scalability. PRMF is comparable to other state-of-the-art robust matrix factorization methods in terms of accuracy and outperforms them particularly for large data matrices.

- **Bayesian Robust Matrix Factorization**: The outliers which correspond to moving objects in the foreground usually form groups with high within-group spatial or temporal proximity. However, PRMF treats each pixel independently without clustering effect. To address this problem, Wang


---

**Table 11 – Subspace tracking models: An overview.** The first column indicates the name of each method. Their corresponding acronym is indicated in the first parenthesis and the number of papers counted for each method in the second parenthesis. The second column gives the name of the authors and the date of the related publication.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Authors - Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grassm. Robust Adaptive Subspace Tracking Algorithm (GRASTA)(^a)[2]</td>
<td>He et al. (2013) [105]</td>
</tr>
<tr>
<td>t-GRASTA(^b)[1]</td>
<td>He et al. (2013) [106]</td>
</tr>
<tr>
<td>$l_0$-norm Robust Online Subspace Tracking (pROST)(^b)[2]</td>
<td>Hage and Kleinsteuber (2012) [107]</td>
</tr>
<tr>
<td>Grassm. Online Subspace Updates with Structured-sparsity (GOSUS)(^a)[1]</td>
<td>Xu et al. (2013) [108]</td>
</tr>
</tbody>
</table>


---

the Grassmannian manifold of subspaces in order to simultaneously estimate a decomposition of a collection of images into a low-rank subspace, a sparse part of occlusions and foreground objects, and a transformation such as rotation or translation of the image. t-GRASTA is four times faster than state-of-the-art algorithms, has half the memory requirement, and can achieve alignment in the case of camera jitter.

Although the $l_1$-norm in GRASTA leads to favorably conditioned optimization problems it is well known that penalizing with non-convex $l_0$-surrogates allows reconstruction even in the case when $l_1$-based methods fail. Therefore, Hage and Kleinsteuber [106] proposed an improved GRASTA using $l_0$-surrogates solving it by a Conjugate Gradient method. Seidel et al. [107] recently improve this model by approaching the problem with a smoothed $l_0$-norm in their algorithm called pROST ($l_p$-norm Robust Subspace Tracking). Experimental results [107] show that pROST outperforms GRASTA in the case of multi-modal backgrounds.

Recently, Xu et al. [108] developed a Grassmannian Online Subspace Updates with Structured-sparsity (GOSUS), which exploits a meaningful structured sparsity term to significantly improve the accuracy of online subspace updates. Their solution is based on Alternating Direction Method of Multipliers (ADMM), where most key steps in the update procedure reduce to simple matrix operations yielding real-time performance.

Table 11 shows an overview of the subspace tracking models.

### 6.6. Low rank minimization

Low Rank Minimization (LRM) methods are extremely useful in many data mining tasks, yet their performances are often degraded by outliers. However, recent advances in LRM allow us robust matrix factorization algorithm that is insensitive to outliers. Robust LRM is then formulated as a matrix approximation problem with constraints on the rank of the matrix and the cardinality of the outlier set. In addition, structural knowledge about the outliers can be incorporated to find outliers more effectively. The main approaches are the following:

- **Contiguous Outliers Detection**: Zhou et al. [109] proposed a formulation of outlier detection in the low-rank representation, in which the outlier support and the low-rank matrix are estimated. This method is called
and Yeung [112] proposed a full Bayesian robust matrix factorization (BRMF). For the generative process, the model parameters have conjugate priors and the likelihood or noise model takes the form of a Laplace mixture. For Bayesian inference, an efficient sampling algorithm is used by exploiting a hierarchical view of the Laplace distribution. Finally, the BMRF is extended by assuming that the outliers form clusters which correspond to moving objects in the foreground. This extension is obtained by placing a first-order Markov random field (MRF) and is called Markov BRMF (MBRMF).

Table 12 shows an overview of the low-rank minimization models.

### 6.7 Sparse models

The sparse models can be classified in the following categories: structure sparsity models [114], dynamic group sparsity models [302] and dictionary models [303,115,116,304]. In compressive sensing, the sparsity assumption is made on the observation data. In the other models, it is the foreground objects that are supposed to be sparse.

- **Compressive sensing models:** Compressive sensing measurements for an image are obtained by its K-sparse representation. Then, this sparse representation can be done on each frame of a video. In this context, Cevher et al. [113] considered the background subtraction as a sparse approximation problem and provided different solutions based on convex optimization [305] and total variation [306]. So, the background is learned and adapted in a low dimensional compressed representation, which is sufficient to determine spatial innovations. This representation is adapted over time to be robust against variations such as illumination changes. Foreground objects are directly detected using the compressive sensing measurements. In another way, Wang et al. [308] used a background modeling scheme, in which background evaluation is performed on the measurement vectors directly before reconstruction. The estimated background measurement vector is constructed through average, running average, median and block-based selective method, respectively. Then the estimated background image is reconstructed using the background model measurements through the gradient projection for sparse reconstruction (GPSR) algorithm [309]. Experimental results show a similar performance to the spatial domain features. Although compressive sensing background subtraction algorithms are being created, no study has been made of the effect of recovery algorithms on the performance of background subtraction. In this context, Davies et al. [310] considered it by applying both Basis Pursuit (BP) and Orthogonal Matching Pursuit (OMP) to the PETS 2001 dataset and comparing their accuracy. The experimentation showed that BP outperformed OMP across the foreground thresholds.

- **Structured Sparsity:** Huang et al. [114] proposed to achieve background and foreground separation using Structured Sparsity (SS), which is a natural extension of the standard sparsity concept in statistical learning and compressive sensing. The structured sparsity problem seeks an optimization problem which can be approximated by a convex relaxation of \( L_0 \) regularization to \( L_1 \) regularization such as Lasso [302]. Another algorithm is the Orthogonal Matching Pursuit (OMP) [303]. These algorithms only address the sparsity but, in practice, the structure is known in addition to sparsity. Furthermore, algorithms such as Lasso and OMP do not correctly handle overlapping components, in that overlapping components are over-counted. To address this issue, Huang et al. [114] developed an algorithm called structured OMP (StructOMP). Experimental results show that StructOMP outperforms Lasso and OMP in the case of background and foreground separation.

- **Dynamic Group Sparsity:** Huang et al. [115] used a learning formulation called Dynamic Group Sparsity (DGS). The idea is that in sparse data the nonzero coefficients are often not random but tend to be a cluster such as in foreground detection. According to compressive sensing, a sparse signal can be recovered from linear random projections. This problem is NP hard. Practically, efficient algorithms were developed to approximate the sparsest solution but all of these algorithms do not consider sparse data priors other than sparsity. However, the nonzero sparse coefficients are often not randomly distributed but group-clustered such as in foreground detection. So, Huang et al. [115] developed a dynamic group sparsity

### Table 12 – Low rank minimization models: An overview. The first column indicates the name of each method. Their corresponding acronym is indicated in the first parenthesis and the number of papers counted for each method in the second parenthesis. The second column gives the name of the authors and the date of the related publication.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Authors - Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contiguous Outliers Detection (DECOLOR$^a$) (1)</td>
<td>Zhou et al. (2013) [109]</td>
</tr>
<tr>
<td>Direct Robust Matrix Factorization (DRMF$^b$) (1)</td>
<td>Xiong et al. (2011) [110]</td>
</tr>
<tr>
<td>Direct Robust Matrix Factorization-Row (DRMF-R$^b$) (1)</td>
<td>Xiong et al. (2011) [110]</td>
</tr>
<tr>
<td>Probabilistic Robust Matrix Factorization (PRMF$^c$) (1)</td>
<td>Yang et al. (2012) [111]</td>
</tr>
<tr>
<td>Bayesian Robust Matrix Factorization (BRMF$^d$) (1)</td>
<td>Yang and Yeung (2013) [112]</td>
</tr>
</tbody>
</table>

$^c$[http://winsty.net/prmf.html](http://winsty.net/prmf.html).
$^d$[http://winsty.net/brmf.html](http://winsty.net/brmf.html).
Sparse error estimation: the proposed method.

The experimental results show the robustness performed on public videos against three state-of-the-art algorithms. The basic assumption of the sparse representation is that the current image can be represented as a linear combination of vectors which forms a dictionary. In order to recover the background image, a functional is minimized. The initialization begins with a known or at least estimated version of the background image. David et al. [116] used a patch-wise average of recent frames as an initial background estimate. The success of this approach depends on the capacity to robustly train the dictionary and find suitable sets of coefficients for estimating the background. In training the dictionary, David et al. [116] exploited the redundancy of consecutive frames by training an initial global dictionary by a k-means classifier. The set of coefficients are obtained by applying a basis matching pursuit. Sivalingam et al. [304] estimated the sparse foreground during the training and performance phases by formulating it as a Lasso [302] problem, while the dictionary update step in the training phase is motivated from the K-SVD algorithm [311]. This method works well in the presence of foreground in the training frames, and also gives the foreground masks for the training frames as a by-product of the batch training phase. In a similar way, Huang et al. [312] initialized the data dictionary with compressive sensing measurement values and sparse basis. In order for initialization to work with corrupted training samples, Zhao et al. [199] proposed a Robust Dictionary Learning (RDL) approach, which automatically prunes unwanted foreground objects out in the learning stage. In another way, Zhou [313] considered a nonparametric Bayesian dictionary learning for sparse image representation using the beta process and the dependent hierarchical beta processing. The results show more robustness than the RPCA [306]. Sang et al. [314] used a dictionary for each position of the video by using the temporal–spatial information of the local region. The proposed method was performed on public videos against three state-of-the-art algorithms. The experimental results show the robustness of the proposed method.

Dictionary learning: David et al. [116] used a sparse representation over a learned dictionary. In the case of a static background, a current frame is decomposed as a sum of the static background image and the foreground image. The basic assumption of the sparse representation is that the current image can be represented as linear combination of vectors which forms a dictionary. In order to recover the background image, a functional is minimized. The initialization begins with a known or at least estimated version of the background image. David et al. [116] used a patch-wise average of recent frames as an initial background estimate. The success of this approach depends on the capacity to robustly train the dictionary and find suitable sets of coefficients for estimating the background. In training the dictionary, David et al. [116] exploited the redundancy of consecutive frames by training an initial global dictionary by a k-means classifier. The set of coefficients are obtained by applying a basis matching pursuit. Sivalingam et al. [304] estimated the sparse foreground during the training and performance phases by formulating it as a Lasso [302] problem, while the dictionary update step in the training phase is motivated from the K-SVD algorithm [311]. This method works well in the presence of foreground in the training frames, and also gives the foreground masks for the training frames as a by-product of the batch training phase. In a similar way, Huang et al. [312] initialized the data dictionary with compressive sensing measurement values and sparse basis. In order for initialization to work with corrupted training samples, Zhao et al. [199] proposed a Robust Dictionary Learning (RDL) approach, which automatically prunes unwanted foreground objects out in the learning stage. In another way, Zhou [313] considered a nonparametric Bayesian dictionary learning for sparse image representation using the beta process and the dependent hierarchical beta processing. The results show more robustness than the RPCA [306]. Sang et al. [314] used a dictionary for each position of the video by using the temporal–spatial information of the local region. The proposed method was performed on public videos against three state-of-the-art algorithms. The experimental results show the robustness of the proposed method.

Sparse error estimation: Static background and dynamic foreground can be considered as samples of signals that vary slowly in time with sparse corruption due to foreground objects. Following this idea, Dikmen et al. [117] achieved the background subtraction as a signal estimation problem, where the error sparsity is enforced through minimization of the $l_1$ norm of the difference between the current frame and estimated background subspace, as an approximation to the underlying $l_0$ norm minimization structure. The minimization problem is solved by a conjugate gradient [97] for memory efficiency. So, background subtraction is then solved as a sparse error recovery problem. Then, different base construction techniques have been compared and discussed in [315]. However, the sparse assumption on the total error in dynamic backgrounds may be inaccurate which degrades the detection performance. To address this problem, Xue et al. [200] proposed an approach that detects foreground objects based on robust linear regression model (RLM). Foreground objects are considered as outliers and the observation error is composed of foreground outlier and background noise. In order to reliably estimate the background, the outlier and noise are removed. Thus, the foreground detection task is converted into outlier estimation problem. Based on the observation that foreground outlier is sparse and background noise is dispersed, an objective function simultaneously estimates the coefficients and sparse foreground outlier. Xue et al. [200] then transformed the function to fit the problem by only estimating the foreground outlier. Experimental results show that this method outperforms the MOG [17], the KDE [36] and the BS [117] in the presence of dynamic backgrounds.

Table 13 shows an overview of the sparse models.

### 6.8. Domain Transform Models

The idea is to separate the background and the foreground in a different domain. For this, different transformation can be used such as the Fast Fourier Transform [127], or the Discrete Cosine Transform [128], or the Walsh Transform [129], or the Wavelet Transform [130] or the Hadamard Transform [131].

#### 6.8.1. Fast Fourier Transform (FFT)

Wren and Porikli [127] estimated the background model that captures spectral signatures of multi-modal backgrounds by using FFT. Those signatures are then used to detect changes in the scene that are inconsistent with these signatures. Results show robustness to low-contrast foreground objects in dynamic scenes. This method is called Waviz.

#### 6.8.2. Discrete Cosine Transform (DCT)

Porikli and Wren [128] developed an algorithm called as Wave-Back that generated a representation of the background using the frequency decompositions of the pixels history. The Discrete Cosine Transform (DCT) coefficients 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 case of waving trees. Another approach developed by Zhu et al. [316] used two features: the DC parameters and the low frequency AC one, each of which focuses on intensity and texture information, respectively. The AC feature parameter consists of the sum of the low frequency coefficients. So, its distribution is assumed to be a single Gaussian which caused false decision under dynamic backgrounds. In another way, Wang et al. [317] utilized only the information from DCT coefficients at block level to construct background models at pixel level. They implemented the running average, the median and the MOG in the DCT domain. Evaluation results show that these algorithms have much lower computational complexity in the DCT domain than in the spatial domain with the same accuracy.

### 6.8.3. Walsh Transform (WT)

Tezuka and Nishitani [129] modeled the background using the MOG applied on multiple block sizes by using Walsh transform (WT). Feature parameters of WT applied to the MOG are determined by using the vertical, horizontal and diagonal direction coefficients, having strong spatial correlation among them. Then, four neighboring WTs are merged into a WT which are four times wider without using the inverse transform. The WT spectral nature reduces the computational steps. Furthermore, Tezuka and Nishitani [129] developed a Selective Fast Walsh Transform (SFWT).

### 6.8.4. Wavelet Transform (WT)

Wavelet transform can be used in background modeling and foreground detection as follows:

- **Marr wavelet:** Gao et al. [130] proposed a background model based on Marr wavelet kernel and used a feature based on binary discrete wavelet transforms to achieve foreground detection. The background model keeps a sample of intensity values for each pixel in the image and uses this sample to estimate the probability density function of the pixel intensity. The density function is estimated using a Marr wavelet kernel density estimation technique. Practically, Gao et al. [130] used difference of Gaussians (DOG) to approximate Marr wavelet. The background is initialized with the first frame. For the foreground detection, the background and current images are transformed in the Binary Discrete Wavelet domain and then the difference is performed in each sub-band. Experiments show that this method outperforms the MOG in traffic surveillance scenes, even though the objects are similar to the background.

- **Dyadic Wavelet:** Guan [318] proposed to use the Dyadic Wavelet (DW) to detect foreground objects. The difference between the background and the current images is decomposed into multi-scale wavelet components. The features are the HSV components. The value component is used to achieve foreground detection and the saturation component is used to suppress moving shadows. However, dyadic wavelet transformation is achieved through tensor product which depends on the horizontal and vertical direction of the image signals. So, it will cause the cut-off of horizontal and vertical direction.

- **Orthogonal non-separable Wavelet:** To avoid the cut-off effect in DW, Gao et al. [319] used orthogonal non-separable wavelet transformation of the training frames and extracted the approximate information to reconstruct the background. Non-separable wavelet transformation is the real multi-dimensional wavelet transformation and processes signal image as a block rather than separate rows and columns. For the background maintenance, a running average scheme is used if the background has a gradual change, otherwise if the background has a sudden change, the background is replaced by the current frame.

- **Daubechies complex wavelet:** Jalal and Singh [320] used the Daubechies complex wavelet transform because it is approximately shift-invariant and has better directionality information with respect to DWT [130]. The background model is the median which is updated over time. For the foreground detection, the threshold is a pixel-wise one and is also updated over time following the activity of the pixel.

### 6.8.5. Hadamard Transform (HT)

Baltieri et al. [131] proposed a fast background initialization method made at block level in a non-recursive way to obtain the best background model using the minimum number of frames as possible. For this, each frame is split into blocks, producing a history of blocks and searching among them for the most reliable ones. In this last phase, the method
works at a super-block level evaluating and comparing the spectral signatures of each block component. These spectral signatures are obtained with the Hadamard Transform which is faster than DCT.

Table 14 shows an overview of the domain transform models.

### 7. Resources, datasets and codes

#### 7.1. Background subtraction web site

This website contains a full list of references in the field, links to available datasets and codes. In each case, the list is regularly updated and classified following the background models as in this paper. An overview of the content of the Background Subtraction Website is given at the home page.

In addition to the sections which concern the steps and issues of background subtraction, this website gives references and links to surveys, traditional and recent datasets, available implementations, journals, conferences, workshops and websites.

#### 7.2. Datasets

Several datasets available to evaluate and compare background subtraction algorithms have been developed in the last decade. We classified them in traditional and recent datasets.

The traditional datasets provide videos with several challenges and some ground-truth images but none of them addresses all the challenges. On the other hand, the recent datasets provide realistic large-scale videos with accurate ground truth giving a balanced coverage of the range of challenges present in the real-world. All these datasets are publicly available and their links are provided on the Background Subtraction Web Site in the section Available Datasets.

##### 7.2.1. Traditional datasets

There are 9 main traditional datasets which are the following ones:

- **Wallflower dataset**: The dataset Wallflower was provided by Toyama et al. [19] and gives representation of real-life situations typical of scenes liable to be met in video surveillance. Moreover, it consists of seven video sequences, with each sequence presenting one of the difficulties a practical task is likely to encounter (i.e. illumination changes, dynamic backgrounds). The size of the images is 160 × 120 pixels. It is one of the most used dataset but its main drawback is that there is only one ground-truth image per sequence.
- **I2R dataset**: The dataset I2R provided by Lin and Huang [147] consists of nine video sequences, with each sequence presenting dynamic backgrounds or illumination changes. The size of the images is 176 × 144 pixels. The advantage of this dataset is that 20 ground-truth images are provided for each sequence. These ground-truth images are not consecutive and are taken when a critical situation occurs.
- **Carnegie Mellon dataset**: The sequence of CMU which is provided by Sheikh and Shah [300] involves a camera mounted on a tall tripod. The wind caused the tripod to sway back and forth causing nominal motion in the scene. Furthermore, some illumination changes occur. The ground-truth is provided for all the frames allowing a reliable evaluation.
- **LIUM dataset**: The Laboratory of Image and Media Understanding (LIUM) provides 5 sequences with ground truth images made every 15 frame. Furthermore, three sequences from PETS 2001 are given with the ground-truth images are not consecutive and are taken when a critical situation occurs.
- **UCSD dataset**: This dataset consists of 18 video sequences from the Statistical Visual Computing Lab (SVCL). The frames of each sequence are provided in JPEG format. The ground truth mask is given in the form of a 3D array in Matlab, where 1 indicates foreground and 0 indicates background. For some sequences, the number

---

5. [http://sites.google.com/site/backgroundsubtraction/](http://sites.google.com/site/backgroundsubtraction/)

6. [http://sites.google.com/site/backgroundsubtraction/test-sequences](http://sites.google.com/site/backgroundsubtraction/test-sequences)
of frames of the ground truth mask is smaller than the number of frames in the sequence. But the ground truth is provided for frames starting from frame 1 of the sequence.

- **SZTAKI surveillance dataset**: This benchmark set contains raw video frames and binary foreground with shadow ground-truth masks, which were used for validation in publications [321,322]. From this dataset, five evaluated sequences can be downloaded: two of them (Laboratory, Highway) come from the ATON benchmark set [182], but the enclosed ground truth was generated by C. Benedek. The three remaining sequences (Sepm, Seam, Senoon) are outdoor surveillance videos which were captured and evaluated by the employees of the MTA-SZTAKI and the Pazmany Peter Catholic University Budapest. Not all frames of the video sequence have ground truth masks. Corresponding images have the same ordinary number.

- **VSSN 2006 dataset**: This dataset consists of different categories of video classified as follows: (1) Vacillate background, gradual illumination changes and shadows, (2) Sudden changes in illumination, (3) Bootstrapping and (4) Two or more cameras with many people. In each video, the moving objects are synthetic ones in a real backgrounds. So, the ground-truth images are very precise.

- **OTCBVS 2006 dataset**: This benchmark related to the conference “Object Tracking and Classification in and Beyond the Visible Spectrum” (OTCBVS12) contains sequences for person detection and face detection. Three sequences are then interesting for background subtraction: (1) Dataset 01 (OSU Thermal Pedestrian) which concerns person detection in thermal imagery, (2) Dataset 03 (OSU Color-Thermal Database) on fusion-based object detection in color and thermal imagery and (3) Dataset 05 (Terravic Motion IR Database) which focus on detection and tracking with thermal imagery.

- **PETS datasets**: These datasets related to the conference “Performance Evaluation of Tracking and Surveillance” (PETS) consist of different datasets such as PETS 2001, PETS 2003 and PETS 2006. They are more adapted for tracking evaluation than for background subtraction directly in the sense that the ground-truth is provided as bounding boxes.

In summary, the Wallflower and I2R datasets provide videos with different representative challenges susceptible to meet in video surveillance. The CMU dataset focuses on the camera jitter. The OTCBVS dataset regards infrared videos. The ATON dataset is limited to shadows. Finally, none of these datasets are a realistic large-scale dataset with accurate ground-truth providing a balanced coverage of the range of challenges present in the real world.

### 7.2.2. Recent datasets

There are several recent datasets which are the following ones:

- **ChangeDetection.net dataset**: The CDW13 dataset [22] presents a realistic, large-scale video dataset consisting of nearly 90,000 frames in 31 video sequences representing 6 categories selected to cover a wide range of challenges in 2 modalities (color and thermal IR). The main characteristic of this dataset is that each frame is annotated for ground-truth foreground, background, and shadow area boundaries. This allows an objective and precise quantitative comparison of background subtraction algorithms.

- **BMC 2012 dataset**: The BMC14 (Background Models Challenge) [24,323] is a workshop organized within ACCV (Asian Conference in Computer Vision) about the comparison of background subtraction techniques with both synthetic and real videos. This benchmark is first composed of a set of 20 synthetic video sequences with the corresponding ground truth, frame by frame, for each video (at 25 fps). The first part of 10 synthetic videos are devoted to the learning phase of the proposed algorithms, while the 10 others are used for evaluation. BMC also contains 9 real videos acquired from static cameras in video-surveillance contexts for evaluation. This dataset has been built in order to test the algorithms’ reliability during a certain time and in difficult situations such as outdoor scenes. Furthermore, this dataset allows us to test the influence of some difficulties encountered during the foreground detection step, as the presence of waving trees, cast shadows or sudden illumination changes in the scene.

- **SABS dataset**: The SABS15 (Stuttgart Artificial Background Subtraction) dataset [23] represents an artificial dataset for pixel-wise evaluation of background models. Synthetic image data generated by modern ray-tracing makes realistic high quality ground-truth data available. The dataset consists of video sequences for nine different challenges of background subtraction for video surveillance. These sequences are further split into training and test data. For every frame of each test sequence ground-truth annotation is provided as color-coded foreground masks. This way, several foreground objects can be distinguished and the ground-truth annotation could also be used for tracking evaluation. The dataset contains additional shadow annotation that represents for each pixel the absolute luminance distance between the frame with and without foreground objects. The sequences have a resolution of 800 × 600 pixels and are captured from a fixed viewpoint.

- **MAR dataset**: Maritime Activity Recognition (MAR16) [152] is a dataset containing data coming from different video sources (fixed and Pan-Tilt-Zoom cameras) and from different scenarios. There are 10 videos from fixed cameras with ground-truth images and 15 form PTZ cameras. The aim of this dataset is to provide a set of videos that can be used to develop intelligent surveillance system for maritime environment.

- **RGB-D object detection dataset**: The RGB-D17 dataset [324] provides five sequences of indoor environments, acquired with the Microsoft Kinect RGB-D camera. Each sequence contains different challenges such as cast shadows, color

---

12 http://www.vcipl.okstate.edu/otcbvs/bench/.
13 http://www.changedetection.net/.
16 http://labroco.co.dis.uniroma1.it/MAR/.
17 http://www.gti.ssr.upm.es/~mac/
and depth camouflage. For each sequence a hand-labeled ground truth is provided.

- **Citic RGB-D dataset**: The Citic dataset contains sequences recorded with rectified stereo cameras, and some frames have been hand-segmented to provide ground-truth information.

- **Fish4Knowledge dataset**: The Fish4knowledge dataset is an underwater benchmark dataset for target detection against complex background which consists of 14 videos categorized into seven different classes representing complex challenges in background modeling.

- **Aquatheque dataset**: The Aquatheque dataset contains four different image sequences from the “Aquatheque” project, and one sequence from the experimental study of turbulent flow. The fish need to be well detected in order to allow features extraction which are used to recognize the species of the fish. One sequence is used for the experimental study of turbulent flow in vertical slot fishways to optimize their protection. Each of the five sequences shows different situations and challenges such as bootstrapping, occlusion, camouflage, light changes and dynamic background.

In summary, there are two recent large-scale datasets: (1) the ChangeDetection dataset which contains visible and infrared videos and the BMC dataset which contains synthetic and real videos. The SABS dataset concerns synthetic videos with different scenarios. On the other hand, the MAR dataset concerns maritime scenes. Finally, two datasets allow the evaluation when the depth information is used as a feature, and two datasets concern fish detection in underwater scenes.

### 7.3. Background Subtraction (BGS) libraries

There are several libraries where background subtraction algorithm are available. The first one is under OpenCV, but there are only two algorithms: the MOG and the foreground detection method developed by Li et al. Then, Parks and Fels evaluated background subtraction algorithms using Visual Microsoft C and OpenCV but their code, available at the personal website of Parks, cannot be directly integrated under OpenCV. More recently, Laurence Bender developed a first library using OpenCV at the Laboratory of Arte Electrónico e Inteligencia Artificial of the Universidad Nacional de Tres de Febrero. This library called Scene is an open source multi-platform computer vision framework that performs background subtraction using two traditional algorithms (SG, MOG) and three recent algorithms based on fuzzy classification rules and neural networks (Fuzzy Gaussian, SOBS, SOBS-SC).

It was mainly designed as a toolkit for the development of interactive art projects that investigate the dynamics of complex environments. In 2012, Andrews Sobral developed the BGSLibrary which provides a C++ framework to perform background subtraction. The code works either on Windows or on Linux and requires OpenCV version 2.0 or superior. Currently, the library offers more than 29 background subtraction algorithms. A large number of algorithms were provided by several authors. The source code is available under GNU GPL v3 license, the library is free and open source. Any user can download the latest project source code using a SVN client. In Windows, a demo project for Visual Studio 2010 is provided. An executable version of BGSLibrary is available for Windows 32 bits and 64 bits, and all required DLL’s are included in package. For Linux users, a Makefile can be used to compile all files and generate an executable example. In the BGSLibrary website, the users can also download the BGSLibraryGUI developed in Java. It is a friendly Graphical User Interface to configure and run BGSLibrary. The algorithms are classified according to their similarities: Basic methods, fuzzy based methods, statistical methods using one Gaussian, statistical methods using multiple Gaussians, type-2 fuzzy based methods, statistical methods using color and texture features and neural networks methods. The BGS library stands out as the reference in the field to compare background subtraction algorithms.

### 8. Conclusion

This paper provides a survey of the traditional and recent background models used in background subtraction. It has two important characteristics that make it different and attractive with respect to the other reviews. First, it considers a classification of the background models based on the mathematical tools used. This has never be made before in the literature for all the background models. Second, resources such as traditional and recent datasets, and libraries are presented.

The Gaussian models and support vector models are greatly designed for dynamics backgrounds and subspace learning models for illumination changes. Neural networks models offer a good compromise between performance and computation cost. Algorithms that seems to be able to simultaneously address all the key challenges that accompany real-world videos are the ViBe, PBAS, SOBS, SOBS-SC and 3dSOBS. Their key advantage is that they used robust update models to deal with illumination changes and dynamic backgrounds.

In our opinion, future developments may concern recent advances on RPCA and sparse models which show a great potential to separate the background and the foreground but they actually need investigations to be implemented in both incremental and real-time ways. Other investigations may concern fuzzy models which offer a nice potential too and feature selection which seems to be an open problem in this field.
Acknowledgments

I would like to acknowledge Megan Cook for her help in proofreading this paper.

References


60


[103] X. Shu, Advanced imaging via multiplexed sensing and compressive sensing (Ph.D. thesis), University of Illinois at Urbana-Champaign, 2013.


[106] C. Hage, M. Kleinsteuber, Robust PCA and subspace tracking from incomplete observations using $l_p$-surrogates, Preprint, 2012.


A. Tavakkoli, Novelty detection: an approach to foreground detection in videos, Pattern Recognit. 7 (2010).


A. Tavakkoli, Novelty detection: an approach to foreground detection in videos, Pattern Recognit. 7 (2010).


