HAL
archives-ouvertes.fr

On the Role and the Importance of Features for Background Modeling and Foreground Detection

Thierry Bouwmans, Caroline Silva, Cristina Marghes, Mohamed Sami Zitouni, Harish Bhaskar, Carl Frélicot

▶ To cite this version:

HAL Id: hal-01724317

https://hal.archives-ouvertes.fr/hal-01724317

Submitted on 6 Mar 2018

On the Role and the Importance of Features for Background Modeling and Foreground Detection

Thierry Bouwmans · Caroline Silva · Cristina Marghes · Mohammed Sami Zitouni · Harish Bhaskar · Carl Frelicot

Abstract Background modeling has emerged as a popular foreground detection technique for various applications in video surveillance. Background modeling methods have become increasing efficient in robustly modeling the background and hence detecting moving objects in any visual scene. Although several background subtraction and foreground detection have been proposed recently, no traditional algorithm today still seem to be able to simultaneously address all the key challenges of illumination variation, dynamic camera motion, cluttered background and occlusion. This limitation can be attributed to the lack of systematic investigation concerning the role and importance of features within background modeling and foreground detection. With the availability of a rather large set of invariant features, the challenge is in determining the best combination of features that would improve accuracy and robustness in

T. Bouwmans
Lab. MIA, Univ. La Rochelle, France
Tel.: +05.46.45.72.02
E-mail: tbouwman@univ-lr.fr

C. Silva
Lab. MIA, Univ. La Rochelle, France
E-mail: lolyne.pacheco@gmail.com

C. Marghes
Movidius
E-mail: cristina.marghes@gmail.com

M. Sami Zitouni
Visual Signal Analysis and Processing (VSAP) Research Center
Department of Electrical and Computer Engineering, Khalifa University, U.A.E E-mail: mohammad.zitouni@kustar.ac.ae

H. Bhaskar
Visual Signal Analysis and Processing (VSAP) Research Center
Department of Electrical and Computer Engineering, Khalifa University, U.A.E E-mail: harish.bhaskar@kustar.ac.ae

C. Frelicot
Lab. MIA, Univ. La Rochelle, France
E-mail: cfrelicot@univ-lr.fr

detection. The purpose of this study is to initiate a rigorous and comprehensive survey of features used within background modeling and foreground detection. Further, this paper presents a systematic experimental and statistical analysis of techniques that provide valuable insight on the trends in background modeling and use it to draw meaningful recommendations for practitioners. In this paper, a preliminary review of the key characteristics of features based on the types and sizes is provided in addition to investigating their intrinsic spectral, spatial and temporal properties. Furthermore, improvements using statistical and fuzzy tools are examined and techniques based on multiple features are benchmarked against reliability and selection criterion. Finally, a description of the different resources available such as datasets and codes is provided.

Keywords Background modeling · Foreground Detection · Features · Local Binary Patterns

1 Introduction

Background modeling and foreground detection are important steps for video processing applications in video-surveillance [86], optical motion capture [68], multimedia [20], teleconferencing and human-computer interface. The aim is to separate the moving objects, called "foreground", from the static information, called "background". For example, Fig. 1 shows an original frame of a sequence from the BMC 2012 dataset [552], the reconstructed background image and the moving objects mask obtained from a decomposition into the low-rank matrix and sparse matrix based model [56]. Conventional background modeling methods exploit the temporal variation of each pixel to model the background and hence use it in conjunction with change detection for foreground extraction. The last decade witnessed very significant contributions to this field [52][48][49] [51][56][57][53][50][191] [55]. Despite these works and advances to background modeling and foreground detection, the dynamic nature of visual scenes attributed by changing illumination conditions, occlusion, background clutter and noise have challenged the robustness of such techniques. Under this pretext, focus has shifted towards the investigation of features and their role in improving both the accuracy and robustness of background modeling and foreground detection. Although fundamental low-level features such as color, edge, texture, motion and stereo have reported reasonable success, recent visual applications using mobile devices and internet videos where the background is non-static, require more complex representations to guarantee robust moving object detection [54]. Furthermore, in order to generalize existing background modeling and foreground detection schemes to real-life scenes where dynamic variations are inevitable and the pose of the camera is little known, automatic feature selection, model selection and adaptation for such schemes are often desired.

Considering the needs and challenges aforementioned, in this paper, a comprehensive review of low-level and hand-crafted features used in background modeling and foreground detection is initiated for benchmarking them against the complexities

Fig. 1 Background Modeling and Foreground Detection: Original image (309), reconstructed background image, foreground mask (Sequences from BMC 2012 dataset [552]).

of typical dynamic scenes. Thus, the aim of this survey is then to provide a first complete overview of the role and the importance of features in background modeling and foreground detection by reviewing both existing and new ideas for (1) novices who could be students or engineers beginning in the field of computer vision, (2) experts as we put forward the recent advances that need to be improved, and (3) reviewers to evaluate papers in journals, conferences, and workshops. In addition, this survey gives a complete overview Moreover, an accompanying website called the Features Website[1] is provided. It allows the reader to have a quick access to the main resources, and codes in the field. So, this survey is intended to be a reference for researchers and developers in industries, as well as graduate students, interested in robust background modeling and foreground detection in challenging environments.

Some of the main contributions of this paper can be summarized as follows:

– **A review regarding feature concepts:** A first complete overview of low-level and hand-crafted features used in background modeling and foreground detection over the last decade concerning more than 600 papers. After a preliminary overview on the key concepts in the field of features in Section 2, a survey of spectral features including color features are detailed in Section 4. Then, spatial features such as edge, texture and stereo features are studied in Section 5, Section 6 and Section 7, respectively. Temporal features such as motion features are reviewed in Section 8. In Section 15, features that are extracted in alternative domains other than the pixel domain are described. Finally, the different strategies of combining multiple features using fusion operators and feature selection mechanisms are discussed in Section 17 and Section 18.

– **A description of the different resources available** to allow fair comparisons of the features. We present the color datasets and recent RGB-D datasets with accurate ground-truth providing a balanced coverage of the range of challenges which are present in the real world. Furthermore, we present the LBP Library which provide a common framework for the implementation of the local texture patterns.

---

[1] https://sites.google.com/site/featuresbackgroundforeground/

The rest of this paper is organized as follows. First, a preliminary overview that investigates the classification of low-level and hand-crafted features by size and type of the features are considered within the context of background modeling and foreground detection in Section 2. Moreover, basic concepts on feature reliability, feature fusion and feature selection are detailed in Section 2.6.1, Section 2.6.2 and Section 2.6.3, respectively. Further, each individual feature is reviewed with their crisp description in Section 4 to Section 8. For each feature, the paper shall also present an investigation of their intrinsic properties that facilitate robustness against the challenges in real-life videos. Also, the paper shall provide an insight on strategies to combine multiple features using fusion operators (Section 17) and hence apply feature selection using boosting algorithms (Section 18). Finally, a description of the different resources available such as datasets and codes is provided in Section 19. Section 21 concludes with remarks on future research directions.

2 A Preliminary Overview

Features (descriptors or signatures) characterize a picture element captured in the current frame of a video sequence and are compared against a known background model to classify it as either foreground or background. Features can be **(1)** low-level features directly obtained from the sensors as color features, **(2)** low-level computed features as gradients, **(3)** hand-crafted features as texture features, and **(4)** features learned by machine learning methods such as deep learning method which is able to learn deep and hierarchical features, which turn out to be much more powerful than classical hand-crafted features (also called hand-design features) for comparing image patches [667][60][582]. Furthermore, feature representations can take multiple forms and can be computed for and from: a pixel, a block around the central pixel and a cluster (a region with the same value of feature than the current pixel). Practically, there are several types of features which can be computed either in the spatial or transform domains. Some of the features commonly used within the background modeling literature includes: color features, edge features, stereo features, motion features, texture features, local histogram features and Haar-like features [557]. These different features have intrinsic properties that allow the model to take into account spectral, spatial and/or temporal characteristics. Furthermore, these features use mathematical concepts in their design that facilitate computing them using well-known statistical or fuzzy concepts. Thus, features used in background/foreground separation can be classified from four different view points: their size, their type in a specific domain, their intrinsic properties and their mathematical concepts (Section 2.1 to Section 2.4). Then, we investigate how features can be used in terms of reliability, fusion and selection in Section 2.6.

2.1 Classification by Size

The size of the picture element chosen for interpreting necessary features that faithfully represent its characteristics plays a crucial role in modeling. As mentioned earlier, features can be computed from and for a pixel [512], a block [147] or a cluster

[41]. That is, the size of the picture element that is used to model the background and hence for comparing the current image frame to the background model, can either be a pixel [512], a block [147], a region (Regions of difference [338], shape [241], behavior [265], cluster [41], super-pixel [253], global appearance [680]) with a feature value. During practical implementations, a feature value at a given pixel can either depend on the feature value at the pixel itself or on the feature values around a predefined neighborhood in the form of a block or a cluster.

– **Pixel-based Features:** These features, otherwise known as point features, concern only the pixel at a given location (x, y). This is the case of intensity and color features but in some cases include stereo features too. The background model applied in this case of pixel-based modeling and comparison is an independent process on each individual pixel. Practically, these features are used in uni-modal or multi-modal pixel-wise background modeling and foreground detection. Furthermore, pixel-based feature can be used to compute the mean or an other statistic value over spatial and/or temporal neighborhood to take into account spatial and/or temporal constraints. Then, the statistic value is assigned to the central pixel. For example, Varadarajan et al. [554][553] proposed a region-based Mixture of Gaussians called (R-MOG) instead of a pixel based MOG. Each region is a square neighborhood which is effectively a block of size 4×4. Then, the color mean obtained from the neighborhood is assigned to the central pixel.

– **Block-based Features:** This category of features is a generalization of the pixel-type, where in the element size a block of 1×1 or any arbitrary block size $m \times n$ it represents an individual feature. In contrast to the previous case of pixel-based feature, which equally applies, spatial and/or temporal information can also be computed depending on the spatial and temporal interaction of the element to its neighborhood as in edge, texture and motion features. To completely exploit their potential, the spatial and/or temporal properties of these features need to be taken into account in all the background subtraction steps to be fully addressed. Practically, these block-based features can be assigned to a central pixel of a block (or neighborhood), or to all the block. For example, textures such as Local Binary Pattern can be assigned at each central pixel of a block size 3 by moving this block all over the frame, or to all the block as in the works of Heikkila and Pietikainen [206], and Heikkila et al. [207] which used a pixel-wise LBP histogram based one (LBP-P) and a block-wise LBP histogram based approach (LBP-B), respectively. Thus, the block-based features can be used in pixel-wise or block-wise background modeling and foreground detection. When the features are obtained from the video compressed domain, the approach is mandatory block-based because the block are pre-defined and thus they can not be moved over the frame. However, in block-based modeling and comparison, blocks (also called patches [684][681][146][619]) can overlap or not [183]. A block is usually obtained as a vector of 3×3 neighbors of the current pixel. The advantage is to take into account the spatial dimension to improve the robustness and to reduce the computation time. Furthermore, blocks can be of spatio-temporal type called spatio-temporal blocks [429], spatio-temporal neighborhoods [562],

spatio-temporal patches [321][595][596] or bricks [605][337][678]) that intrinsicly encapsulate temporal information within spatial relationships of a group of pixels. In Pokrajac and Latecki [429], a dimensionality reduction technique is applied to obtain a compact vector representation for each block. These blocks provide a joint representation of texture and motion patterns. One advantage is their robustness to noise and to the movement in the background. However, the disadvantage is that the detection is less precise because only blocks are detected, making them unsuitable for applications that require detailed shape information.

– **Region-based Features:** Region-level (cluster-level, superpixel-level) features consider element sizes that are non-uniform across the image frame considered, and then specific features are computed on the corresponding element size. First, pixels in an image frame are grouped using an application-specific homogeneity criteria, typically exploiting partitioning mechanisms as follows: **1)** region-based mechanisms as in Lin et al. [338] with the notion of Regions of Difference (RoD), **2)** shape mechanisms as proposed in Jacobs and Pless [241], **3)** behavior mechanisms as in Jodoin et al. [265], **4)** clustering mechanisms as discussed by Bhaskar et al. [41][40][42], and Park and Byun [419], and **5)** super-pixel mechanisms as in Sobral et al. [253], Ebadi et al. [141][140], Zhao et al. [671] and Chen et al. [79]. For example in Bhaskar et al. [41], each cluster contains pixels that have similar features in the color space. Then, the background model is applied on these clusters to obtain cluster of pixels classified as background or foreground. This cluster-wise approach gives less false alarms. Instead of the block-wise approach, the foreground detection is obtained at a pixel-level precision.

Pixel-based features need less time to be extracted than block-based or region-based features which require to be computed. In literature, in general, it can be summarized that the size of the feature and the comparison element determines the robustness of background modeling to noise and the challenges met in the videos, and often controls precision of foreground detection. A pixel-based modeling and comparison gives a pixel-based precision but it is less robust to noise compared to block-based or region-based based modeling and comparison. However, there are several works which combined block-based (or region-based) and pixel-based approaches to reduce computation time by first using a block (or region) approach, and second to obtain a pixel precision by using a pixel-based approach, and they can be classified as follows: **(1)** multi-scales strategies [668][113][112][106][85][180], **(2)** multi-levels strategies [545][249][642][689][405][533][101][102][103][529][623], **(3)** multi-resolutions strategies [353][569][683][599],**(4)** multi-layers strategies [289][286] [664][489][596][597] [521][163][182][612][367][386][259], **(5)** hierarchical strategies [184][185][186][81] [16][616][575][686][94][675], and **(6)** coarse-to-fine strategies [37][36] [157][555][690]. The analysis of these different approaches is out of the scope of this review, and the reader can found details about these strategies in [55].

2.2 Classification by Type

Features can be computed in the pixel domain or in a transform domain. In this section, features those are predominantly computed in each domain and their robustness to critical situations in real videos, are discussed.

2.2.1 Features in the Pixel Domain

Features are popularly computed in the pixel domain as the value of the pixel is directly available. The following features are commonly used:

– **Intensity features:** Intensity features are the most basic features that can be provide by gray-level cameras or infra-red (IR) cameras (See Section 3).

– **Color features:** The color features in the RGB color space are most widely used because it is directly available from the sensor or the camera. But the RGB color space has an important drawback: its three components are dependent to each other which increases its sensitivity to illumination changes. For example, if a background point is covered by the shadow, the three component values at this point could be affected because the brightness and the chromaticity information are not separated. Thus, the three component values increase or decrease together as the lighting increases or decreases, respectively [349]. Alternative color spaces that have also been explored in the literature include YUV or YCrCb spaces. Several comparisons between these color spaces are available in the literature including [298][450][273][28][350] and usually YCrCb is selected as the most appropriate color space. Although color features are often very discriminative features of objects, they have several limitations in the presence of challenges such as illumination changes, camouflage and shadows (See Section 4). In order to solve such issues, authors have also proposed to use other features like edge, texture and stereo features in addition to the color features.

– **Edge features:** The ambient light present in the scene can significantly affect the appearance of moving objects. However, spectral features, are limited by their ability to adapt to such changes in appearance. Thus, edge features emerged as a robust alternative for moving object detection. Edge features are generally computed using a gradient approaches such as Canny, Sobel [240][210][17][349][326] [285][234] or Prewitt [339][579] edge detector. It is commonly believed that edge features can handle local illumination changes, thus eliminating the chances of leaving ghosts when foreground objects begin to move. Despite some compelling advantages, edge features (high pass filters) tend to vary more than other comparable features based on low pass filters [349]. For example, edge features in the horizontal and vertical directions have different reliability characteristics, since textured objects have high values in both directions, whereas homogeneous objects have low values in both directions (See Section 5).

- **Texture features:** Texture features are appropriate to cope with illumination changes and shadows. Some common texture features that are generally used within this domain include the Local Binary Pattern (LBP) [207], and the Local Ternary Pattern (LTP) [332]. Numerous variants of LBP and LTP can be found in the literature as can be seen summarized in Table 5. Furthermore, statistical and fuzzy textures can be used as developed in Section 6.

- **Stereo features:** The extraction of stereo features rely on the need and use of specific acquisition systems such as a stereo, 3D, multiple, Time of Flight (ToF) cameras or RGB-D cameras (Microsoft Kinect[2], or Asus Xtion Pro Live[3]) to obtain the disparity information that usually represent the depth in the visual scene. It has become well-known that stereo features allow the model to deal with the camouflage in color [144][167][202][494][66][148][153] (See Section 7).

- **Motion features:** Motion features are usually obtained via optical flow but with the limitation of the computational time. Motion features allow the model to deal with irrelevant background motion and clutter [532][230][229][228][231][227][226] (See Section 8).

- **Local histogram features :** Local histograms are usually computed on color features [324][363][406][407][323][247][318][660][661][687]. But, local histograms can also be computed on edge features [252][145][218][386][415] to obtain Histograms of Oriented Gradients (HOG) (See Section 9).

- **Local histon features:** Histon [90] is a contour plotted on the top of the histograms of three primary color components of a region in a manner that the collection of all points falling under the similar color sphere of predefined radius, called similarity threshold, belongs to one single value. The similar color sphere is the region in RGB color space such that all the colors falling in that region can be classified as one color. For every intensity value in the base histogram, the number of pixels falling under similar color sphere is calculated, and this value is added to the histogram value to get the histon value of that intensity. Histon can be extended to 3D histon and 3D Fuzzy histon as developed by Chiranjeevi and Sengupta [90] (See Section 10).

- **Local correlogram features:** Correlogram was originally proposed for computer vision applications like object tracking [676]. Since, correlogram captures the inter-pixel relation of two pixels at a given distance, spatial information is obtained in addition to the color information. Thus, correlograms can efficiently alleviate the drawbacks of histograms, which only consider the pixel intensities for calculating the distribution. The main drawback of correlograms is their computation time due to their size of $256^3 \times 256^3$ in RGB, and 256×2563 in grey level. Hence, the single channel is quantized to a finite number of levels l. Due to this, the correlograms' size is further reduced to $l \times l$ with $l \ll 256$. Correlogram can be extended to fuzzy correlogram [87] and multi-channel fuzzy correlogram

[91] (See Section 11).

- **Haar-like features:** Some authors [293][294][195][675], used the Haar-like features [557]. Haar-like features are features defined in real-time face detector and based on the similarity with Haar wavelets. Haar-like features are computed from adjacent rectangular areas at a given location in a detection window by adding the pixel intensities in each area and by calculating the difference between these sums. The main advantage of Haar-like features is their computation speed. With the use of integral images, Haar-like features of any size can be computed in constant time (See Section Section 12).

- **Location features:** The location (x,y) can be used as a feature to exploit the dependency between the pixel [481][479][394][395][396] (See Section 14).

Table 3, Table 4, Table 5, Table 6 and Table 7 present an overview of the features in the pixel domain. Pixel domain features are generally robust and perform well provided more accurate representation of the visual scene is available. However, its high computational complexity restricts its real-time use in some applications. The features in the pixel domain are analyzed in details from Section 4 to Section 14.

2.2.2 Features in a Transform Domain

In order to accomplish some of the real-time demands of visual scene analysis, feature computation in a transform domain has gained importance.

- **Frequency domain:** The frequency domain offers a good framework to detect periodic processes that appear in dynamics backgrounds such as waving trees and waves in the ocean. For this, there is a need to transform the data values in the pixel domain into the frequency domain via a transformation such as the Fourier Transform (FFT) [591][547], Discrete Cosine Transform (DCT) [437], Wavelet Transform [158][175][156], Curvelet Transform [278], Walsh Transform [534][536][535], Hadamard Transform [30], Slant Transform [193] and Gabor Transform [607][609][588]. Practically, FTT processes blocks much faster in comparison with DCT. But, DCT outperforms slightly FFT in terms of precision, similarity and F-measure [555].

- **Video compressed domain:** As videos are usually compressed before transmission and storage, a number of compressed domain approaches have also been developed to improve the computational complexity of feature extraction. To obtain moving objects, the compressed video stream is partially decoded. Thus, the compressed domain data, such as motion vectors (MVs), transform coefficients, are employed to extract moving objects. Initially, compressed domain algorithms focussed attention on the MPEG standard. According to the classification of moving object detection, those methods can be mainly divided into three groups:

---

[2]http://www.microsoft.com/en-us/kinectforwindows/

[3]http://www.asus.com/Multimedia/

MVs based method [18], coefficients based method [645] [577], and combining MV and coefficients based methods [435] [436]. Other compressed domain algorithms used the video coding standard such as the H.264/AVC which process each video frame in units of a MacroBlock (MB) [541]. Thus, these MacroBlocks could be used as features. A Rate Distortion Cost (RDCost) value for each MB which changes depending on the frame content can be used as an indicator of changes. Typically, more cost will be spent on high motion and/or detailed MBs and less cost for low motion and/or homogenous MBs, which was verified in the experiments of [541]. The reasons for the effectiveness of the RDCost for foreground/background separation can be attributed to the following reasons. First, RDCost reflects the overall coding cost of a MB, which considers the effect of each factor on coding efficiency during video coding, such as prediction mode, MB partition size, motion vectors, residuals, etc. So RDCost can reflect true motion. Second, RDCost is in unit of MB. Only one MB has an RDCost value, as the basic coding unit in H.264/AVC is MB. Thus, RDCost is less affected by noise when compared with MVs [18]. Finally, compressed domain algorithms recently focused attention on the HEVC standard [674][673][73].

Table 5 shows an overview of the features in a transform domain. The features in the transform domain are analyzed in details in Section 15.

2.3 Classification by Intrinsic Properties

According to Li et al. [326], features can be classified by their intrinsic properties into the following categories:

– **Spectral features:** The intensity or color features are directly available from the images. Spectral features can easily detect changes if the difference in color between the foreground and the background are sufficient. However, spectral features produce *1)* false positive detections particularly when there are illumination changes, and *2)* false negative detections when foreground objects have similar color to the background (camouflage). Spectral features do not take into account or exploit the neighbourhood relationship of the considered pixel to deal with its poor robustness. Further, stereo features can also be considered as spectral features in the depth domain.

– **Spatial features:** Spatial features are edge and texture features. These features help to detect foreground objects that camouflage with the background and suppress shadows. Spatial features are however not applicable to non-stationary background objects at pixel level since the corresponding spatial features vary over time.

– **Temporal features:** Temporal features concern the motion between consecutive image frames. One way to obtain temporal features is to estimate the consistency

of optical flow over a short duration of time.

In order to achieve robust background modeling and foreground detection, features in each category are required to deal with a number of challenges commonly encountered in video surveillance as indicated in Bouwmans [51]. Table 1 shows an overview of the features classified following their intrinsic properties.

2.4 Classification by Mathematical Concepts

Some of the mathematical concepts that underlie during the computation of robust features can present an other useful categorization of features into crisp, statistical and fuzzy types.

– **Crisp Features:** Crisp features are those features which are computed without the use and need of any statistical or fuzzy concepts. It is the case of the color features (RGB, YUV, HSV, etc...), edge features obtained by a filter (Canny, Sobel, Prewitt), motion features obtained through optical flow or temporal operator, and stereo features.

– **Statistical Features:** Statistical features can be obtained by exploiting some of the statistical properties of the representation of the visual. The first work developed by Satoh et al. [462] proposed a Radial Reach Correlation (RRC) feature which has several variants: Bi-polar Radial Reach Correlation (BP-RCC) [457], Fast Radial Reach Correlation (F-RRC) [235][236], and Probabilistic Bipolar Radial Reach Correlation (PrBP-RCC) [631]. In a similar way, Yokoi [629] used Peripheral Ternary Sign Correlation (PTESC). Recently, Yoshinaga et al. [634][635] proposed the Statistical Local Difference Pattern (SLDP). The aim of these statistical features is to be more robust to illumination changes and dynamic backgrounds. Thus, SLDP integrates both pixel-based multi-modal model with color feature and spatial-based unimodal model with texture feature in one model feature taking into account the advantages of their respective robustness.

– **Fuzzy Features:** Fuzzy features are used to take into account the imprecision and the incertitude in features that represent a visual scene. For example, Chiranjeevi and Sengupta introduced fuzzy 3D Histons [90], fuzzy correlograms [87][91] and fuzzy statistical texture features [89][93]. The aim is to deal with illumination changes and dynamic backgrounds.

Table 2 shows an overview of the features classified following their mathematical concepts.

2.5 Exhaustive Overview of all the Features

The reader can refer to Table 3, Table 4, Table 5, Table 6 and Table 7 for an exhaustive overview of all the features. 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. Furthermore, prospective features not currently used for background modeling and foreground detection but in other computer vision applications are indicated in each table.

Fig.3 shows the distribution of research papers which used the corresponding type of features in their background modeling and foreground detection. It shows the domination of following types of features: **1)** *texture features* due their robustness in presence of shadows and gradual illumination changes, and sometimes in dynamic backgrounds, **2)** *features in a transform domain* due to the fact they allow to reduce computation time, and **3)** *multiple features* because it combines features from other categories and allow to combine their advantages.

For texture features, we can see in Fig. 4 that *local pattern textures* such as LBP and LTP are the most investigated. They are followed by *spatio-temporal patterns* and *statistical texture features*. *Fuzzy texture features* are less investigated although they appear to be very suitable and robust in presence of dynamic backgrounds and illumination changes as can be seen in the work of Chiranjeevi and Sengupta [89].

This exhaustive overview shows the activity and the importance of the research on features in the field of background modeling and foreground detection. To have a quantitative view of the activity, we have counted the number of papers in terms of publication in conferences and journals. From Fig. 2, we can see how the research focus on the development of background modeling and foreground detection methods has grown in the last 20 years. However, we can note that compared to the number of papers (more than 15000 in the field), this field is less investigated although the role and the importance of the features is primordial in the robustness of the methods against challenges as developed in Section 2.6.5.
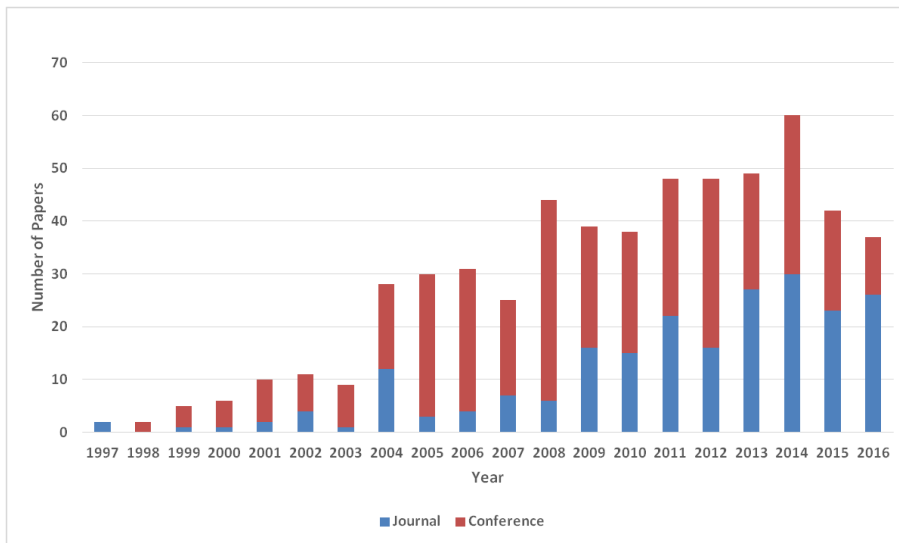
Fig. 2 Number of papers by years. Note that the ratio between journal and conference publications increased in favor of journal publications, and thus testifies the importance of this field.
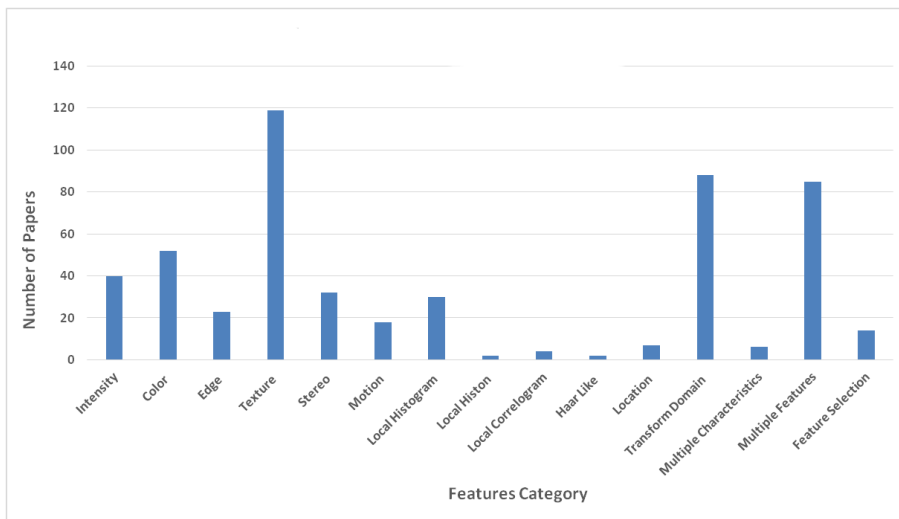

Fig. 3 Number of papers by type of features: Texture features, features in a transform domain, multiple features are the most predominant features investigated.
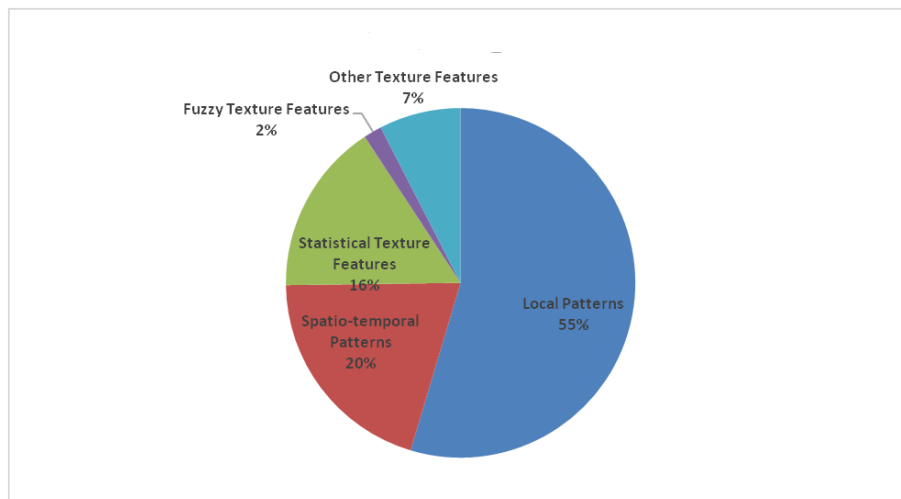
Fig. 4 Number of papers by type of textures. Local pattern textures, spatio-temporal patterns and statistical texture feature are the most investigated.

Spectral Features	Spatial Features	Temporal Features	Spectral/Spatial Features	Spatial/Temporal Features
Color features	**Edge Features**	**Motion Features**	**Local Histogram**	**Texture Features**
RGB, Normalized RGB	Canny, Sobel, Prewitt	Optical Flow	LCH,LK-CH	STLBP, *SCS-LBP*, ST-CS-LBP
YUV, HSV	**Texture Features**	Temporal Operator	HOG, LOH, LK-HOG	MV-LBP, SLBP-AM
HSI, Luv	PISC, PTESC		LDH, LHFG	CS-ST-LTP, ST-SITLP
Improved HLS, Ohta	LBP, ELBP		**Texture Features**	CS-SILTP
YCrCB, Lab, Lab2000HL	CS-LBP,HCS-LBP		SC-LBP, *SCS-LBP*, OC-LBP	
	DLBP, ULBP		iLBP	
	ExtLBP, RI-LBP		MC-SITLP	
	LN-LBP, SALBP		**Frequency Features**	
	WB-LBP		FFT, DCT, WT	
	LTP, SILTP		HT, GT	
	SCS-LTP		**Video Compressed Features**	
	SILS		MPEG (MVs, Coefficients, Mixed)	
	Stereo features		H.264/AVC (MBs)	
	Disparity, Depth		HEVC (MBs)	

Table 1 Classification by Intrinsic Properties: An Overview.

2.6 Features and Their Usage

Different features have different properties that enable them to handle critical situations such as illumination changes, motion changes and structure background changes, differently. Therefore there is a need to characterize features so that benchmarking becomes possible. In this context, reliability has been considered. Furthermore, each feature can be characterized by its reliability (Section 2.6.1). If more than one feature is required to be used, then there is a need for fusion schemes that can aggregate the results of each one (Section 2.6.2). Finally, feature selection can be used to optimize the discrimination between the background and foreground classes (Section 18).

Crisp Features	Statistical Features	Fuzzy Features
Color features	**Intensity features**	**Color Coherence Vector**
RGB, Normalized RGB	Rank Order Statistic [598]	FCC [438][438]
YUV, HSV, HSI	OSID [531]	**Local Color Histogram**
Luv, Improved HLS	**Color features**	LFCH [290][272][299][624][190]
Ohta, YCrCb, Lab, Lab2000HL	Standard Variance Feature [685]	LFCDH [416]
Edge features	Color co-occurences [326]	**Local Gradient Histogram**
Canny [249], Sobel [240], Prewitt [339]	Entropy [355][262][76]	HOFG [455]
Motion Features	IISC [291]	**Local Histon**
Optical Flow, Temporal Operator [682]	NCC [428]	3D Fuzzy Histon [90]
Stereo Features	**Edge features**	**Local Correlogram**
Disparity [237], Depth [202]	Gradient Deviations [269]	Fuzzy Correlogram [87]
LCH [363], LK-CH [406]	Projection Gradient Statistics [656]	Multi-channel Kernel FC [91]
LCH [363], LK-CH [406]	**Texture features**	**Local Fuzzy Pattern** [412]
LCH [363], LK-CH [406]	PISC [459], PETSC [629]	**Fuzzy Statistical Texture**
HOG [252], LOH [247], LK-HOG [407]	RRC [462], BP-RCC [457]	FST [89][93]
LDH [660], LHFG [687]	F-RRC [235], PrBP-RCC [631]	
Local Histon	ABP-RCC [374]	
Histon [90], 3D Histon [90]	RRF [464]	
Local Correlogram	RPF [375], MRPF [376]	
Correlogram [676]	SRF [292]	
Location [481]	ST [92]	
Haar like Features [293]		

Table 2 Classification by Mathematical Concepts: An Overview.

2.6.1 Feature Reliability

Although efforts have been made to perform reliable foreground detection, in reality, results are not 100% reliable. This is mainly because, applications are often considered in controlled environments, where closed world assumptions could be applied. However, in general, one has to deal with a real-world scenario, which means that the background or foreground may significantly change dynamically making it literally impossible to predict possible changes. For example, the light conditions may suddenly fluctuate in some parts of the image, video compression or transmission artefacts may cause noise, a wind may cause a stationary camera to tremble, and so on. The fundamental problem lies in building an appropriate and robust background modeling that is capable of accurate detect moving objects without compromising model against background changes, and thus fail to detect all changes.

One potential solution is in monitoring the reliability of the features by analyzing their general properties. For example Latecki et al. [430] considered statistical properties of feature value distributions as well as temporal properties as a source of tracking feature reliability. The proposed strategy is to estimate the deviations of the feature to the distribution and if found reliable then compute feature properties, else detect as being unreliable. As computed features will never be 100% reliable, it is interesting to compute reliability measures. This way decisions will only be made when features are sufficiently reliable. This means that in addition to feature computation, an instantaneous evaluation of their reliability should also be made, and then adapt the decision in accordance to the detected level of reliability. For example, if the goal of the application is to monitor motion activity, and to signal an alarm if the activity

Pixel Domain Features	Categories	Authors - Dates
Intensity features	**1) Well-Known Intensity Features**	
	Visible cameras	
	Intensity (4)	Silveira et al. (2005) [493]
	IR cameras	
	Intensity (5)	Davis and Sharma (2004) [320]
	2) Designed Illumination Invariant Features	
	Illumination Ratio (1)	Paruchuri et al. (2011) [423]
	Brightness (1)	Wang et al. (2015) [564]
	Reflectance (7)	Toth et al. (2000) [544]
	Surface Spectral Reflectance (SSR) (1)	Sedky et al. (2014) [469]
	Radiance (1)	Xie et al. (2004) [598]
	Photometric Variations (1)	Di Stefano et al. (2007) [513]
	3) Statistical Intensity Features	
	Rank Order Statistics (1)	Xie et al. (2004) [598]
	Gray level Co-occurrence Matrix (GLCM) (2)	Qu et al. (2010) [439]
	Local GLCM (1)	Subudhi et al. (2016) [516]
	4) Prospective Intensity Features	
	Ordinal Spatial Intensity Distribution (OSID)	Tang et al. (2014) [531]
Color features	**1) Well-Known Color Spaces**	
	RGB	Stauffer and Grimson (1999) [512]
	Normalized RGB	Xu et Ellis (2001) [601]
	YUV	Harville et al. (2001) [202]
	HSV	Sun et al. (2006) [520]
	HSI	Wang and Wu (2006) [576]
	Luv	Yang and Hsu (2006) [622]
	Improved HLS	Setiawan et al. (2006) [471]
	Ohta	Zhang and Xu (2006) [654]
	YCrCb	Baf et al. (2008) [655]
	YIQ	Thangarajah et al. (2015) [538]
	Lab	Balcilar et al. (2013) [27]
	Lab2000HL	Balcilar et al. (2013) [27]
	CIEDE2000	Thangarajah et al. (2016) [538]
	2) Designed Shape Color Spaces	
	Cylinder Color (CY) (5)	Horprasert et al. (1999) [212]
	Hybrid Cone-Cylinder Color (2)	Doshi and Trivedi (2006) [131]
	Arbitrary Cylinder Color (ACY) (2)	Zeng and Jia (2014) [646]
	Ellipsoidal Color (EC)(1)	Sun et al. (2011) [519]
	Box-based Color (1)	Tu et al. (2008) [549]
	Cubic Color (1)	Noh and Jeon (2011) [403]
	Spherical Color (1)	Hu et al. (2012) [216]
	Double-Trapezium Cylinder Color (DTC)	Huang et al. (2015) [222]
	Conical Color	-
	3) Designed Illumination Invariant Color Features	
	Color Illumination Ratio (2)	Pilet et al. (2008) [428]
	RGB-HSV (1)	Takahara et al. (2012) [522]
	RGB-YCrCb (1)	Sajid and Cheung (2014) [454]
	Color Illumination Invariant (1)	Yeh et al. (2012) [627]
	4) Color Filter Array (CFA) Features	
	CFA patterns (1)	Suhr et al. (2011) [518]
	Bayer CFA patterns (1)	Suhr et al. (2011) [518]
	5) Statistics on Color Features	
	Standard Variance Feature (1)	Zhong et al. (2010) [685]
	Color co-occurrences (2)	Li et al. (2004) [326]
	Entropy (5)	Ma and Zhang (2001) [355]
	IISC (1)	Kim and Kim (2016) [291]
	Normalized Cross Correlation (NCC) (2)	Pilet et al. (2008) [428]
	6) Multiscale Color Features	
	Multiscale Color Description (1)	Muchtar et al. (2011) [385]
	7) Fuzzy Color Features	
	Fuzzy Color Coherence Vector (1)	Qiao et al. (2014) [438]
	7) Prospective Color Features	
	Scale-Invariant Feature Transform (SIFT)	-
	Speeded Up Robust Features (SURF) (1)	Shah et al. (2014) [473]
	Weber Local Descriptor (WLD)	Chen et al. (2009) [77]
Multispectral features	Multi-spectral features (3)	Benezeth et al. (2014) [35]
Edge Features	**1) Crisp Edge**	
	Gradient (Magnitude/Direction) (3)	Javed et al. (2002) [249]
	Sobel Edge Detector (7)	Jabri et al. (2000) [240]
	Prewitt Edge Detector (2)	Lindstrom et al. (2006) [339]
	Local Hybrid Pattern (LHP) (1)	Kim et al. (2015) [285]
	Local Directional Number (LDN) (1)	Roy et al. (2017) [452]
	2) Statistical Edge	
	Gradient Deviations (1)	Kamkar-Parsi (2005) [269]
	Projection Gradient Statistics (1)	Zhang (2012) [656]

Table 3 Features in the Pixel Domain: An Overview (Part 1).

Pixel Domain Features	Categories	Authors - Dates
Motion Features	Optical Flow (9)	Tang et al. (2007) [532]
	Temporal Operator (4)	Zhong et al. (2008) [682]
	SIFT flow (1)	Dou and Li (2014) [133]
	Magno channel (1)	Martins et al. (2016) [362]
Stereo Features	**1) Disparity**	
	Disparity (5)	Ivanov et al. (1997) [237]
	Variational Disparity (1)	Javed et al. (2015) [250]
	2) Depth	
	Depth (Stereo System) (4)	Harville et al. (2001) [202]
	Depth (ToF) (6)	Silvestre (2007) [494]
	Depth (Microsoft Kinect) (13)	Camplani and Salgado (2013) [66]
Local Histograms Features	**1) Local Histograms of Color**	
	1.1) Crisp Local Histograms of Color	
	Local Color Histogram (LCH) (1)	Mason and Duric (2001) [363]
	Local Kernel Color Histograms (LK-CH) (1)	Noriega et al. (2006) [406]
	Estimated Local Kernel Histogram (ELKH) (1)	Li et al. (2008) [323]
	Local Color Difference Histograms (LDCH) (1)	Li (2009) [324]
	Local Dependency Histograms (LDH) (2)	Zhang et al. (2008) [660]
	Spatiotemporal Condition Information (SCI)(1)	Wang et al. (2014) [562]
	1.2) Fuzzy Local Histogram of Color	
	Local Fuzzy Color Histograms (LFCH) (6)	Kim and Kim (2012) [290]
	Local Fuzzy Color Difference Histograms (LFCDH) (1)	Panda et al. [416]
	2) Local Histograms of Gradient	
	2.1) Crisp Local Histograms of Gradient	
	Local Histogram on Gradient (LGH) (1)	Mason and Duric (2001) [363]
	Local Histogram of Oriented Gradient (L-HOG) (3)	Fabian (2010) [145]
	Local Adaptive HOG (LA-HOG) (1)	Hu et al. (2010) [218]
	Local Orientation Histograms (LOH)(1)	Jang et al. (2008) [247]
	Local Kernel Histograms of Oriented Gradients (LK-HOG) (1)	Noriega and Bernier (2006) [407]
	Prospective Histograms	
	2.2) Fuzzy Local Histograms of Gradient	
	Local Histograms of Fuzzy Gradient (HOFG) (1)	Salhi et al. (2013) [455]
	3) Local Histograms of Figure/Ground	
	Local Histogram of Figure/Ground (LHFG) (1)	Zhong et al. (2009) [687]
Local Histon Features	**1) Crisp Local Histon**	
	Histon (1)	Chiranjeevi and Sengupta (2012) [90]
	3D Histon (1)	Chiranjeevi and Sengupta (2012) [90]
	2) Fuzzy Local Histon	
	3D Fuzzy Histon (2)	Chiranjeevi and Sengupta (2012) [90]
Local Correlogram Features	**1) Crisp Local Correlogram**	
	Correlogram	Zhao and Tao (2005) [676]
	2) Fuzzy Local Correlogram	
	Fuzzy Correlogram (FC) (1)	Chiranjeevi and Sengupta (2011) [87]
	Multi-channel Kernel Fuzzy Correlogram (MKFC) (1)	Chiranjeevi and Sengupta (2013) [91]
Location Features	Location Features (6)	Sheikh and Shah (2005) [481]
	Invariant Moments (1)	Marie et al. (2011) [361]
Haar-like Features	Haar-like features (7)	Klare (2008) [293]
Haralick Features	Haralick features (1)	Subudhi et al. (2016) [516]

Table 4 Features in the Pixel Domain: An Overview (Part 2).

is high, the system is allowed to make reliable decisions only if there exist a subset of the computed motion activity features that is sufficiently reliable. The monitoring of features reliability and adjusting the system behaviour accordingly, seems to be the best mechanism to deploy autonomous video surveillance systems.

To determine whether a particular feature is reliable, Latecki et al. [430] assumed that the feature bears more information if its distribution differs more significantly from a normal (Gaussian) distribution. The assumption is that the feature becomes unreliable if an addition random noise is superimposed, which would lead the distribution of such noisy feature to become more Gaussian like. Hence, by measuring to what extent a feature distribution differs from a Gaussian distribution, it would be possible to get an notion of how useful the feature is and importantly detect if such

Transform Domain Features	Categories	Authors - Dates
Frequency Features	**Fourier Transform (FT)(2)**	
	Fast Fourier Transform (FFT)(1)	Wren and Porikli (2005) [591]
	Discrete Fourier transform (2D-DFT)(1)	Tsai and Chiu (2008) [547]
	Discrete Cosine Transform (DCT) (10)	Porikli and Wren (2015) [437]
	Wavelet Transform (WT) (28)	
	Discrete Wavelet Transform (DWT) (3)	Huang and Hsieh (2003) [221]
	Binary Discrete Wavelet Transform (BDWT) (4)	Gao et al. (2008) [158]
	Modified directional lifting-based 9/7 DWT (MDLDWT)(1)	Hsia and Guo (2014) [215]
	Orthogonal non-separable Wavelet (OW) (1)	Gao et al. (2008) [156]
	Wavelet multi-scale Transform (2D dyadic WT) (5)	Guan et al. (2008) [175]
	Daubechies Complex Wavelet Transform (DCWT)(9)	Jalal and Singh (2011) [245]
	Multi-Resolution Wavelet Transform (MRWT) (1)	Mendizabal and Salgado (2011) [365]
	Undecimated wavelet transform (2D-UWT) (3)	Antic et al. (2009) [15]
	Three-Dimensional Discrete Wavelet Transform (3D-DWT) (1)	Han et al. (2016) [197]
	Curvelet Transform (CT) (1)	Khare et al. (2013) [278]
	Walsh Transform (WalshT) (3)	Tezuka and Nishitani (2008) [534]
	Hadamar Transform (HT) (1)	Baltieri et al. (2010) [30]
	Gabor Transform (GT) (3)	Xue et al. (2010) [607]
	Slant Transform (ST) (2)	Haberdar and Shah (2013) [193]
	Prospective Frequency Features	
	Sparse FFT	-
Video Compressed Features	**MPEG domain**	
	MVs based Features (1)	Babu et al. (2004) [18]
	Coefficient based Features (2)	Zeng et al. (2003) [645]
	MVs and Coefficient based Features (2)	Porikli (2014) [435]
	H.264/AVC domain	
	MacroBlock based Features (MB) (2)	Dey and Kundu (2013) [124]
	Enhanced MacroBlock based Features (EMB) (1)	Dey and Kundu (2016) [125]
	HEVC domain	
	MVs based Features (3)	Zhao et al. (2013) [674]
Compressive Features	**Compressive measurements**	
	Orthonormal basis (4)	Cevher et al. (2008) [70]
	Linear compressive measurements (5)	Needell and Tropp (2008) [397]
	Cross-validation measurements (ARCS-CV) (2)	Warnell et al. (2012) [585]
	Low-resolution measurements (ARCS-LR) (1)	Warnell et al.(2012) [585]
	Orthonormal wavelet basis (1)	Li et al.(2010) [325]
	Wavelets transform (5)	Wang et al. (2015) [581]
	Canonical sparsity basis (1)	Xu and Lu (2011) [602]
	Random projections basis (1)	Shen et al. (2016) [482]
	Walsh-Hadamard measurements (1)	Liu and Pados (2016) [345]
	Random Gaussian or Fourier Scrambled matrices (1)	Li and Qi (2014) [328]
	Three-dimensional Compressive Sampling (3DCS) (2)	Shu and Ahuja (2011) [488]
	Three-Dimensional Circulant CS (3DCCS) (2)	Kang et al. (2015) [271]

Table 5 Features in a Transform Domain: An Overview.

usefulness drops. Latecki et al. [430] proposed an entropy-based technique for feature reliability assessment. The proposed parametric negentropy estimation, inspired from information theory, can be efficiently used to evaluate the usability of a motion measure employed for the detection of moving objects. This approach is useful for one-dimensional features and works under assumption that the noise, which corrupts the observed feature, is additive Gaussian. Some future extensions of this work could be to generalize it for the multidimensional case and for the feature reliability detection with non-Gaussian and non-Additive noise.

In an an other study, Latecki et al. [310][314][315] described a simple temporal method to determine the reliability of motion features. The input motion feature has binary values for each 8×8 block with 1 for "motion detected" and 0 for "no mo-

Textures	Methods-Acronym-Number of papers)	Authors - Dates
1) Crisp Textures (Part 1)		
Local Binary Pattern (25 variants)	Local Binary Pattern (LBP) (9)	Heikkila et al. (2004) [207]
	Spatio-temporal Local Binary Pattern (STLBP) (2)	Zhang et al. (2008) [659]
	Epsilon Local Binary Pattern (ϵ-LBP) (2)	Wang and Pan (2010) [568]
	Center-Symmetric Local Binary Patterns (CS-LBP)(3)	Tan et al. (2010) [523]
	Space-Time Center-Symmetric Local Binary Patterns (ST-CS-LBP) (1)	Li et al. (2011) [322]
	Spatial Extended Center-Symmetric Local Binary Pattern (SCS-LBP) (1)	Xue et al. (2010) [608]
	Hybrid Center-Symmetric Local Binary Pattern (HCS-LBP) (1)	Xue et al. (2011) [606]
	Spatial Color Binary Patterns (SCBP) (1)	Zhou et al. (2011) [692]
	Opponent Color Local Binary Patterns (OCLBP) (1)	Lee et al. (2011) [316]
	Double Local Binary Pattern (DLBP) (1)	Xu et al. (2009) [600]
	Uniform Local Binary Patterns (ULBP) (1)	Yuan et al. (2011) [636]
	Extended Local Binary Patterns (Ext-LBP) (1)	Yu et al. (2011) [637]
	Rotation Invariant Local Binary Patterns (RI-LBP) (1)	Yu et al. (2011) [637]
	Larger Neigborhood Local Binary Patterns (LN-LBP) (1)	Kertesz (2011) [275]
	Motion Vectors Local Binary Patterns (MV-LBP) (3)	Yang et al. (2012) [617]
	Scene Adaptive Local Binary Pattern (SALBP) (1)	Noh and Jeon (2012) [404]
	Stereo Local Binary Pattern based on Appearance and Motion (SLBP-AM) (1)	Yin et al. (2013) [628]
	Window-Based LBP (WB-LBP) (1)	Kumar et al. (2014) [300]
	Intensity Local Binary Pattern (iLBP)(2)	Vishnyakov et al. (2014) [558]
	BackGround Local Binary Pattern (BGLBP) (1)	Davarpanah et al. (2015) [111]
	eXtended Center-Symmetric Local Binary Pattern (XCS-LBP) (1)	Silva et al. (2015) [490]
	Uniform XCS-LBP (UXCS-LBP) (1)	Du and Qin (2016) [134]
	Local SVD Binary Pattern (LSVD-BP)(1)	Guo et al. (2016) [187]
	Multi-Block Temporal-Analyzing LBP (MB-TALBP) (1)	Chen et al. (2016) [80]
	Perception-based Local Binary Pattern (P-LBP) (1)	Chan. (2016) [75]
	Extended Symmetrical-Diagonal Hexadecimal Pattern (ES-DHP)) (1)	Jeyabharathi and Dejey (2017) [256]
	Prospective LBP	
	Multi-scale Region Perpendicular LBP (MRP-LBP) (1)	Nguyen and Miyata (2015) [399]
	Scale-and Orientation Adaptive LBP (SOA-LBP) (1)	Hegenbart and Uhl (2015) [205]
Local Ternary Pattern (7 variants)	Local Ternary Pattern (LTP) (1)	Liao et al. (2010) [332]
	Scale Invariant Local Ternary Pattern (SILTP) (1)	Liao et al. (2010) [332]
	Scale-invariant Center-symmetric Local Ternary Pattern(SCS-LTP) (2)	Zhang et al. (2011) [670]
	Multi-Channel Scale Invariant Local Ternary Pattern (MC-SILTP) (1)	Ma and Sang (2012) [354]
	Center Symmetric Spatio-temporal Local Ternary Pattern (CS-ST-LTP) (2)	Xu (2013) [605]
	Spatio Temporal Scale Invariant Local Ternary Pattern (ST-SILTP) (1)	Ji and Wang (2014) [258]
	Center-Symmetric Scale Invariant Local Ternary Pattern (CS-SILTP) (1)	Wu et al. (2014) [592]
Local States Pattern	Scale Invariant Local States (SILS) (1)	Yuk et al. (2011) [638]
Local Derivative Pattern	Center-Symmetric Local Derivative Pattern (CS-LDP) (1)	Xue et al. (2011) [606]
	Uniform CS-LDP (UCS-LDP) (1)	Du and Qin (2016) [134]
	SpatioTemporal Center-Symmetric Local Derivative Pattern (STCS-LDP) (1)	Jmal et al. (2016) [263]
Local Difference Pattern	Local Difference Pattern (LDP) (1)	Yoshinaga et al. (2010) [633]
	Statistical Local Difference Pattern (SLDP) (2)	Yoshinaga et al. (2011) [634]
Local Self Similarity	Local Self Similarity (LSS) (1)	Jodoin et al. (2012) [264]
	Local Similarity Statistical Descriptor (LSSD) (1)	Zeng et al. (2017) [644]
Local Similarity Binary Pattern	Local Similarity Binary Pattern (LSBP) (6)	Bilodeau et al. (2013) [43]
	Uniform LSBP (U-LBSP) (1)	Yan et al. (2016) [613]
	Color LSBP (C-LBSP) (1)	Zhang et al. (2017) [658]
Directionnal Rectangular Pattern	Directionnal Rectangular Pattern (DRP) (1)	Zhang et al. (2009) [649]
Local Color Pattern	Local Color Pattern (LCP) (3)	Chua et al. (2012) [95]
Local Neigborhood Pattern	Local Neigborhood Pattern (LNP) (1)	Amato et al. (2010) [10]
Local Ratio Pattern	Local Ratio Patterns (LRP) (1)	Zaharescu and Jamieson (2011) [641]
Local Ray Pattern	Local Ray Pattern (LRP) (2)	Shimada et al. (2013) [484]
Spatio-Temporal Vector	Spatio-Temporal Vector (STV) (11)	Pokrajac and Latecki (2003) [429]
Spatio-Temporal Texture	Space-Time Patch (ST-Patch) (2)	Yumiba et al. (2011) [639]
Spatio-Temporal Features	Spatio-Temporal Features (STF) (3)	Nonaka et al. (2012) [530]

Table 6 Features in the Pixel Domain (Texture Features): An Overview (Part 3).

Textures	Methods-Acronym-Number of papers	Authors - Dates
1) Crisp Textures (Part 2)		
Texture Pattern Flow	Texture Pattern Flow (TPF) (3)	Zhang et al. (2011) [648]
Binary Map	Binary Map (BM) (1)	Lai et al. (2013) [306]
Textons	Textons (T) (2)	Spampinato et al. (2014) [504]
Galaxy Pattern	Galaxy Pattern (GP) (2)	Liu et al. (2013) [344]
Bayer-Pattern	Bayer-Pattern (BP) (1)	Suhr et al. (2011) [518]
Structure-Texture Decomposition	Structure-Texture Decomposition (STD) (2)	Elharrouss et al. (2015) [143]
2) Statistical Textures		
Peripheral Increment Sign Correlation	Peripheral Increment Sign Correlation (PISC) (1)	Satoh et al. (2004) [459]
Peripheral TErnary Sign Correlation	Peripheral TErnary Sign Correlation (PTESC) (2)	Yokoi (2006) [629]
Radial Reach Correlation	Radial Reach Correlation (RRC) (7)	Satoh et al. (2002) [462]
	Bi-Polar Radial Reach Correlation (BP-RCC) (2)	Satoh (2005) [457]
	Fast Radial Reach Correlation (F-RRC) (2)	Itoh et al. (2008) [235]
	Probabilistic Bi-Polar Radial Reach Correlation (PrBP-RCC) (1)	Yokoi (2009) [631]
	Adaptive Bi-Polar Radial Reach Correlation (ABP-RCC) (1)	Miyamori et al. (2011) [374]
Radial Reach Filter	Radial Reach Filter (RRF) (5)	Satoh et al. (2002) [464]
Radial Proportion Filter	Radial Proportion Filter (RPF) (1)	Miyamori et al. (2012) [375]
	Multi Radial Proportion Filter (MRPF) (1)	Miyamori et al. (2012) [376]
Statistical Reach Feature	Statistical Reach Feature (SRF) (2)	Iwata et al. (2009) [292]
Statistical Texture	Statistical Texture (ST) (2)	Chiranjeevi and Sengupta (2014) [92]
3) Fuzzy Textures		
Local Fuzzy Pattern	Local Fuzzy Pattern (LFP) (2)	Ouyang et al. (2012) [412]
4) Fuzzy Statistical Textures		
Fuzzy Statistical Texture	Fuzzy Statistical Texture (FST) (2)	Chiranjeevi and Sengupta (2012) [89]

Table 7 Features in the Pixel Domain (Texture Features): An Overview (Part 5).

tion detected". Let $f(n)$ be the number of 1s in the frame number n, i.e., $f(n)$ is the number of detected moving blocks as function of frame number. The finite difference approximation of first derivative of f is used to monitor the reliability of the motion feature. If the jump in values of f is above a certain threshold for a given time interval, the binary feature is unreliable in this interval. The threshold necessary to detect the unreliable features is not static and is determined by a dynamic thresholding algorithm.

Harville et al. in [144] detected invalid depth and chroma components in the manner as follows: *1)* The chroma (U and V) components become unstable when the luminance is low. So, Harville et al. [144] defined a chroma validity test based on luminance and set a predicate that operated it on a Gaussian model by applying it to the luminance mean of the distribution. When one of the test is failed, the chroma components of the current observation or the Gaussian distribution are not used, and *2)* The depth computation relies on finding small area correspondences between image pairs, and therefore does not produce reliable results in regions of little visual texture and in regions, that are visible in one image but not the other. Most stereo depth implementations attempt to detect such cases and label them with one or more special values. Harville et al. [144] relied on the stereo depth system to detect invalid depth data and define a depth validity predicate. When the test is false, the depth is not used.

In an other work, Mittal et al. [373] briefly described methods that might be used

for obtaining measurements of normalized RGB and optical flow features, and their associated uncertainties. But, the uncertainties were finally not used in the model.

According to the literature in this area, feature reliability has been less investigated and only the works of Latecki et al. [430][310][314][315] specifically addressed this problem. Thus, the determination of the features' reliability and then when to use them is still an open problem and may be one of the main future developments in this field.

2.6.2 Feature Fusion

It has been discussed that the use of multiple features could bring complementary advantages to the modeling and detection techniques. However, the feature aggregation procedure requires reliable operators that can efficiently fuse features. There are several operators which can be used for feature aggregation. These operators can either be of basic (logical AND, logical OR), statistical or fuzzy types (Sugeno integral [654], Choquet integral [21], interval valued Choquet integral [93]) as follows:

1. **Basic operators:** Logical operators such as OR and AND are the simplest ways to combine different results. AND is more useful when the aim is to suppress false positive detections which appear in one mask and not in the other. But it presents the disadvantage to suppress true positives which appear in one mask and not in the other one. On the other hand, OR is more suitable when the aim is to suppress false negative detections which appear in one mask and not in the other one. But it presents the disadvantage to add false positive which appear in one mask and not in the other one. Other basic operators include mean, median, minimum and maximum operators, as well as some generalizations like the Ordered Weighted Average (OWA) having the minimum and the maximum as particular cases.

2. **Statistical operators:** The simplest way to statistically combine features consists of a basic product formulation of the likelihoods [670][326] but it has the limitation that a single close to zero probability in one of the sensors may lead to the cancellation of the overall combination. In order to avoid the zero probability problem, which could lead to critical misclassification errors, Logarithmic Opinion Pool can be used as proposed by Gallego and Pardas [153]. Thus, by taking logarithms, a weighted average of the log-likelihoods can be obtained. The weighting factors is central to the correct working of the sensor fusion system. The weighting factors can be according to the reliability that each one of the sensors presents. An an other possible way is to exploit the similarity between foreground and background classes for each one of the sensors, assuming that *(1)* high similarity implies that both classes are modeling the same space in a camouflage situation, and thus, the decision is not reliable, and *(2)* small similarity implies classes separated enough to achieve a correct decision. In this idea, Gallego and Pardas [153] computed the Hellinger distance between the pdf's that model the ith pixel of the jth-sensor color or depth for the foreground and background classes, respectively. Thus, this distance detects the degree of similarity

between foreground and background models that each one of the sensors present. Moreover, the Hellinger distance presents two main characteristics that are very interesting for the application of background/foreground separation: Unlike the Bhattaharyya distance, or the Kullback-Leibler divergence, which give a similarity distance bounded between $[0, \infty)$, the Hellinger distance gives a normalized distance among models bounded between $[0, 1]$. Furthermore, unlike the Kulback-Leibler divergence, the Hellinger distance is symmetric. Finally, the weighting factors are defined as a function of the Hellinger distance. Thus, sensors that present a higher degree of similarity between foreground and background classes have a close-to-zero weight, thus avoiding misclassification errors in case of color or depth camouflage problems.

3. **Fuzzy operators:** The family of fuzzy integrals is a generalization of the weighted average technique using the Choquet integral, as well as the minimum and the maximum using the Sugeno integral. The advantage of fuzzy integrals is that they take into account the importance of the coalition of any subset of criteria. A brief summary of the basic concepts around fuzzy integrals (Sugeno and Choquet) is described below:

   – **Sugeno and Choquet integrals:** Let μ be a fuzzy measure on a finite set X of criteria and $h : X \to [0, 1]$ be a fuzzy subset of X.

   Definition 1 *The Sugeno integral of h with respect to μ is defined by:*

   $$S_\mu = Max\left(Min\left(h\left(x_{\sigma(i)}\right), \mu\left(A_{\sigma(i)}\right)\right)\right) \tag{1}$$

   where σ is a permutation of the indices such that
   $h_{\sigma(1)} \leq \ldots \leq h_{\sigma(n)}$ *and* $A_{\sigma(i)} = \{\sigma(1), \ldots, \sigma(n)\}$
   Definition 2 *The Choquet integral of h with respect to μ is defined by:*

   $$C_\mu = \sum_{i=0}^{n} h\left(x_{\sigma(i)}\right)\left(\mu\left(A_{\sigma(i)}\right) - \mu\left(A_{\sigma(i+1)}\right)\right) \tag{2}$$

   with the same notations as above.

   An interesting interpretation of the fuzzy integrals arises in the context of source fusion. The measure μ can be viewed as a factor that describes the relevance of the sources of information where h denotes the values that the criteria has reported. The fuzzy integrals then aggregates nonlinearly the outcomes of all criteria. The Choquet integral is adapted for cardinal aggregation while Sugeno integral is more suitable for ordinal aggregation. Additionally, Sugeno integral calculates only minimum and maximum weightage and the Choquet integral has the same functionality as Sugeno integral but it also uses additional operations like arithmetic mean and Ordered Weighted Averaging (OWA). Thus, Choquet integral is more suitable for background/foreground separation.

– **Fuzzy measures:** While fusing different criteria or sources, fuzzy measures take on an interesting interpretation. A pixel can be evaluated based on a criteria or sources providing information about the state of the pixel whether pixel corresponds to background or foreground. The more the criteria provides information about the pixel, the more relevant the decision of pixel's state. Let $X = \{x_1, x_2, x_3\}$, with each criterion, a fuzzy measure is associated, $\mu(x_1) = \mu(\{x_1\})$, $\mu(x_2) = \mu(\{x_2\})$ and $\mu(x_3) = \mu(\{x_3\})$ such that the higher the $\mu(x_i)$, the more important the corresponding criterion in the decision. To compute the fuzzy measure of the union of any two disjoint sets whose fuzzy measures are given, an operational version proposed by Sugeno which called λ-*fuzzy measure* can be used. To avoid excessive notation, let denote this measure by μ_λ-*fuzzy measure*, where λ is a paramater of the fuzzy measure used to describe an interaction between the criteria that are combined. Its value can be determined through the boundary condition, i.e. $\mu(X) = \mu(\{x_1, x_2, x_3\}) = 1$. The fuzzy density values over a given set $K \subset X$ is computed as:

$$\mu_\lambda(K) = \frac{1}{\lambda}\left[\prod_{x_i \in K}(1 + \lambda\mu_\lambda(x_i)) - 1\right] \tag{3}$$

– **Interval-valued fuzzy integrals:** Although discrete (or real-valued) Sugeno and Choquet integrals defined in Equation 1 and Equation 2 can be used as a decision making operator to fuse the information from multiple sources, it does not consider the uncertainty. But in practice information sources have wide range of possible values (i.e., high uncertainty) and hence cannot be represented by a single number. To solve this problem, interval-valued fuzzy sets (IVFSs) [248][621][652] can be used to model the uncertainty in these values. Thus, the value are represented as an interval. In this context, an aggregation operator is needed to integrate the information sources, represented by IVFSs. Thus, modifications of the discrete integrals called interval-valued integrals are used as in Chiranjeevi and Sengupta [93].

Dempster-Shafer theory can be also used in feature fusion as in the work of Munteanu et al. [388]. According to the literature in this area, these different feature fusion schemes have been applied when multiple features are used as can be seen in Section 17 and Table 12.

2.6.3 Feature Selection

As seen in Section 2.2.1, there is not a unique feature that performs better than any other feature independently of the background and foreground properties because each feature has its strengths and weaknesses against each challenge. Thus, a way to take advantage of the properties of each feature is to perform feature selection. The aim is to use the best feature or the best combination of features on a per-pixel

[170][172][171][417][491][492] or per-block [316] basis. A set of feature could be composed of **(1)** homogeneous features that are feature from the same category, and then the idea is to reinforce the reliability fo the concerned type of features, or **(2)** heterogeneous features to complement the features each others [567]. Once the set of features is determined, ensemble learning methods such as the boosting classifier can be used for feature selection. Boosting algorithms usually generate a weighted linear combination of some weak classifiers that perform only a little better than random guess. So, weak classifiers can be learned from the feature values at a pixel and combined to perform better than the others alone. This combination produces a strong classifier. Thus, this method can effectively select different features at each pixel to distinguish foreground objects from the background.

Extensions of this conventional algorithm are available in the form of on-line boosting algorithms [170][172][171] which use several classifier pools, and each pool contains several weak classifiers. Once an input image is given, each classifier pool selects the best classifier for the given image. The selected classifiers form a strong classifier group, and the final classification is performed using those strong classifiers. At the same time, each classifier pool selects the worst classifier as well. The worst classifier is replaced with a randomly selected classifier so that a better classifier can be included in the classifier pool. Instead of selecting the best classifier from each classifier pool as the previous method does, an improvement according to [316] selects several good classifiers from each pool. While the previous method replaces the worst classifier in each pool, instead this improvement replaces several bad classifiers.

According to the literature in this area, feature selection has been less investigated in background modeling and foreground detection methods with only 9 papers. Practically, only five approaches have so far been used in the literature: *(1)* Adaboost [150] used with the classifier-based background model [170][172][171][316], *(2)* Realboost [467] used with the KDE model [417], *(3)* dynamic feature selection [254] with OR-PCA model [255], *(4)* generic feature selection [59] with the ViBe model [32], and *5)* One-class SVM [491][492]. These different approaches and their characteristics are analyzed in Section 18.

2.6.4 Feature Relevance and Learning

To choose the most discriminative features in a multiple features or feature selection scheme, feature relevance may be addressed. More generally, feature relevance can be determined in feature learning scheme which can be classified as developed in Zhong et al. [688]:

1. **Traditional feature learning:** This category includes linear algorithms and their kernel extension, and manifold learning method. Practically, an learning algorithm can be linear or nonlinear, supervised or unsupervised, generative or discriminative, global or local. For example, Principal Component Analysis (PCA) is a linear, unsupervised, generative and global feature learning method, while Linear Discriminant Analysis (LDA) is a linear, supervised, discriminative and global method. Global methods aim to preserve the global information of data

in the learned feature space, but local ones focus on preserving local similarity between data during learning the new representations. For instance, unlike PCA and LDA, Locally Linear Embedding (LLE) is a locality-based feature learning algorithm. Locality-based feature learning like LLE as manifold learning, since it is to discover the manifold structure hidden in the high dimensional data.

2. **Deep learning algorithms:** Deep learning models includes models like Convolutional Neural Network (CNN) [468] and Recurrent neural network (RNN). A survey of deep learning models can be found in Schmidhuber [468].

Feature relevance has been less investigated in background modeling and foreground detection methods than manual image feature methods, such as Local Binary Patterns (LBP) [207], histogram of oriented gradients (HOG) [252], and Scale-Invariant Feature Transform (SIFT) [133]. For *traditional feature learning*, the one work which concerns feature relevance is the work of Molina-Giraldo et al. [378][379]. The feature relevance analysis is made through a Principal Component Analysis (PCA), searching for directions with greater variance to project the data. Thus, the relevance of the original features is quantified with weighting factors. Finally, Molina-Giraldo et al. [378][379] developed a background subtraction method based a multi-kernel learning in which the weight are selected from the feature relevance analysis. Experimental results [378][379] on the I2R dataset [326] show that the proposed Weighted Gaussian Kernel Video Segmentation (WGKVS) model outperforms SOBS [356]. For *deep learning algorithms*, the approaches available in literature can be classified as follows: *1)* Deep Auto-encoder Networks (DAN) [603][604][667], *2)* Convolutional Neural Networks (CNN) [60][582][33][440], *3)* Neural Reponse Mixture (NeREM) [472].

2.6.5 Features and Challenges

In this section, we grouped all the advantages and disadvantages of the different features in terms of robustness against the different challenges met in video and detailed in Bouwmans [51], and they can be summarized as follows:

– **Color features:** Although intensity and color features are often very discriminative features and allow basic foreground detection, they are not robust in challenges such as illumination changes, foreground aperture, camouflage in color and shadows. However, intensity can be used in complementarity of color to deal with different color problems such as dark foreground and light foreground. Furthermore, this combination solves saturation problems and minimum intensity problems [233], and reduces the number of false negatives, false positives and increase true positives. But, the intensity as colors can not work with intense shadows and highlight that often occur in indoor and outdoor scenes, and in presence of gradual or sudden illumination changes [449]. Then, different strategies can be found in literature to alleviate the limitations of the basic color spaces: **(1)** the use

of well-known color spaces which separate the luminance and the chrominance information such as HSV and YCrCb, **(2)** the use of designed shape color space models such as the cylinder color model [212][213][287][646], the hybrid cone-cylinder [131][132], the ellipsoidal color model [519], the box-based color model [549], and the double-trapezium cylinder model [222], **(3)** the use of characteristics in addition of the intensity or color value (mean, variance, minimum, maximum, etc..) (See Section 16), **(4)** the use of designed illumination invariant intensity or color features obtained by normalization [428][566][522][423], **(5)** the use of illumination compensation methods [556][281][280][83][548][372][496][122][498][418], and **(6)** the addition of other features (See Section 17). Normalization based features sacrifice discriminability while texture features cannot operate on textureless regions. Both types of features produce large missing regions in the foreground mask.

– **Edge features:** Edge features are obtained with edge detectors which operate on the difference between neighboring pixels, hence an edge detector should be reasonably insensitive to global shifts in the mean level, i.e. to global illumination changes. Therefore it is interesting to run background/foreground separation algorithms on the output from edge detectors, hopefully reducing the effects of rapid illumination changes. So, the edge could handles the local illumination changes but also the ghost leaved when waking foreground objects begin to move. However, edge features are not sufficiently good to segment the foreground objects isolatedly. Indeed, edge features can sometimes handle dark and light camouflage problems and it is less sensitive to global illumination changes than color feature [234]. Nevertheless, problems like noise, false negative edges due to local illumination prob-lems, foreground aperture and camouflage do not allow an accurate foreground detection. Furthermore, due to the fact that it is sometimes difficult to segment the foreground object borders, it is not possible to fill the objects, and solve the foreground aperture problem. Since it is not possible to handle dark and light camouflage problems only by using edges due to the foreground aperture difficulty, the brightness of color model is used to solve this problem and help to fill the foreground objects.

– **Texture features:** Texture features allow to be robust in presence of shadows and gradual illumination changes, and sometimes in dynamic backgrounds. Texture features can produce false detections due to textures induced by local illumination effects like in cast shadows. Furthermore, an algorithm based only on texture may cause detection errors in regions of blank texture and heterogeneous texture.

– **Motion features:** Motion features can handle irrelevant background motion and clutter such as waving trees and waves.

– **Stereo features:** Stereo features allow the model to deal with the camouflage in color but not in depth.

Thus, multiple features approaches with two, three or a set of features obtained from a bag-of features or by feature selection are suitable to address multiple challenges in the same video (See Section 17). A representative work developed by Li et al. [326] consists in a sets of features built following the type of background (static or dynamic) as follows:

– **Features for static background pixels:** For modeling pixels belonging to a stationary background object, the stable and most significant features are its color and local structure (gradient). As the gradient is less sensitive to illumination changes, the two types of feature vectors are integrated under the Bayes framework in the basic product formulation of the likelihoods.

– **Features for dynamic background pixels:** For modeling dynamic background pixels associated with non stationary objects, color co-occurrences are used as their dynamic features. This is because the color co-occurrence between consecutive frames has been found to be suitable to describe the dynamic features associated with non stationary background objects, such as moving tree branches or a flickering screen.

2.6.6 Features and Strategies

There are several strategies in literature such as multi-scales strategies, multi-levels strategies, multi-resolutions strategies, multi-layers strategies, hierarchical strategies, and coarse-to-fine strategies (See Section 2.1). Practically, different features can be used following the scale, the level or the resolution. For example, a feature can be used at the block level (such as Haar-like features in [675]), and other features can be used at the pixel level (such as RGB in [675]). Thus, these strategies employed multiple features schemes. Please see Table 9, Table 10, and Table 11 for a quick overview.

2.6.7 Features and Similarities

The foreground mask is usually obtained from a similarity/dissimilarity measure between *1)* the direct value of the feature in the background model and the current frame, or *2)* a value computed from the direct value of the feature (mean, variance, probability, etc...) in the background model and the current frame. This value can be a scalar (intensity value, mean, probability, etc..), a vector (2D spatial vector, 3D spatiotemporal vector, etc...) or a histogram (correlogram, etc..). Practically, comparison of features can be made by using similarity/dissimilarity measures obtained with **1)** a crisp, statistical or fuzzy distance for scalar cases, **2)** a ratio for scalar cases, **3)** linear dependence measure for vector cases, and **4)** a intersection measure for histogram (correlogram) case. The choice of the suitable similarity/dissimilarity measure is guided by the properties and the distribution of the concerned features. Furthermore, spatial and temporal features such as LBP and LTP need also measures for their computing as follows: **1)** a measure for the distance in the spatial neighborhood, and **2)** a measure for the distance in the temporal neighborhood. Thus, for

spatial and temporal features like texture, it needs to choose three distances. We list below the different similarity/dissimilarity measures used in the literature for foreground detection (See Table 8 for a quick overview):

A) Similarities for scalar case: Scalar value is the most common case in the literature and the similarities used can be classified as follows:

– **Difference:** The difference computed in a pixel-wise manner between the feature in the background model and the current frame is the most measure used. So, the difference is obtained by a distance and then a threshold is used to classify the pixel as background or foreground as follows:

$$distance(B(x,y) - I(x,y)) < T \tag{4}$$

where $B(x,y)$ and $I(x,y)$ are the values of the feature in the background image and in the current image, respectively. $distance(,)$ is a distance function. Several distance functions have been used in the literature and they can be classified as follows:

1. **Crisp distance:** The most common distance function used for intensity/color values is the absolute distance [5][544]. Aach et al. [6] used a total least squares distance measure. In an other work, Yadav and Sing used a quasi-euclidian distance. To compare Spatiotemporal Condition Information (SCI), Wang et al. [562] designed a specific measure called Neighborhood Weighted Spatiotemporal Condition Information (NWSCI). Using compressive features [84], Yang et al. [618] developed a (Pixel-to-Model) P2M distance.

2. **Statistical distance:** To compare the K distribution in the original MOG, Stauffer and Grimson [512] used the Mahalanobis distance with the RGB features. An alternative to the Mahalanobis distance is the Kullback-Leibler (KL) divergence used in Makantasis et al. [359] with the infrared features and Patwardhan et al. [424] with the RGB features. In a further work, Pavlidis et al. [381] claimed that the MOG algorithm needs a divergence measure between two distributions so that if the divergence measure between the new distribution and one of the existing distributions is "too small", these two distributions could be merged together. Thus, Pavlidis et al. [381] used the Jeffreys divergence measure to check if the incoming pixel value can be ascribed to any of the existing K Gaussians. Experimental results presented by Pavlidis et al. [381] show that the false positives are reduced in comparison with the Mahalanobis distance and the KL divergence. In an other work, Santoyo-Morales and Hasimoto-Beltran used the Chi-2 distance with YUV features instead of the Mahalanobis distance. In a non parametric model based on KDE, Ko et al. [296] choose the Bhattacharyya distance due to its low computational cost. In an other work, Mukherjee et al. [386] developed a distance measure based on support weight to compare RGB features. St-Charles and Bilodeau [505]

employed the Hamming distance to compare LSBPs.

3. **Order-Consistency Measure:** Xie et al. [598] used an explicit model for the camera response function, the camera noise model, and illumination prior. Assuming a monotone and nonlinear camera response function, Xie et al. [598] show that the sign of the difference between two pixel measurements is maintained across global illumination changes. Noise statistics are used to transform each frame into a confidence frame where each pixel is replaced by a probability that it is likely to keep its sign with respect to the most different pixel in its neighborhood. Hence, an order consistency measure is defined as a distance between two distributions. Xie et al. [598] used the Bhattacharyya distance due to its properties to the Bayes error. Finally, an Illumination Invariant Change Detector via order consistency (IICD-OC) is developed. Experimental results [598] on videos taken by an omni-directional camera show the robustness of IICD-OC against illumination changes. But, the ordinal measure required a reordering of blocks and it is computationally expensive. To solve this problem, Singh et al. [495] explicitly modeled noise under which rank-consistency is tested, and used a probabilistic generative model under which frame blocks are generated. The order-consistency is posed as a hypothesis validation problem using fast significance testing based on PAV. In a further work, Parameswaran et al. [418] used the same order-consistency measure in an illumination compensation approach.

– **Ratio:** The ratio computed in a pixel-wise manner between the feature in the background model and the current frame is also used in several works. Thus, the ratio is obtained by a division and then a threshold is used to classify the pixel as background or foreground:

$$ratio(B(x,y), I(x,y)) < T \qquad (5)$$

where $B(x,y)$ and $I(x,y)$ are the values of the feature in the background image and in the current image, respectively. For example, Baf et al. [21][24][22] computed the ratio of color components and LBP that are further aggregated with the Choquet integral. In a further work, Baf et al. [23] used the ratio of the IR intensity and LBP. In a similar approach, Ding et al. [130][127] computed the ratio between the difference and the number of gray level for the color components and LBP. In a further work, Ding et al. [128] developed a specific ratio measure to compare gradients. This measure is a ratio between a product and a sum of the gradient in the background and the current images. Azab et al. [17] used a similar scheme than Ding et al. [130][127] but with statistical values. In a change detection approach, Aach et al. [4] and Aach an Kaup [3] used a ratio of probabilities.

B) Similarities for vector case: Linear dependence measures and colinearity measures are the most used similarity measure to compare vector. First, Durucan and Ebrahimi [136] proposed to use a linear dependence measure (colinearity measure) to test the depence/independence properties of each vector that represented information on the neighborhood region of the concerned pixel. The idea is that illumination

changes do no change the colinearity of the vector. Thus, a Linear Dependence Detector (LLD) which is invariant to transformations was defined for change detection, and applied to the vector composed by the reflectance for change detection. Experimental results [136] show that LDD outperforms the Statistical Change Detector (SCD) developed by Aach et al. [5] in presence of noise as well as global illumination changes and local shadows and reflection. LLD presents the advantage to detect semantic objects but object interiors maybe not well detected and it requires high computational complexity. To solve these problems, Durucan and Ebrahimi [137] used a more rigorous test based on the Wronskian determinant. The corresponding detector called Wronskian Detector (WD) outperforms both SCD and LDD. In a further work, Durucan and Ebrahimi [138][139] used an other test based on the Gramian determinant which provides low computational cost and is easy to implement. Thus, Durucan and Ebrahimi [138][139] proposed a Gramian Detector (GD) which can be applied on color images in the RGB color space. In practice, other than illumination changes, change detection is also influenced by the noises of cameras and reflections but the previous linear algebra detectors have intrinsic weakness in case of noises [155]. To solve this problem, Gao et al. [155] developed a Linear Approximation Detector (LAD). Experimental results on the PETS 2001 dataset show that LAD show more robustness against noise than LDD and WD. In an other work, Ming et al. [370][371] proposed a local linear dependence based Cauchy Statistical Model (LLD-CSM). Experimental results [370][371] demonstrate that LLD-CSM can tolerate the local or global slow or sudden illumination changes, noise due to small motion in the background.

C) Similarities for histogram case: Intersection measures are the most used similarity measure to compare histograms or correlograms. First, Mason and Duric [363] used an intersection measure to compare Local Color Histograms (LCH). In an other work, Fabian [145] first used simple metric to compare Histograms of Oriented Gradients (HOG) in a discrete metric space. But, this metric made no difference between two different bins. Due solve to this problem, Fabian [145] proposed a complex metric. Several works [90][406][323] [660][661][687][218] used the Bhattacharyya distance as can be seen in Table 8. The chi-squared measure is used to compare local gradient histograms in Mason and Duric[363]. The Bhattacharyya distance is the most statistical used distance and was employed for the following histograms: 3D HRI [90], Local Kernel Color Histograms [406], ELKH [323], LDH [660][661], LH-FGs [687] and adaptive HOG [218]. In an other approach, Mukherjee et al. [386] developed a HoG distance.

In the following sections, we present and analyze the different features currently used in background modeling and foreground detection in terms of robustness against the challenges met in videos taken by a fixed cameras.

Categorie	Sub-categories	Features
Scalar Case		
Difference	**1) Crisp Measure**	
	Absolute Distance	Reflectance [544], Intensity, Color
	Total Least Squares Distance	Reflectance [6]
	Logarithmic Distance	Reflectance [594]
	Quasi-Euclidian Distance	RGB [610]
	Cosine Measure	Gradient [654], Texture [134]
	Exponential Cosine Measure	Gradient [654]
	Angular Deviation	Intensity Vector [269]
	P2M Distance	Compressive Feature [84][618]
	Neighborhood Weighted Spatiotemporal Condition Information (NWSCI)	SCI [562]
	2) Statistical Measure	
	Mahalanobis distance	RGB [512]
	Bhattacharyya distance	RGB [296]
	Chi-2 distance	YUV [456]
	Kullback-Leibler divergence	IR [359], RGB [424]
	Jeffreys divergence	RGB [381]
	Distance Measure based on Support Weight	RGB [386]
	Hamming distance	LSBP [505]
	l_1 distance	RGB [563]
	l_2 distance	RGB [563]
	3) Fuzzy Measure	
	Interval-valued fuzzy similarity	FST Features [93]
Ratio	**1) Direct values**	IR Intensity [23]
		RGB [24], YCrCb [24]
		LBP [24], ULBP [351]
		Intensity Vector Magnitude [269]
	2) Differences	*Crisp values:*
		RGB [130], YCrCb [130], HSI [130]
		LBP [130]
		Statistical values:
		RGB [17], LBP [17]
	3) Other values	*Product/Difference:*
		Gradient [128], LBP [130]
		LBP [130]
		Probabilities: [3]
		[3]
Vector Case		
Linear Dependence Measure	**1) Linear Dependence Measure**	
	LDD [136]	Reflectance (Vector)
	Wronskian Detector (WD) [137]	Reflectance (Vector)
	Gramian Detector (GD) [138][139]	Color (Vector)
	3) Linear Approximation Measure	
	Linear Approximation Detector (LAD) [155]	Reflectance (Vector)
	4) Local Linear Dependence Measure	
	Local Linear Dependence Measure (LLD-CSM) [370][371]	YCbCr (Vector)
	5) Spectral Distance Measure	
	Spectral Distance Measure [35]	Spectral Feature (Vector)
Order Consistency Measure	**Order-Consistency Measure**	
	Bhattacharyya distance [598]	Reflectance
	PAV [495]	Color
	qPAV [495]	Color
Histogram Case		
	1) Intersection Measure	LCH [363]
	Simple metric (discrete metric space)	HOG [145]
	Complex metric (discrete metric space)	HOG [145]
	l_1 distance metric	HSV [565]
	Earth movers distance metric	HSV [565]
	Normalized Histogram Intersection	LFCH [290]
	2) Proximity Measure	LBP [17]
	3) Bhattacharyya distance	3D HRI [90], Local Kernel Color Histograms [406]
		ELKH [323], LDH [660][661], LH-FGs [687]
		adaptive HOG [218]
	4) Chi-Squared Measure	LGH [363]
	5) HoG distance	HOG [386]

Table 8 Similarities Measures: An Overview. The first column indicates the type of the measure for each category of feature in terms of value (scalar, vector and histogram). The second column gives the name of the measures and the third column indicates the corresponding features on which the measure is applied.

3 Intensity Features

Intensity features are the most basic features that can be provide by gray-level cameras [493][242][267][47] or infra-red (IR) cameras [116][117][115][346][359][402][611][478][34][107][360][626][614][448]. Practically, several works combined both of them by using a fusion scheme strategy [118][119][121][120][392][393][39][100][98][99][196][320] [319] to combine their

respective advantages. In other works, several designed illumination invariant features based on intensity have been developed as follows:

- **Reflectance:** Several reflectance models have been developed to describe the reflectance due to normal, forescatter and backscatter distributions such as the Lambertian model, the Phong shading model, the dichromatic reflection model [295] and the Ward model [583]. Toth et al. [544] were the first authors who used the reflectance in change detection and they proposed to obtain the illumination and the reflectance components with a homomorphic filtering [413] based on the *Lambertian model*. Then, only the reflectance components are compared by using a sum of absolute differences within a sliding window for foreground detection between consecutive frames. Toth and Aach [542] used this approach to detect and further distinguish between humans, vehicles, and background clutters. In further works, Toth et al. [543][2] improved this method by using a Bayesian framework. In an other way, Aach et al. [6] used a total least squares distance measure instead of sum of absolute differences. Aach and Condurache [1] used a threshold transform by significance invariance.

- **Surface Spectral Reflectance:** Sedky et al. [469] developed a method called Spectral-360 which adopted a physics-based model called the *dichromatic reflection model* [295]. This approach is different from the previous work, in that it relies on models, which can represent wide classes of surface materials. Then, the feature used is the Surface Spectral Reflectance (SSR). Thus, the background model based on the mean SSR as well as its maximum correlation and its minimum correlation, is built during a training step and updated continuously. Experimental results [469] on the ChangeDetection.net 2014 dataset show that Spectral-360 outperforms MOG [512] and KDE [142] in terms of F1-measure. In IR camera, Nadimi and Bhanu [392] obtained the reflectance throught the *dichromatic reflection model* too, and they combined visible and IR intensity feature via a physics-based fusion. In further works, Nadimi and Bhanu [393][39] used an evolutionary-based fusion method.

- **Radiance:** Based on the *Phong shading model*, Xie et al. [598] used the radiance obtained by considering both ambient and diffuse reflection, and thus take into account an explicit model for the camera response function, the camera noise model, and illumination prior. Then, Xie et al. [598] developed an illumination-invariant foreground detection via order consistency.

- **Photometric Variations:** Di Stefano et al. [513] adopted a visual correspondence measure called Matching Function (MF) [513]. This measure matches corresponding blocks of two images by checking an ordering constraint. Since photometric variations tend to respect the ordering of intensities in a neighborhood of pixels, MF allows to handle sudden and strong illumination variations between the background scene and the current frame. Practically, MF matched the high contrast regions corresponding to the intensity edges of two blocks, since only

high intensity differences can provide high contributions to the correlation score. In a second step, Di Stefano et al. [513] used a tonal alignment technique. Hence, the background is tonally aligned to the current frame by applying a histogram specification method [166], and then Di Stefano et al. [513] obtained a background image where the photometric distortions have been removed. Finally, a pixel-wise difference between the background image and the current frame extracts the foreground regions.

4 Color Features

Color features provide *spectral information*, and are the natural features as they are directly available from the sensor or the camera. Although they are often used for facilitating easy discrimination between the background and the foreground, color features are generally not robust against illumination conditions and shadows cast by the moving objects. Furthermore, similar colors between background and foreground lead to the well-known problem of camouflage in color.

4.1 Features in Well-Known Color Spaces

Several color features in different color space have been proposed in the literature and are described as follows:

- **RGB color space:** The RGB color space is the most popularly used feature due to their direct availability from the sensor or the camera. Red, Green, and Blue channels of each pixel are usually measured with 8-bits resolution, where 0 is no color (black) and 255 is the maximum color (white), therefore, a total of 24-bit true color definition. But the RGB color space has several limitations: *1)* it is well known that the RGB color space does not reflect the true similarities among colors, *2)* depending on the scene one color component could be more informative than the others, so it should be given more importance than others, *3)* the three components are dependent on each other which increase its sensitivity to illumination changes. For example, global illumination changes shift the mean level of the entire RGB image, possibly with shifts of different magnitude for each color component, and *4)* as the three channel components are correlated, there is a need to compute inter-correlation terms in the covariance matrix which shall be incorporated into existing background models such as in the MOG model [512]. Stauffer and Grimson [512] demonstrated that by not computing these inter-correlations terms, computational speed improves with increased false detections.

- **Normalized RGB (rgb) color space:** The normalized RGB space is derived from the traditional RGB color space to be illumination invariant. Xu and Ellis [601] used the normalized RGB to allow the MOG to be robust to fast illumination

changes in an outdoor environment lit by sunlight and shadowed by clouds.

- **YUV color space:** The YUV space separates luminance and chroma and so it is more suitable for improving the robustness of the model against illumination changes. For example, Wren et al. [589] used the normalized components, U/Y and V/Y to remove shadows in a relatively static indoor scene. Using the MOG model, Harville et al. [202] defined a chroma validity test based on the luminance Y as the chroma (U and V) components become unstable when the luminance is low. When the test is not verified, the chroma components of the current observation are not used and so are its current Gaussian distributions. Furthermore, the detection in luminance was combined with the detection in depth improve robustness to color camouflages.

- **HSV color space:** The HSV color space is used to improve the discrimination between shadow and object, classifying shadows as those pixels having approximately the same hue and saturation values compared to the background, but lower luminosity. For example, Sun et al. [520] used the Hue-Saturation-Value (HSV) color space, because the likelihood term in the MOG model shows stronger contrast in HSV space rather than the RGB space, especially for objects that share similar appearance to the background (Camouflage in color). In an other study, Kanprachar and Tangkawanit [273] compared the RGB and HSV color spaces. Experimentally, HSV color space was very suitable under low illumination intensity conditions. Kanprachar and Tangkawanit [273] showed that not all three parameters in HSV color system are useful for the detection. Saturation (S) and Value (V) are the two key parameters to be used.

- **HSI color space:** HSI color space is closer to human interpretation of colors in the sense that brightness, or intensity, is separated from the base color. HSI uses polar coordinates. In the original MOG model, shadows are extracted as part of object mask when using the RGB color space. To address this problem, Wang and Wu [576] used the HSI color space which tends to be shadow-removable. However, the obtained results are not satisfactory due to the fragmented segmentation obtained by using hue and saturation. In order to achieve both "shadow-rejection" and "segmentation stability over time", Wang and Wu [576] employed the MOG on chroma (hue and saturation) and luma (intensity) separately. The fused results obtained by combining chroma and luma is prepared using two criteria. This scheme reserves the advantage of using chroma (i.e. avoiding shadow) and that of luma (i.e. stability of segmentation). The foreground mask is refined using an Hidden Markov Model (HMM) for improved performance.

- **Luv color space:** Yang and Hsu [622] used the Luv components assuming independence in the computation of covariance matrix required in the MOG model. Then, Yang and Hsu [622] built an hybrid feature space with spatial and color features to obtain a 6-dimensional hybrid feature vector for each pixel. A mean-shift procedure classified each hybrid feature vector to its corresponding local maximum along the gradient direction. Thus, a set of neighbouring pixels associated

with the same local maximum (i.e. mode) is highly similar in this hybrid feature space. Yang and Hsu [622] then assign pixel-wise background likelihood for each pixel using the MOG likelihood, and further obtain a smoothed version of MOG in terms of spatial and color coherency.

– **Improved HLS color space:** Setiawan et al. [471] proposed to use the IHLS color space which has the following advantage against the RGB color space. That is to identify shadow region from object by utilizing luminance and saturation-weighted hue information directly, without any calculation of chrominance and luminance. By exploiting this color space in the MOG model, Setiawan et al. [471] obtained good sensitivity to color changes and shadow.

– **Ohta color space:** The axes of the Ohta space are the three largest eigenvectors of the RGB space, found from the principal components analysis of a large selection of natural images. This color space is a linear transformation of RGB. Using the mean model, Zhang and Xu [654][655] used the Ohta color space. The three orthogonal color features of Ohta color space are important components for representing color information. Good results in the case of illumination changes and shadows in outdoor scenes are achieved by using only the first two components which are combined with the texture feature (LBP).

– **YCrCb color space:** YCbCr uses Cartesian coordinates. El Baf et al. [21][24] used the YCrCb color space combined with the texture feature (LBP) to be robust to illumination changes and shadows. Experimental results in [21][24] showed that YCrCb color space is more robust in these cases than the Ohta and HSV color spaces.

– **Lab/Lab2000HL color space:** Lab color space is a color space which indicates proper changes in the direction of human color perception. Its components are the lightness of the color and two color opponent dimensions. Lab2000HL color space, which is an improved version of Lab color space, was introduced and is thought to perform a better modeling of human perception. Particularly, Lab2000HL color space has linear hue band. So, Balcilar et al. [27] investigated the performance of the Lab2000HL color space. The average precision value of Lab2000HL is the greatest in all videos in comparison to all other color spaces. In terms of the computational costs for each color space (YCrCb, Luv, Lab, Lab2000HL), RGB color space leads the lowest. The reason is that it does not require any transformation since the information gathered from the camera sensors is directly in RGB. Lab2000HL color space, on the other hand, has the most computational cost, since a computationally intensive procedure is required to apply first the Lab transformation, and then the computation of transformation value with respect to the transition map using interpolation. The conclusion is that the Lab2000HL color space increases foreground detection rate significantly, in spite of its high computational cost. Balcilar et al. [26] improved the performance obtained by the Lab2000HL color space with a spatial and temporal smoothing

scheme.

– **CIEDE2000 color space:** Thangarajah et al. [537] used the CIEDE2000 color-difference described in Sharma et al. [477].

Several comparisons on the color spaces can be found in the literature [298][450] [273][28][350][475][108][301]. First, Kumar et al. [301] studied five color spaces using the single Gaussian model: RGB, XYZ, YCrCb, HSV and the normalized rgb. YCrCb obtained the highest true detection rate and thelowest miss rate on traffic video sequences.

The most complete comparison made by Kristensen et al. [298] compared the following color spaces using the MOG model: RGB, HSI, YCbCr, rgb, $C_1C_2C_3$ and $l_1l_2l_3$. rgb, $C_1C_2C_3$ and $l_1l_2l_3$ are all invariant to changes in brightness, a color space that should decrease the sensitivity to shadows. This investigation showed that the HSI, rgb, $C_1C_2C_3$, $l_1l_2l_3$, and $m_1m_2m_3$ are noisy but less sensitive to shadows than RGB and YCbCr. YCbCr is less noisy than RGB, due to its more independent color channels. Even though both HSI and $l_1l_2l_3$ are sensitive to changes close to the gray scale, the result is much worse for $l_1l_2l_3$. The light invariant color spaces do not detect as much shadows as the other color spaces, at a cost of missed detection of bright areas. With the $m_1m_2m_3$ color space the segmentation algorithm becomes more of an edge detector, since it is based on two neighboring pixels. Overall, the most suitable color space for the segmentation algorithm is YCbCr. It is least sensitive to noise, due to numerical stability and more independent color channels. No information is lost when it is transformed from RGB in comparison to other normalized color spaces in which brightness information is usually lost. Finally, YCbCr is affected by shadows and compared to RGB, is too insensitive in some cases but Kristensen et al. [298] presented compensation methods for these two cases.

In the case of self-organizing neural network model [356], Lopez et al. [350] compared five different color spaces : RGB, Lab, Luv, HSV and YCrCb. Furthermore, they proposed a color component weighting selection process to take into account the different importance of each component in a color space. A set of 22 different configurations was then evaluated on the I2R dataset [326]. The performance of color spaces has been different in each sequence but YCrCb is the best choice in most cases, and it remains as an interesting option. YCrCb gives the best results when setting the weight to 1 or close to this value for Y, which indicates that the most important channel for a robust detection is the one that corresponds to lightness, while the remaining channels often add nothing to the result.

For moving shadow elimination, Shan et al. [142] determined the optimal color space in which to remove shadow among the following set of color spaces: RGB, HSV, YCrCb, XYZ, L*a*b*, c1c2c3, $l_1l_2l_3$, and normalized RGB. Practically, there are three key points with moving shadow detection and elimination [142]: **(1)**It exists different classes of shadow due to the variety of scenes and different properties of shadow, **(2)** shadows in each color spaces are different, which provide different fore-

ground detection results, and **(3)** there might be not one but different optimal color space following the prominent properties of the shadow. Shan et al. [142] focused in detecting two kinds of shadows: shadows illumination and shadows reflection. Then, shadows are reclassified as invisible and visible. Experimental results [142] show that every space cannot be suitable for all kinds of shadows. HSV, c1c2c3, normalized RGB spaces are appropriate for visible shadow, and YCbCr, L*a*b* spaces are appropriate for invisible shadow. Even if all types of shadows can be removed in one color space with special methods, different application can selected suitable color space, which can provide twice the result with half effort.

4.2 Features in Designed Shape Color Spaces

In literature of background modeling and foreground detection, several color space models were designed with a particular shape for the test volume around a background pixel, whose space defines the cluster associated with that pixel. In other words, the test pixels which is inside this volume are associated with the corresponding background. The shape of the volume determines its robustness to highlight and shadows and the different shapes can be classified as follows:

– **Cylinder Color Model:** Horprasert et al. [212][213] separated the brightness and the chromaticity components to deal with shading and shadows. Thus a brightness distortion and a color distortion are computed from the RGB components. This color model is called the cylinder color model,and allows to discriminate a RGB color pixel into shadow, highlight, background, and foreground under static background. In the case of the codebook model, Kim et al. [287] used also a cylinder color model which separated the color and brightness component to cope with the problem of illumination changes such as shading and highlights. Thus, a color distortion feature is computed from the current RGB components and the RGB components stored in the codebook. For brightness changes, Kim et al [287] computed a brightness distortion computed from statistics on the intensity to allow the brightness to vary in a certain range that limits the shadow level and highlight level. In an other work, Guo and Hsu [184][185] improved this cylinder color model by using an adaptive threshold instead of a fixed threshold to increase its ability to distinguish highlight and shadow.

– **Arbitrary Cylinder Color Model:** However, the CY model is valid only if the spectrum components of the light source change in the same proportion. In fact, it is not true in many practical scenes, where the variations of each spectrum component of the light source would not be in proportion. In these cases, the CY model is inaccurate and much less efficient. To solve this problem, Zeng and Jia [646][647] used an Arbitrary CYlinder based color (ACY) model. The ACY model uses cylinders whose axes needs not going through the origin, so that the CY model is extended to more general cases. Furthermore, the ACY model reduces the false classification rate of CY model by more than 50% without loss of real-time performance. Practically, the CY models only a special case of the

ACY model.

– **Hybrid Cone-Cylinder Color Model:** In the cylinder color model, almost every low-intensity test pixel are assigned to the cylinder of a background pixel of very small intensity. Thus, as two similarly grouped pixels increase intensity, they have less chance of being in the same cluster, even though their respective chrominance remains the same. As a more suitable shape, a cone corrects these problems and more precisely covers the color space but a pure conical highlight detector attracts too many pixel values within its space and thus increases the false negative rate. For sensitive detections, the highlight volume should be limited to a cylinder. Thus, Doshi and Trivedi [131][132] presented a Hybrid Cone-Cylinder Codebook (HC3) model for a 24-7 long-term surveillance system. Experimental results [131] on videos taken by a fixed camera and an omni-directional camera show that the hybrid cone-cylinder color model in HSV outperforms the cylinder color model in RGB.

– **Ellipsoidal Color Model:** In the CY model, the two opposite margins of the cylinder cell are not sufficient to model the background pixel with changed illumination. Indeed, it is not optimal in case of illumination changes to use cylindrical cells of the same size with the same threshold. Furthermore, false negative detections may occur due to the dark background pixels of which the cylinder codewords are located near the origin. The closer to the origin the cylinder codeword is, the larger possibility that the corresponding dark background pixels may be easily miss-classified as a foreground pixel because of the small illumination changes. To handle this problem, Sun et al. [519] described the color distribution of each background pixel by ellipsoidal shape codewords based on multiple 3D Gaussian distributions. The variation of every stationary background pixel is limited so that they may be described by only one ellipsoidal cell. Experimental results [519] show that the ellipsoidal color model outperforms the cylinder color model.

– **Box-based color model:** To reduce computation time, Tu et al. [549] used a box-based color model which using box-based subspace to represent a codeword. The three edge-lengths values of the box are usually set to the same in the RGB color space, so the box is a cubical box. This model reduces computation complexity. Letting a 24-bit color image sequences, its color mode consisting of RGB uses 8 bits to represent red, blue and green separately. Thus, the color space can be considered as a large cubical box 256×256. Then the input pixel value is encoded into the center of cluster when its value lies within a cubical box. So, the pixel value is represented by the cluster center.

– **Double-Trapezium Cylinder Color Model:** Huang et al. [222] proposed a Double-Trapezium Cylinder Color Model on YUV, called DTCC-YUV. First, Huang et al. [222] located the shadow detection area in the lower part of the model based on the low luminance of shadows in comparison with the background. Then, a trapezium cylinder is built as the structure of the shadow detection area by us-

ing the variable chrominance of shadows. Second, Huang et al. [222] located the highlight detection area in the upper part of the model based on the high luminance of highlights in comparison with the background. An inverted trapezium cylinder is built as the structure of the highlight detection area based on the variable chrominance of highlights. Finally, the main background area is built with a cylinder structure in the middle part of the model and then DTCC-YUV model is completel built. Experimental results [222] on the PETS 2001 dataset show that the DTCC-YUV model outperforms the cylinder color model [212] and the hybrid cone-cylinder codebook model [131].

4.3 Designed Illumination Invariant Color Features

Design of illumination invariant color features was investigated to be robust with illumination changes because the well-known color spaces show limitations in this case. So, Pilet et al. [428] adopted a color illumination ratio as feature to deal with sudden illumination changes. This illumination ratio does not depend of the surface albedo. It depends of the surface orientation and on the illumination environment. Therefore, Pilet et al. [428] used the MOG model with this color illumination ratio instead of RGB with a spatial feature which is the Normalized-Cross Correlation (NCC). Experimental results [428] show that MOG with the color illumination ratio is more robust to illumination changes than the improved MOG of Zivkovic [695] with RGB. In a further work, Wang and Yagi [566] used the same ratio but with a more efficient learning model which improved the robustness against illumination changes. In a multi-backgrounds approach, Takahara et al. [522] used both RGB and HSV features be robust againts various illumination changes while Sajid and Cheung [454] used both RGB and YCrCb features. In an other work, Yeh et al. [627] developed a color illumination-invariant.

4.4 Color Filter Array (CFA) Features

A color image is composed of three channels per pixel which are usually acquired by using three spatially aligned sensors. But, this method present two main disadvantages: 1) it increases the camera size, and cost, and 2) it requires a complicated pixel registration procedure. So, most digital color cameras employ a single image sensor with a Color Filter Array (CFA) in front. Each pixel measures only one color and spatially neighboring pixels which correspond to different colors are used to estimate unmeasured colors. Among all the CFA patterns, the Bayer CFA pattern is the most widely used pattern [518]. Practically, the Bayer CFA pattern is a 2×2 pattern which has two green components in diagonal locations, and red and blue components in the other locations. An image based on this pattern is called a Bayer-pattern image. The interpolation process which allow to obtain a full-color image is called "demosaicing". Suhr et al. [518] used the Bayer pattern instead of RGB in the original MOG

[512]. This method presents almost the same accuracy as the original MOG using RGB features while maintaining computational resources in terms of time and memory similar to the original MOG using grayscale feature.

4.5 Statistics on Color Features

Statistics on color feature can be defined and used as features such as:

– **Standard Variance:** Zhong et al. [685] divided each image into patches by representing each image patch as a standard variance feature computed on the intensity. Then, assuming that standard variance feature fits a MOG distribution, Zhong et al. [685] used the MOG model [512]. The advantages of using the standard variance feature as co-occurrence statistics features for dynamic background modeling are the following: *1)* It explicitly considers correlation between pixels in the spatial vicinity. Indeed, an image patch's center pixel in current frame would be a neighbouring pixel in the next frame due to the small movements of objects in dynamic scenes. The center pixel's intensity will change non-periodically. However, the image patch's standard variance feature is unchanged due to the spatial co-occurrence correlations between the center pixel and its neighboring pixels. *2)* Image noise is largely filtered out with the average filter during the computation of standard variance feature; *3)* The standard variance feature is invariant to mean changes such as identical shifting of intensities. This is very suitable in the case of illumination changes; *4)* The standard variance feature results in a low dimensional scalar representation of each image patch. This avoids expensive computation during the background modeling phase. Experimental results [685] on several dynamic backgrounds show that the MOG model is more robust with this feature than RGB features.

– **Color co-occurrences:** Li et al. [326] used in feature selection scheme the color co-occurrences which are more significant features for dynamic background pixels (See Section 2.6.5). In an other work, Adam et al. [7] used co-occurrences of intensities in the spatio-temporal neighborhood of a pixel for dynamic background modeling in an ocean scenes. There are three approaches to used intensities and color co-occurrences [7]: *1)* The intensities are collected into a vector. Then similarity between two spatio- temporal neighborhoods is measured by the norm of the difference vector or by correlation between the two vectors. This approach is too strict for dynamic environments because we cannot expect a good correlation between two space-time volumes of a waving tree or an ocean for example, *2)* Based only on the histogram of the intensities, the two spatio-temporal neighborhoods are compared with their histograms. Instead of the first approach, the order of the samples is completely irrelevant. Furthermore, it ignores important spatio-temporal relationships between pixels, and *3)* a compromise between the two previous approaches which robusly combines each corresponding fragments similarity scores to obtain an overall similarity measure between the spatio-

temporal neighborhoods.

- **Entropy:** Entropy is a measure of uncertainty. The pixel value changes over time from frame to frame due to the following reasons: (1) changes in the background as in the case of illumination changes and dynamic backgrounds, and (2) Motion objects which make the pixel value change from background to object or from object to background. Thus, in Ma and Zhang [355], the change of pixel value is considered as the state transition of pixel, e.g. in a 256 level gray image, each pixel has 256 states. Pixel states change brought by noises would be in a small range, but those brought by motion will be large. So the diversity of state at each pixel can be used to characterize the intensity of motion at its position. With this assumption, the probability distribution of each pixels state is observed along temporal axis. Ma and Zhang [355] used a temporal histogram to present state distribution in a sliding window. The probability density function of pixels state is obtained by histogram normalization. Once the histogram is obtained, the corresponding probability density function for each pixel is computed as follows:

$$P_{i,j,q} = \frac{H_{i,j,q}}{N} \tag{6}$$

where $H_{i,j,q}$ denotes the spatial-temporal histogram i.e. $w \times w \times L$ pixels are accumulated to form the histogram of pixel (i,j), q denotes the bins of the histogram, the total number of bins is Q, N is the total number of pixels in the histogram and $\sum_{q=1}^{Q} P_{i,j,} = 1$. Once the pdf of the pixel is known, the state diversity level of this pixel is computed using entropy definition as follows:

$$E_{i,j} = -\sum_{q=1}^{Q} P_{i,j,q} log(P_{i,j,q}) \tag{7}$$

where $E_{i,j}$ is called the spatial-temporal entropy of pixel (i,j). $E_{i,j}$ is quantized into 256 gray levels to form an energy image, named as Spatial Temporal Entropy Image (STEI). In STEI, the lighter the pixel is, the higher its energy is, and the more intensive its motion is. But, STEI cannot differentiate the high entropy caused by motion and spatial structure of the image. Thus, Jing et al. [262] proposed a method based on difference image. The entropy images formed this way is denoted as Difference-based Spatial Temporal Entropy Image (DSTEI). Experimental results on the PETS 2001 dataset show that DSTEI shows more robustness than STEI in presence of gradual illumination changes and dynamic backgrounds but false detections occur in the case of shadows or sudden illumination changes. In an other way, Chang and Cheng [76] used the entropy image with an adaptive state-labeling technique. Similar to the STEI and DSTEI methods, Chang and Cheng [76] construct a spatio-temporal sliding window for each pixel (i, j) to acquire the corresponding spatio-temporal histogram but the histogram is based on the distribution of pixels state labels instead on the distribution of pixels intensity. A state label by a simple three-frame differencing rule is assigned to each pixel, and the state label might be changed adaptively when the algorithm proceeds over time. This algorithm leads to lower computational complexity compared to

to STEI and DSTEI. The detection has a more precise outline and the algorithm is more robust and less sensitive to the changes of parameters. In an other approach, Park et al. [422][421] used an entropy-based adaptive Gaussian mixture model based on the idea that the increment of entropy generally means the increment of complexity. Because objects in unstable conditions cause higher entropy variations, pixels with large changes in entropy in moments can be considered as moving objects. Therefore, Park et al. [422][421] applied the Clausius entropy theory to convert the pixel value in an image domain into the amount of energy change in an entropy domain. Second, Park et al. [422][421] applied a MOG algorithm on these entropy features. Experimental results [422][421] show that the entropy-based method outperforms the original MOG in presence of dynamic backgrounds and illumination changes.

– **Illumination-Invariant Structural Complexity (IISC):** Based on the idea that the orthogonal decomposition is effective for separately handling structural features from the illumination effects in a small local region, Kim and Kim [291] used the singular value decomposition (SVD). Thus, SVD coefficients normalized by the largest singular value provide the illumination-invariant feature space. Due to the fact that pixel values in the small local region are mainly determined by the illumination, its variation is well revealed in the largest singular value in the SVD-based scheme. Then, Kim and Kim [291] defined the unit brightness level in which all the singular values are divided by the largest one. The illumination-invariant structural information are revealed by such normalized singular values. Therefore, the sum of those values is employed as feature called Illumination-Invariant Structural Complexity (IISC). The background model used the traditional on-line updating scheme, i.e the running average. Experimental results [291] on the PETS2001 dataset and OTCBVS dataset show that IISC has the ability to discriminate structural changes due to moving objects from those due to illumination effects.

4.6 Discussion on Intensity and Color Features

According to the literature, it appears that **1)** Spectral-360 [469] based on the dichromatic reflection model is the best reflectance approach for intensity features, **2)** YCrCb color space seems to be the most suitable space for category of color features in well-known color spaces as demonstrated in case of the MOG model and the self-organizing neural network model [350] even if it is affected by shadows and it is too insensitive in some cases. But, compensation methods [298] can be employed for these two cases, **3)** for features in designed shape color spaces, the double-trapezium cylinder color model [222] in YUV color space appears as the most suitable shape model for the codebook model, **4)** color illumination ratio [428] provides a robust illumination-invariant color features, and **5)** When a single image sensor with a Color Filter Array (CFA) is used , Bayer CFA pattern [518] allows the MOG model to keep the same accuracy as the RGB features while maintaining computational resources

in terms of time and memory similar to the gray scale feature. Furthermore, statistics on color features are useful to be more robust in presence of dynamic backgrounds and illumination changes.

All the intensity and color features can be combined with edge, texture, motion or stereo features to take advantage of these features. In literature, edge features are one the first features used in addition of color features and are detailed in the following section.

5 Edge Features

5.1 Intensity Edge Features

Edge based on intensity features give ***spatial information***, and are computed using a gradient approach such as Canny, Sobel [240][210][17][349][326][234] or Prewitt [339][579] edge detector. The gradients can be calculated from the gray level image or in each components of a color space. The edge features are generally used in addition to intensity or color features [240][249][339][452] or alone [243] [390][444][283][391][282][443][284][211] as follows:

– **In addition with intensity or color features:** First, Jabri et al. [240] used in addition to the intensity features the intensity gradient obtained by the Sobel edge detector. Large changes in either intensity or in edges are fused. However, the involvement of the intensity model retains the sensitivity to sudden changes in illumination. In Javed et al. [249], the edge and color information obtained from pixel level is integrated at the region level. The basic idea is that any foreground region that corresponds to an actual object will have high values of gradient based background difference at its boundaries. It requires significant changes in both the intensity and intensity gradient. The use of a gradient feature removes many false alarms due to small illumination changes. However, intensity gradients arising from large illumination changes can still generate false detections. As edge detection consists of first filtering the image using a suitable approximation of derivatives followed by a thresholding, it gives an image where the pixel values come from a binary distribution. This image is difficult to model using Gaussian mixtures. Skipping the thresholding gives pixel values that come from a continuous distribution. However, since a majority of the picture usually contains no edges such a distribution will be extremely skewed, and thus still difficult to model using Gaussian mixtures. To solve this problem, Lindström et al. [339] proposed to use a Prewitt edge detector without the thresholding independently to each color component followed by a log-transformation, to reduce skewness, gives a color edge image with pixel values that can be modeled using Gaussian mixtures. Experimental results [339] shows better performance against illumination changes for the log-transformed detection using the Prewitt edge detector. In an other work, Kim et al. [285] used edge and non-edge (texture) features in a

hybrid background model to generate the background model. Thus, theses features are encoded into a coding scheme called Local Hybrid Pattern (LHP). LHP selectively models edges and non-edges features of each pixel. Then, each pixel is modeled with an adaptive code dictionary to take into account the background dynamism. In the background maintenance, stable codes are added in the model while unstable ones are discarded. The incoming codes that deviate from the dictionary are classified as edge or inner region. Experimental results [285] on the ChangeDetection.net dataset show that this Adaptive Dictionary Model (ADM) with LHP features outperforms the original MOG [512], the original LBP [206] and SALBP [404]. ADM achieves similar results than ViBe [32] and PBAS [209]. In other works, Li et al. [330] and Gu et al. [174] used the gradient feature with the KDE model [142]. The reader can see how the color features and edge features are fused in these different approaches in Section 17.

– **Edges alone:** First, Kim and Hwang [279] proposed to use only edges to model the background, and thus this approach used binary information as existence of edge on each pixel. But, regions in consecutive frames may not have exactly the same edge position, and have shape and length changes due to presence of noise. This strategy may generate many false alarms in the foreground mask due to edge distortion from consecutive frames. To solve the edge-distortion problem, edge-segment-based methods have emerged to take advantage of the edge existence and its shape information [214]. An edge-segment approach consists of the concatenation of adjacent edges, and it inherits the problems of edges: shape and position changes. Thus, basic comparison of edge-segments produces similar results as edge-pixel-based approaches. To solve this problem, statistical edge-segment-based methods extract movement of edge-segments including edge distortion [390][444][283][391][282][443][284]. Thus, these methods solve the edge-variation problem by accumulating edge existence from a training sequence [285]. Practically, each accumulated region represents an edge- segment distribution. Each region refines their statistical properties after each frame to provide a stable background model. Since edge-based and edge-segment-based methods detect foreground as edges, these methods depend on a post-processing to extract the regions defined by the detected edges. Moreover, these methods have problems updating their background model to adapt the background.

5.2 Subpixel Edge-Maps Features

A key limitation of intensity gradient features is that they do not take spatial interactions into account. Alternatively, edge maps tag those background pixels which maximize local gradient in a neighborhood of pixels. This tagging increases selectivity which in turn reduces both the number of pixels which would have been discarded from the background model and the pixels would have been erroneously labeled as foreground [625][657]. Discretization errors in pixel-based edge maps generate unnecessary broad background models: a background edge halfway between the pixels

will require both pixels modeled as background, thus unnecessarily "blurring" the background model, which in turn reduces sensitivity to detecting regions. Instead of intensity gradient features, Jain et al. [243] proposed to use subpixel edge-maps by modeling the position and the orientation of subpixel edges which disambiguates between edges of the same orientation but at different positions and vice versa. Subpixel edge-maps has high precision and accuracy with invariance to illumination changes and suitable for small translations. An other advantage is that it requires fewer frames to build the background model even in case of slow moving objects and bootstrapping when edges from a region can share the same pixel as well as the same orientation.

5.3 Statistics on Gradient Features

Statistics on gradient features are also used as follows:

– **Gradient deviations:** Kamkar-Parsi [269] proposed to model the probability of appearance and disappearance of edges due to moving objects in the scene. This probability model is similar to the order consistency criteria described in [598] applied to gradient deviations instead of intensity values.

– **Projection Gradient Statistics:** Zhang [656] used projection gradient statistics as features. In the projection statistics curves, the largest change of the statistical value is caused by the moving objects. Thus, the gradient of the curve can reflect the statistical curve change clearly but also the position of the moving objects. First, the background is identified by projection statistics and projection gradient statistics of sub-images. Second, the background is reconstructed according to the results of each sub-image. Experimental results [656] show that this approach has a good anti-interference to low intensity vehicles in highway traffic scenes, and the processing time is less as well.

5.4 Discussion on Edge Features

The influence of the edge detector (Canny, Sobel, Prewitt) on robustness have not be studied in literature. Edges based approaches used in addition with intensity or color features are the most investigated approaches and allow to combine the advantage of the two features. For the approaches based on edges alone, only two main works emerged that are the works of based on edge segments[390][444][283][391][282][443][284] and the work based on subpixel edge [243]. These approaches appear to be relevant too and merit to be more investigated. Statistics on gradient features are useful to be more robust in presence of low intensities.

6 Texture Features

Texture features give ***spatial information***, and are the most investigated features in the field as can be seen in Table 5. The Local Binary Pattern (LBP) [207], and the Local Ternary Pattern (LTP) [332] are the most used with numerous variants. Furthermore, some improvements are developed in literature or can be developed to allow texture features to integrate both:

1. **Spectral and temporal information** such as Spatial-Color Binary Patterns (SCBP) [692], Opponent Color LBP (OCLBP) [316], Uniform LBP (ULBP) [636], Scene Adaptive LBP (SALBP) [404], Intensity LBP (iLBP)[558], eXtended Center-Symmetric [490] and Multi-Channel Scale Invariant LTP (MC-SILTP) [354].

2. **Spatial and temporal information** such as Spatial-Temporal LBP [486][659], Space-Time Center-Symmetric LBP (ST-CS-LBP) [322], Spatial Center-Symmetric LBP (SCS-LBP) [608], Motion Vectors Local Binary Patterns (MV-LBP) [617][573][572], Stereo LBP based on Appearance and Motion (SLBP-AM) [628], Space-Time Center-Symmetric Local Binary Patterns (ST-CS-LBP) [322], Center Symmetric Spatio-temporal LTP (CS-ST-LTP) [605], Spatio Temporal Scale Invariant LTP (ST-SILTP) [258], Texture Pattern Flow (TPF) [648][650] and dynamic texture patterns [628][74][387][96][694][257].

3. **Spectral, spatial and temporal informations:** There are no improvements in this category. Thus, investigation can be made in this direction.

Other approaches [568][570][593][523][606][300][638][43] reduce memories and computation cost of texture features. Moreover, other improvements was made by the use of statistical or fuzzy concepts. We classified them as follows :

1. **Crisp features:** such as local patterns [207][332][638][264][43][649][95][10][484] and spatio-temporal patterns [429][312][313][639][640].

2. **Statistical texture features:** such as peripheral patterns [459][629][630] and reach patterns [462][465][457][235][236][631][374].

3. **Fuzzy texture features:** Local Fuzzy Pattern (LFP) [340][412] and Fuzzy Statistical Texture (FST)[89][93].

In the following sub-sections, we briefly describe each texture feature in each category and the reader can see how the texture features are fused with other features in Section 17.

6.1 Local Patterns

6.1.1 Local Binary Pattern (LBP)

Heikkila et al. [207] proposed to use a texture feature called Local Binary Pattern (LBP) computed from the intensity feature. Based on this texture feature, a block-wise LBP histogram based approach (LBP-B) [207] and pixel-wise LBP histogram based one (LBP-P) [206] were developed. These methods can deal with gradual and sudden illumination changes due to the robustness of the texture feature to illumination variations. But these methods only use one learning rate that is a trade-off between different rates of change in background. When a high learning rate is used, the model updates quickly and slow-moving objects are incorporated into the background model which causes in a high false negative rate. When a low learning rate is used, these methods cannot handle sudden changes in background which causes in a high false positive rate. To address this problem, Goyal and Singhai [168] proposed a LBP texture-based algorithm which uses adaptive learning rate to deal with different rates of change in background. In an other approach, intead of computing LBP from the intensities, Satpathy et al. [466] proposed to apply LBP on the background and current edge images obtained by a Difference of Gaussians (DoG) edge detectors. Li et al. [321] combined the LBP difference between the background and the current images, the single Gaussian [590] and the codebook [288] for foreground detection to combine complementary advantages of each method. In a further work, Wu et al. [595][596] proposed a layered background modeling. First, every block on the first layer is modeled via texture based on local binary pattern (LBP) operators. Then, the modeling granularity is deflated onto the second layer to model via codebook. Layered match is done from top down when a new video frame enters. Experimental results [595][596] show that this approach efficiently avoids the false negative detection rate in the pixel-based background modeling when the object color is similar to the background, and also stops the false positives occurring on the contour areas of the moving objects due to the block model. In an other work, Zhong et al.[663] used intensity, LBP code and the norm of the first order derivative of the intensity with respect to x and y in a covariance based model to deal with dynamic backgrounds. In a further work, Zhong et al. [682] developed a background subtraction algorithm, which takes both texture and motion information into account. Texture information is represented by local binary pattern (LBP), which is tolerant of illumination changes and is computational simplicity. Assuming that there is significant structure in the correlations between observations across time, an operator to extract motion information is used. Then, each pixel is modeled as a group of texture pattern histograms and motion pattern histograms respectively. Finally, the texture pattern-based and motion pattern-based background model are combined using a weighted rule. A mixture factor γ is used to control the influence of the texture pattern-based background model and the motion pattern-based background model. Practically, it was taken to 0.5. This combination is more robust to dynamic backgrounds. These methods applied the original LBP in different ways compared to Heikkila et al. [207]. However, there are numerous variants of LBP and some of them have been developed for back-

ground/foreground separation. They can be classified as follows:

– **Spatial-Temporal LBP (STLBP):** A spatial background model, like LBP [207] (and RRF [460]), has the following problem. Their method uses spatial invariant features to monotonic changes in a local area. When pixel values change in a part of the local area, spatial invariant features are no longer good. Such a situation often occurs especially in outdoor scene changing of weather conditions. To address this problem, Shimada and Taniguchi [486] proposed an invariant feature using both spatial invariance and temporal invariance called Spatial-Temporal LBP (STLBP) suitable for outdoor scene in which the illumination condition can change gradually. In a similar way, Zhang et al. [659] extended the ordinary LBP from spatial domain to spatio-temporal domain, and proposed a new online dynamic texture extraction operator, named spatio-temporal local binary patterns (STLBP). Then, Zhang et al. [659] developed a dynamic background modeling and subtraction based on STLBP histograms which combine spatial texture and temporal motion information together. However, the computational load was increased and comparative results with the original LBP [207] were not presented.

– ϵ**-LBP:** The main limitation of LBP is that both memories and computation costs increase greatly with the increasing of the images resolution. To solve it, Wang and Pan [568] proposed a fast background subtraction method based on the novel LBP called ϵ-LBP. Furthermore, it overcomes two drawbacks brought by the LBP operator, i.e. the neighboring pixels are conditional independent under the center pixel, and it is weakly to measure the difference between the center pixel and its neighborhood. Compared with the LBP, ϵ-LBP improves greatly memories and computation efficiency by a simple measurement, which is linearly proportional of the images resolution. But, ϵ-LBP needs a threshold which is empirically selected as a global constant. Thus, this method only performs well when the illumination variation is global. To address this problem, Wang et al. [570] improved th ϵ-LBP by adding local adaptive property. The threshold is adaptively selected for each pair of two neighboring pixels. With two evaluation criterions, i.e. the description stability and the discriminative ability, a simple yet effective approach is presented to adaptively estimate the threshold by classifying all the pixels into two groups, i.e. the edge pixels and the texture pixels. In background modeling procedure, a naive Bayesian technique is adopted to effectively model the probability distribution of local patterns in the pixel level. The utilization of single ϵ-LBP (pixel level) improves the robustness to the illumination variation and reduces the computation cost compared with LBP.

– **Center Symmetric Local Binary Patterns (CS-LBP):** Tan et al. [523] proposed to use the Center Symmetric Local Binary Patterns (CS-LBP) [208], which effectively decreases the influence of noise, illumination changes and shadows. This method firstly computes the CSLBP feature images, then it calculates integral histogram as basic feature of background model, and finally establishes a background model containing K models. Finally, it classifies pixels as background

or foreground and updates the background model by comparing the similarity of histogram within L square neighborhood. This methods increases processing speed effectively by the use of one integral histogram and two integral histograms, respectively. To decrease the computation time, Wu and Zhu [593] proposed to compute the CS-LBP eigenvalue in the current image and to subtract it with the eigenvalue of the same pixel in the background image. The difference is then thresholded to classify pixel as background or foreground. This method could meet the requirements of real-time detection. In a further work, Zhang et al. [653] used an adaptive strategy based on CS-LBP. Then, foreground detection is used for detecting moving object by using a confidence factor to determine whether the current pixel is a background or foreground pixel. The confidence factor of the current pixel is computed in terms of the difference values of its neighbourhood pixels. Experimental results made with PETS 2009, BMC 2012 [552] and SABS [61] datasets demonstrate that this approach can robustly detect moving object under various scenes. An other approach proposed by Li et al. [322] consists in a space-time symmetrical ST-CS-LBP operator which integrates time prediction and texture information. A ST-CS-LBP histogram is built and merges the advantages of time-domain statistics and spatial distribution. Results show more robustness to long and short luminance changes than the CSLBP.

In an other way, Xue et al. [608] proposed to use a Spatial Center-Symmetric Local Binary Pattern (SCS-LBP). This operator not only has the property of illumination invariance, but also produces short histograms and be more robust to noise. So, Xue et al. [608] extended the CS-LBP operator from spatial domain to spatial-temporal domain and proposed a texture operator named SCS-LBP which extracts spatial and temporal information simultaneously. Then, combining the SCS-LBP operator with an improved temporal information estimation scheme, Xue et al. [608] obtained a background modeling approach which reach high accurate detection in dynamic scenes while reducing the computational complexity compared to the LBP based method. The LBP operator produces long feature set since it only adopts the first-order gradient information between center pixel and its neighbors. Xue et al. [606] developed a second-order center-symmetric local derivative pattern (CS-LDP) operator which extracts more detail local information than CS-LBP. Then, Xue et al. [606] concatenated the CS-LBP histogram and CS-LDP histogram to get a new hybrid feature. This LBP pattern is called Hybrid Center-Symmetric LBP (HCS-LBP) Experiments on challenging sequences indicate that this method can produce comparable results while using less computation time (25%) compared to the LBP based method.

– **Spatial-Color Binary Patterns (SCBP):** Zhou et al. [692] proposed an spatial-color feature extraction operator named spatial-color binary patterns (SCBP). It extracted spatial texture and color information. In addition, a refine module is designed to refine the contour of moving objects. For each pixel, first, a histogram of SCBP is extracted from the circular region, and then a model consist of several histograms is built. For a new observed frame, each pixel is labeled either background or foreground according to the matching degree between its SCBP histogram and its model, then the label is refined and finally the model of this

pixel is updated.

– **Opponent Color Local Binary Pattern (OC-LBP):** Lee et al. [316] developed a novel training process for classifiers which use block based Opponent Color Local Binary Pattern (OC-LBP). So, the pixel based OC-LBP [168] was extended to to block-level.

– **Double Local Binary Pattern (DLBP):** Xu et al. [600] developed a Double Local Binary Pattern (DLBP).

– **Uniform Local Binary Pattern (ULBP):** Yuan et al. [636] proposed in a K histograms model a combination of color feature, i.e hue, and an uniform local binary pattern (ULBP) texture to be robust to shadows. The results from each features are combined with a weighted average. ULBP with the hue outperforms LBP [207] and DLBP [600].

– **Extended and Rotation Invariant LBP (Ext-LBP):** Yue et al. [637] developed an extended LBP (Ext-LBP) and a Rotation Invariant LBP (RI-LBP). Ext-LBP is obtained by expanding the adjacent area of original LBP. Combined RI-LBP with MOG effectively removes shadow influences.

– **Larger neighborhood LBP (LN-LBP):** Kertesz [275] used a larger neighborhood than the original LBP (LN-LBP). It is calculated inside a 5×5 neighborhood, therefore, it is not defined two pixels wide on the image borders. The algorithm specified a correction value to the LN-LBP calculation in order to handle the flat color areas where the color values almost do not change. Furthermore, Kertesz [275] used a Markov Random Field (MRF) as a higher level classification of the LN-LBP histograms into foreground and background. LN-LBP with MRF is quite invariant for the resolution and provides better performance at higher resolutions than the original LBP [206].

– **Motion Vectors Local Binary Patterns (MV-LBP):** Yang et al. [617], Wang et al. [573] and Wang et al. [572] proposed a moving object detection method towards H.264 compressed surveillance videos. First, the motion vectors (MV) are accumulated and filtered to achieve reliable motion information. Second, considering the spatial and temporal correlations among adjacent blocks, spatiotemporal Local Binary Pattern (LBP) features of MVs are extracted to obtain coarse and initial object regions. Finally, a coarse-to-fine segmentation algorithm of boundary modification is conducted based on the DCT coefficients.

– **Scene Adaptive Local Binary Pattern (SALBP) :** Noh and Jeon [404] developed a texture operator namely, Scene Adaptive Local Binary Pattern (SALBP) that provides more consistent and accurate texture-code generation by applying scene adaptive multiple thresholds. A background subtraction framework employing diverse cues (pixel texture, pixel color and region appearance) is presented. The SALBP information of the scene is clustered by the conventional

codebook model [288], and utilized to detect initial foreground regions. Background statistics of the color cues are also modeled by the codebook model and employed to refine the texture-based detection results by integrating color and texture characteristics.

– **Stereo Local Binary Pattern based on Appearance and Motion (SLBP-AM) :** Yin et al. [628] proposed a Stereo Local Binary Pattern based on Appearance and Motion (SLBP-AM) descriptor. The motion of pixels is represented as dynamic texture in ellipsoidal domain. Then, Yin et al. [628] combined texture histograms in the XY, XT, YT planes in the ellipsoid. SLBP-AM is more robust to slight disturbance, but also adapts quickly to the large-scale and sudden changes.

– **Window-based LBP (WB-LBP):** The histogram computation and construction of LBP is a very time-consuming and complex process. Kumar et al. [300] proposed to reduce the complexity to a large extent by using a window-based LBP (WB-LBP) subtraction method. Moreover, the efficacy of the WB-LBP in terms of correct classification is quite satisfactory as compared to the other LBP-based methods.

– **Intensity Local Binary Patterns (iLBP):** The original LBP [207] has a main drawback. Indeed, it ignores the intensity information when comparing LBP descriptors. Because of this, there could be a paradoxical situation which generates wrong pixel comparison result when intensity values of pixels differ drastically, but their LBP descriptors are identical. To solve it, Vishnyakov et al. [558] proposed a intensity Local Binary Patterns (iLBP) descriptor and built a fast background model on its basis. This feature allows stabilizing the value of the descriptor and constructing a background model that is robust to lighting conditions changes in the scene and is applicable for the real time multi-camera setup. In a further work, Vishnyakov et al. [559] used iLBP in a background model based on regression diffusion maps. This approach allows objects that move with different speed or even stop for a short while to be uniformly detected.

– **BackGround Local Binary Pattern (BGLBP):** Davarpanah et al. [111] developed a BackGround LBP (BGLBP) which has been designed to inherit the positive properties of Direction Local Binary Pattern (D-LBP), CS-LBP [523], ULBP [636], and RI-LBP [637]. Experimental results on the I2R dataset [326] show that BGLBP outperforms several variants of LBP such as the original LBP, ULBP and RI-LBP in presence of illumination changes and dynamic backgrounds.

– **eXtended Center-Symmetric Local Binary Pattern (XCS-LBP):** Silva et al. [490] presented an extension of CS-LBP called XCS-LBP (eXtended CS-LBP) by comparing the gray values of pairs of center symmetric pixels so that the produced histogram are short as well, but considering the central pixel also. This combination makes the resulting descriptor less sensitive to noise. Experimental results on the BMC dataset [552] show that XCS-LBP outperforms LBP, CS-LBP and CS-LDP. For computation time, XCS-LBP has slightly better time perfor-

mance than both CS-LBP and CS-LDP.

- **Local SVD Binary Pattern (LSVD-BP):** Guo et al. [187] developed the LSBP feature descriptor which has the ability to gain the potential structures of local regions. LSBP also inhibits the effect of illumination changes especially cast shadows and noise. Experimental results on the ChangeDetection.net dataset [169] show that LSBP allows robustness in the "Shadow" and "Thermal" categories while LSBP seems to perform at a level comparable to PBAS [209] the "Dynamic Backgrounds" and "Baseline".

In summary, three main properties are required for a designed LBP version for the application of background modeling and foreground detection [111]: *1)* It should be fast to compute, *2)* the number of bins in the plotted histogram should be the least, and *3)* It should be computed based on whole pixels values belonging to each block. Thus, the original LBP [207][206] and the LBP based on gradient [466] are not optimal. Multi-Block LBP calculates a LBP value for each block instead of each pixel separately. ULBP [636] is not rotation invariant. Finally, there are many bins in the CS-LBP model. Regarding all of these limitations, BGLBP and XCS-LBP seem to be the best LBP variants for background modeling and foreground detection.

6.1.2 Local Ternary Pattern (LTP)

- **Local Ternary Pattern (LTP):** First, Tan and Triggs [524] proposed to include an additional buffering state to solve the instability problem of LBP, and introduced the local ternary pattern (LTP) as a robust extension of LBP. Liao et al. [332] proposed to use the Local Ternary Pattern (LTP) operator [524] for background/foreground separation. Practically, LTP is more robust by introducing a small tolerative range. However, the descriptor is extended from LBP by simply adding a small offset value for comparison, which is not invariant under scale transform of intensity values by a multiplying constant. For example, LTP descriptor can not keep its invariance against scale transform when all local pixel values are multiplied by 2.

- **Scale Invariant Local Ternary Pattern (SILTP):** For background modeling and foreground detection, the intensity scale invariant property of a local comparison operator is very important, because illumination variations, either global or local, often cause sudden changes of gray scale intensities of neighboring pixels simultaneously, which would approximately be a scale transform with a constant factor. Therefore, Liao et al. [332] proposed to extend LTP to Scale Invariant Local Ternary Pattern (SILTP). Furthermore, Liao et al. [332] improved Heikkila et al's region histogram based method by modeling pixel process with a single local pattern instead. However, local patterns are not ordinal numerical values, thus can not be modeled directly into traditional density functions. Therefore, Liao et al. [332] developed a Pattern Kernel Density Estimation (PKDE) technique to effectively model probability distributions of such patterns for handling

complex dynamic backgrounds and a multiscale fusion scheme to consider the spatial scale information. The SILTP operator was presented by 8 bits [332]. This encoding strategy only used 4 pixels of its 8 neighborhoods which might result in loss of the texture information. If all 8 neighboring pixels were used, the length of encoding was 16 bits which led to computational complexity increasing.

- **Scale Invariant Center-symmetric Local Ternary Pattern (SCS-LTP):** Zhang et al. [670] proposed a texture operator named Scale Invariant Center-symmetric Local Ternary Pattern (SCS-LTP), and a corresponding Pattern Adaptive Kernel Density Estimation technique for its probability estimation. Zhang et al. [670] used a simplified Gaussian Mixture Models for intensity feature. Then, the results from texture and intensity are combined in a multi-scale fusion scheme based on the basic product formulation of the likelihoods. Zhang et al. [669] used a similar scheme but based on Gaussian Mixture Models for both texture (SCS-LTP) and color features (Normalized RGB). The results from texture and color are combined in a multi-scale fusion scheme based on the weighted average. The weights are a function of a parameter balancing texture distance and color distance which is empirically set to 0.7. There are two advantages of the SCLTP operator in foreground/background separation. The first one is that the SCLTP operator is robust not only to noise but also to illumination by introducing a scale factor like SILTP. The second one is that the SCLTP operator can represent more texture information with less bits. The SCLTP operator can use 8 bits to express a pixel with all its 8 neighboring pixels because of only comparing center-symmetric pairs of pixels. Experimental results [670][669] on several complex real world videos with illumination variation, soft shadows and dynamic backgrounds (I2R dataset [326]) show that SCS-LTP outperforms slightly SILTP [332].

- **Multi-Channel Scale Invariant Local Ternary Pattern (MC-SILTP):** Ma and Sang [354] proposed to extends the SILTP to feature space and to operate on the three channels of RGB images rather than the only channel of gray images to get the texture pattern. This texture descriptor is called Multi-Channel Scale Invariant Local Ternary Pattern (MC-SILTP).

- **Center Symmetric Spatio-temporal Local Ternary Pattern (CS-ST-LTP):** Xu [605] and Lin et al. [337] decomposed the scene background into a number of regular cells, within which a batch of video bricks (e.g.pixels) are extracted as observations. In order to better represent video bricks and enhance the robustness to illumination variations, Xu [605] proposed a brick-based descriptor, namely Center Symmetric Spatio-temporal Local Ternary Pattern (CS-ST-LTP), instead of using pixel intensity. CS-ST-LTP is inspired by the 2D local pattern descriptor and adapted to characterize video brick..

- **Spatio-Temporal Scale Invariant Local Ternary Pattern (ST-SILTP):** Considering the temporal persistence of texture sequences, Ji and Wang [258] extended the SILTP [332] to the spatio-temporal domain called ST-SILTP. Second, Ji and Wang [258] presented an adaptive fusion approach of color and texture to com-

pensate for their respective defects. Furthermore, since a pixel of foreground or not depends on not only itself but also its neighborhood, a lateral inhibition filter model incorporated the neighborhood information into calculating the pixel's confidence score. Practically, a pixel is classified as background or foreground by using its probability computed from the similarities in each feature, and from confidence of color and texture components. For color or texture feature of a background mode, if its corresponding confidence is a larger value, it means that it plays a more important role in making the decision about the label of the given pixel, i.e., foreground or background. In the computation of updating the background model, the weight and confidences of color and text features are constantly updated in the online learning way. Thus, in estimating the probability of a pixel as background, the contribution of color and texture components can be adaptively adjusted and both kinds of information are fully fused together to segment foreground object. Experimental results [258] on I2R dataset [326] show that ST-SILTP gives silhouettes with less holes than the original SILTP [332].

– **Center-Symmetric Scale Invariant Local Ternary Pattern (CS-SILTP):** Wu et al. [592] extended the SILTP descriptor by introducing a the Center-Symmetric Scale Invariant Local Ternary Pattern (CS-SILTP) descriptor, by exploring spatial and temporal relationships of neighborhood.

6.1.3 Local States Pattern (LSP)

Yuk et al. [638] introduced Scale Invariant Local States (SILS) as texture features for modeling a background pixel, and a pattern-less probabilistic measurement (PLPM) which is derived to estimate the probability of a pixel being background from its SILS. An adaptive background modeling framework was also proposed for learning and representing a multi-modal background model. Practically, considering N neighbors in the texture pattern and K background models, the memory space and operations required for the SILS based method are only $O(3NK)$ comparing to SILTP based method [332] which requires $O(3NK)$. Experimental results [638] on the I2R dataset [326] show that SILS based method runs nearly 3 times faster than SILTP based method [332] with the same accuracy.

6.1.4 Local Difference Pattern (LDP)

Yoshinaga et al. [633] proposed a probabilistic background model which combined pixel-based multi-modal model with color features and spatial-based uni-modal model with texture features by considering the illumination fluctuation in localized regions. Then, Yoshinaga et al. [633] used several pairs of a focused pixel and its peripheral pixels, i.e., its surrounding pixels, in a localized region, and modeled the distribution of the difference between pixel values of each pair with a mixture of Gaussians. These pixel value differences in the localized region is called Local Difference Pattern (LDP). LDP presents several advantages: **(1)** There are little changes in a LDP in presence of sudden illumination changes because the pixel values in a localized

region similarly increase and decrease, and **(2)** LDP can also deal with periodic changes of pixel values because MOG represents multiple hypotheses of the background. Thus, LDP used both properties of pixel-based and spatial-based features, without decreasing the accuracy. In further works, Yoshinaga et al. [634][635] called this feature Statistical Local Difference Pattern (SLDP). Experimental results on the BMC dataset [552] show that SLDP outperforms adaptive RRF [460] in presence of illumination changes and dynamic backgrounds.

6.1.5 Local Self Similarity (LSS)

LBP [207] and SPF [530] performed reasonably well, but the features that are used can only capture change in texture, not change in intensity. Furthermore, because these features are computed based on comparisons with a center pixel, change cannot be detected if the intensity of the center pixel remains larger (or smaller) than each neighboring pixel after a change in a scene. Recently, Jodoin et al [264] proposed to use the local self-similarity (LSS) descriptor, but because the LSS descriptor is calculated on a large region, there are borders of falsely detected pixels around the detected foreground. Some morphological operations were applied to improve the detected foreground by removing extra pixels, but holes or spaces between legs cannot be recovered that way. Furthermore, LSS is slow to compute on complete image.

6.1.6 Local Similarity Binary Pattern (LSBP)

To solve the shortcomings of both LBP and LSS, Bilodeau et al. [43] proposed a descriptor called Local Similarity Binary Pattern (LSBP) which is binary and fast to compute, works on small regions, and captures both change in texture and change in intensity. Practically, a binary vector is built from the comparisons of pixel intensities centered around a point of interest over a small predetermined pattern. Unlike LBP and LSS which compute the difference between two values, the LBSP approach returns whether they are similar or not via absolute difference. Its temporal aspect and sensitivity to illumination variation is due to its ability to use the central pixel intensity of a previous frame for new comparisons. But, one of the disadvantages of LBSP is that it is not a spatio-temporal descriptor because it does not keep both feature information and pixel intensity information jointly up to date. To address this problem, St-Charles et al. [505] presented a spatio-temporal LSBP in the framework of an adaptive background subtraction method called LOcal Binary Similarity segmenTER (LOBSTER). This modification allows the spatio-temporal LSBP to be more suitable in noisy or blurred regions and more robust to high illumination variations than the original LSBP. In further works, St-Charles et al. used this spatio-temporal LSBP in addition with color features with a flexible background subtraction algorithm called Self-Balanced SENsitivity SEgmenter (SuBSENSE) [506][507], and a background subtraction algorithm that analyzes the periodicity of local representations called Pixel-based Adaptive Word Consensus Segmenter (PAWCS) [509][510].
In a further work, Yan et al. [613] improved LSBP by proposing an Uniform LSBP (U-LSBP) which reduces computational complexity and is more robust to shadows

and illumination changes. In an other work, Zhang et al. [658] proposed to combine color and LBSP information to establish the background model.

6.1.7 Directionnal Rectangular Pattern (DRP)

Zhang et al. [649] developed a Directional Rectangular Pattern (DRP) based complex background modeling method to detect the moving objects in a video sequence. Different from LBP encoding the binary result of first-order derivative between the central point and its neighborhoods, Directional Rectangular Pattern encoded the binary result of first and second order derivative direction in all neighborhoods among a rectangular region. To model the distribution of the DRP micro-patterns, Zhang et al. [649] used DRP integral histograms. The local gray-level feature based Gaussian Mixture Model (GMM) is exploited to calculate an adaptive threshold for the histogram similarity measure to decide which part/pixel is background or moving object. Experimental results [649] show the effectiveness of DRP by comparing with LBP.

6.1.8 Local Color Pattern (LCP)

All LBP-based algorithms are often invariant to local illumination changes but they are unable to detect uniform foreground objects in large uniform background except at the objects'edges. To solve this problem, Chua et al. [95] proposed a robust texture-color based background modeling. The texture feature is the LBP histogram and the color feature called Local Color Pattern (LCP) is is formed by concatenating the quantized hue, luminance, and saturation histograms, summed over a structuring element. For the initialization, LBP and LCP histograms are computed and stored as background models for the first N image frames. Since the background model is represented by LBP and LCP histograms, the final similarity is obtained by computing a weighted rule with a weight τ that controls the importance of texture and color features. Higher τ indicates that texture is more important than color features. τ is determined adaptively. The ability to adapt the weight of color and texture information makes the algorithm very suitable for video surveillance applications especially with dynamic scenes. Reckley et al. [446] used LCP features in a sensor selection scheme in which spatiotemporal signatures of moving objects are integrated from different sensing modalities into a video segmentation method in order to improve object detection and tracking in complex scenes such as dynamic backgrounds with moving water and high reflections. Experimental results [446] on two complex datasets demonstrate that this technique significantly improves the accuracy and utility of the original LCP [95][96], and largely outperforms CS-LBP [523].

6.1.9 Local Neigborhood Patterns (LNP)

Amato et al. [10] used two discriminative features based on angular and modular patterns, which are formed by similarity measurement between two sets of RGB color vectors: one belonging to the background image and the other to the current image. This Local Neighborhood Patterns (LNP) improved foreground detection in the presence of moving shadows.

6.1.10 Local Ratio Pattern (LRP)

In LBP and LTP, the very coarse binning of local intensity ratios may result in substantial feature noise when the true ratios fall close to a bin boundary. To solve this problem, Zaharescu and Jamieson [641] proposed the Local Ration Pattern (LRP) in a multi-scale multi-feature codebook-based background subtraction method. LRP is similar to LBP and LTP, but extended to 4 bits, in order to divide the range of possible intensity ratios into 16 bins, instead of 2 or 3. This finer-grained encoding allows the feature to respond more proportionately to a change in the underlying image. Then, a similarity between two LRPs is defined with a ratio. Furthermore, Zaharescu and Jamieson [641] addressed the LRP reliability by evaluating the confidence of a match between two LRP features with their flatness (which are their similarity score to a perfectly flat LRP).

6.1.11 Local Ray Pattern (LRP)

Light field camera was originally proposed for image-based rendering for computer graphics. But, the light field camera has been applied to solve a difficult computer vision and pattern recognition problem too [358]. In this context, Shimada et al. [484][485] proposed a feature representation, called Local Ray Pattern (LRP) to evaluate the spatial consistency of light rays. The combination of LRP and GMM-based background modeling realizes object detection on the infocus plane. Experimental results [484][485] demonstrate the effectiveness and applicability for video surveillance.

6.2 Spatio-temporal Patterns

6.2.1 Spatio-Temporal Vectors

Pokrajac and Latecki [429] proposed to use as features Spatio-Temporal Vectors (STV), that are 3D blocks vectors. So, Pokrajac and Latecki [429] decomposed a given video into spatio-temporal blocks ($8 \times 8 \times 3$ blocks). Then, a dimensionality reduction technique is applied to obtain a compact representation of color or gray level values of each block as vector of just a few numbers. The block vectors provide

a joint representation of texture and motion patterns. Then, the MOG model is used on the spatiotemporal blocks. So, the MOG model [512] is applied on the pixel and region levels with a single level texture representation, that is the 3D block, whereas it has been applied on pixel level in the original MOG. This method implicitly assumes the feasibility of computing projection matrix from blocks that adequately represent the texture. This approach provides good results on videos with comparatively high stationarity in background. Furthermore, improvements are possible if the projection matrix is computed dynamically. However, the techniques for adaptive estimation of projection coefficients in time are not studied.

Texture at a given pixel is very likely to highly vary when a moving object is passing through this location. Therefore, Latecki et al. [311][312][313] proposed to use a local variation of the texture vectors. To robustly measure this variation, Latecki et al. [311][312][313] measure it in a limited and as short as possible window of time, since at a given pixel a moving object can quickly appear or disappear. So, the local variation is defined as the largest eigenvalue of spatio-temporal texture vectors in a small time window. It is computed by applying PCA to the covariance matrix of the SP texture vectors within a small temporal window. This way, Latecki et al. [311] indirectly determined the magnitude of texture variability in the direction of its maximal change. Thus, Latecki et al. [311][312][313] used PCA twice, first time to compute the sp texture vectors, and the second time to compute the variation of a set of texture vectors in a given time window. The decision whether a moving object or a stationary background is identified at a given spatiotemporal location is then made by dynamic thresholding of the obtained eigenvalues.

Zeljkovic et al. [643] and Pokrajac et al. [432] studied the resilience of moving objects detection algorithm based on the spatio-temporal blocks on additive Gaussian noise. Experimental results [643][432] on the PETS 2001 dataset show that the STV is robust to strong additive Gaussian noise.

Pokrajac et al. [431] evaluated the STV on monochrome and multi-spectral IR videos. Experimental results show that the STV can provide low false positive error rates and successful identification of the front edge of the moving object.

Latecki et al. [309] extended the original method of STV [429] by replacing the dynamic threshold with dynamic distribution learning and outlier detection significantly improving the performance of the original approach.

Miezianko and Pokrajac [368] proposed a local background dissimilarity measurement based on wavelet decomposition of localized texture maps. Dynamic threshold of the normalized dissimilarity measurement identifies changed local background blocks, and spatial clustering isolates the regions of interest. The use of STV provides accurate and illumination invariant separation of foreground and background textures. The rough dissimilarity measurement of collected background texture maps and the dynamic thresholding locate regions of interest in the background that exhibit significant changes. The clustering of the blocks in localized spatial regions establish

the boundaries of the background at different times. The texture maps are created with gray level values to increase the computational speed of the wavelet decomposition and background change detection.

Miezianko and Pokrajac [367] developed an effective method for extracting changed backgrounds for regions observed by multiple overlapping cameras. A multi-layers background is constructed, and spatio-temporal texture blocks are used to detect motion. Detected motion delineated temporally cohesive non-moving regions to extract hyperspherical clusters and detect changed backgrounds. The proposed method allows for detecting background changes in crowded environments exhibiting large motion flows and significant occlusion.

6.2.2 Spatio-Temporal Textures

Yumiba et al. [639][640] proposed a spatio-temporal texture called Space-Time Patch (ST-Patch) which describes motion in addition to appearance for detecting moving objects in presence of dynamic changes. This approach can cover global changes by using appearance information as conventional spatial textures. In addition, it can cover local changes by using motion information. Yumiba et al. [639] applied the MOG model with the ST-Patch. Experimental results [639][640] show that ST-Patch allows the MOG to be more robust on videos with dynamic backgrounds and illumination changes.

6.2.3 Spatio-temporal Features

In a multi-level approach, Tanaka et al. [530][528] and Nonaka et al. [405] proposed a integrated background modeling based on spatio-temporal features. First, foreground objects are detected based on the pixel-level background modeling which is a KDE model with the location (x,y) and the RGB color components as features and based on the region-level background modeling with Radial Reach Correlation (RRC) as feature. In the further step, the two foreground masks are combined. That is, pixels which are classified as foregrounds by both of the pixel-level and region-level are classified as foregrounds and other pixels are classified as backgrounds. Then, a frame-level background modeling which is based on brightness normalization of a model background image Fukui et al. [152][151], is used to be robust against sudden illumination changes. Experimental results on the CD.net 2012 dataset show that this multi-level approach with the considered spatio-temporal features outperforms the original RRC [460], the original MOG [512] and the improved KDE [525] in presence of dynamic backgrounds and illumination changes.

6.3 Statistical Texture Features

6.3.1 Peripheral Patterns

– **Peripheral Increment Sign Correlation (PISC):** Satoh et al. [459] proposed Peripheral Increment Sign Correlation (PISC) feature that encodes the intensity differences between a target pixel and its peripheral pixels as a 0/1 binary code similar to the case of Local Binary Pattern (LBP) proposed by Heikkila et al. [207]. However, this leads to increase false positives because the code is reversed easily with slight intensity changes in regions with small intensity differences, for example in plain regions. Plain regions often occupy large spatial region within images, which makes stabilizing on them very important.

– **Peripheral TErnary Sign Correlation (PTESC):** In order to alleviate the stabilization issue of PISC, Yokoi [629] proposed a Peripheral TErnary Sign Correlation (PTESC) features that stabilized the encoding by using -1/0/1 ternary code.

Though these texture-based methods are robust against illumination changes, they cannot cope with regions that have poor texture. Plain foreground objects before plain background with different intensity from foreground cannot be detected by these methods because both foreground and background have the same plain texture. For this, Yokoi [630] proposed a texture descriptor combining Peripheral TErnary Sign Correlation (PTESC) [629] and Bi-Polar Radial Reach Correlation (BPRRC) [457][461]. The idea is to take advantages of the both descriptors. Indeed, Peripheral TErnary Sign Correlation (PTESC) is robust against illumination changes by using -1/0/1 ternary code for encoding the intensity difference between pixels in texture, and Bi-polar Radial Reach Correlation (BPRRC) [457][461] yields high detectability in a region with little texture.

6.3.2 Reach Patterns

– **Radial Reach Correlation (RRC):** Satoh et al. [462][465] defined a texture feature called Radial Reach Correlation (RRC) (also called Radial Reach Filer [458][464][463][460][560])which achieves robust detection of moving objects while remaining insensitive to moving object intensity distribution and intensity changes in scene objects. RRC evaluates foregroundness based on local texture described in the magnitude relation between the center pixel and its neighbor pixel. In principle, this magnitude relation is not affected by the changes of illumination and, thus, RRC is more robust than features which only use distribution information of the center pixel values. However, it cannot handle the changes of the textural information caused by the small background fluctuation such as swaying tree leaves [529]. In further works, Satoh and Sakaue [457][461] proposed Bi-Polar Radial Reach Correlation (BP-RRC) which expands RCC in cases where the image texture is feeble and the intensity distribution is biased. Thus, Satoh and Sakaue [457][461] used a mechanism that simultaneously defines and utilizes intensity differences (positive or negative) relative to a focal

pixel. Experimental results [457][461] with the RCC showed that, even in cases where image texture is weak or intensity distribution is biased, BP-RCC offers superior detection performance and stability. BP-RRC is robust against gradual illumination changes by using texture model but its present several weakness: *1)* it is not robust against background movements because of its rigid texture model [631], *2)* Since BP-RRC uses the texture information, it does not detect offset/gain change, and *3)* it cannot respond to the texture change due to illumination change under multiple light sources. To address the first weakness, Yokoi [631] proposed the Probabilistic BP-RRC (PrBP-RRC) which preserves BP-RRCs robustness against illumination changes and adds the robustness against background movements. PrBPRRC introduces a probabilistic model for background texture and learns a probabilistic background with inputs including background movements and presence of moving objects. Experimental results [631] on ATON dataset [369] and PETS 2007 dataset show that PrBP-RRC outperforms BP-RRC. To solve the second problem and the third problem, Miyamori et al. [374] proposed the Adaptive Bi-Polar Radial Reach Correlation Mixture Model (ABP-RRC), which generates and selects several BP-RRC accommodative to background changes. Thereby, Miyamori et al. [374] developed a stable background/foreground separation method, even when the texture changes occurred which various illumination change causes. To reduce the computation time of RRC, Itoh et al. [235][236] developed the Fast Radial Reach Correlation (F-RRC). As a result, Itoh et al.[235][236] reduce the RRC calculation process to about $1/4$ in comparison with normal RRC. In an other approach, Tanaka et al. [527][526][529] improved the original RRC [462][465] which is further used and combined in pixel-level and region level method.

– **Radial Proportion Filter (RPF):** Miyamori et al. [375] defined the Radial Proportion Filter (RPF). In a further work, Miyamori et al. [376] developed a Multi-RPF (MRPF).

– **Statistical Reach Feature (SRF):** Iwata et al. [292] defined a set of statistical pair-wise features, derived by intensity comparison in a local neighborhood. The classification for each pixel P is achieved by a set of selected N points Q_n ($1 \leq n \leq N$), which are called reference points. Thus, SRF can be described as the selection of Q_n. Three factors impact the search of Q_n : *1)* the absolute value of the intensity difference between P and Q_n must exceed a given threshold T which has an important role because it allows the background model to tolerate noise, *2)* Q_n needs to meet a statistical requirement, that is its intensity remains T units smaller (or larger) than that of P in most images, and *3)* Q_n is searched from the starting point P to the edge of image in N (with $N = 8$) directions. Then, two types of Q_n can be selected. SRF defines the sign between P and Q_n, which satisfies $I_t(P) - I_t(Q_n) \geq T$ in most images, as $SRF(P, Q) = 1$. It also defines the sign between P and Q_n, which satisfies $I_t(P) - I_t(Q_n) \leq T$ in most images, as $SRF(P, Q) = -1$. Then, comparing the signs in the background model with that in the current image allows to classify P as a background or foreground pixel, depending on whether the sign between P and Q_n has changed. Due to the spe-

cial properties of point pairs, SRF works well in object detection, but SRF present three weakness: *1)* SRF may not search a sufficient number of Q_n because SRF searches for Q_n in only eight directions, implying that most pixels are not taken into account. This may lead to an insufficient number of Q_n that causes drawbacks in the foreground detection step, *2)* the searching way of Q_n is not optimal because SRF searches Q_n in the order of space instead of the order of intensity difference, which leads to an incomplete detection. The larger the intensity difference is, the less sensitive the model becomes. The intensity difference between P and Q_n searched by SRF tends to be too large. This leads to incomplete detection, and *3)* the one-sided criterion in the SRF detection step results in false detections. A one-sided criterion means that the two signs of SRF are combined into one binary decision. In the case of presence of a moving object with much brighter color, pixel are misclassified as background. To solve these problems, Zhao et al. [677] improved SRF both in theory and algorithm.

6.4 Fuzzy Texture Features

– **Local Fuzzy Pattern (LFP):** Ouyang and Chen [412] proposed a texture descriptor called the Local Fuzzy Pattern (LFP) histogram. The LFP histogram of each pixel is calculated for each new frame and is compared to its corresponding values in the background model for classifying the pixel into foreground objects. Experimental results [412] show that LFP possesses better adaption capability and tolerance for dynamic conditions such as shadows and illumination variation in comparison to LBP. Liu et al. [340] developed a similar LFP.

– **Fuzzy Statistical Texture (FST):** Chiranjeevi and Sengupta [89] proposed to apply a fuzzy membership transformation on the co-occurrence vector. The idea is to derive a fuzzy transformed co-occurrence vector with shared membership values in a reduced dimensionality vector space, and called Fuzzy Co-occurrence Vector (FCV). FCV can handled better the dynamic backgrounds than the crisp Co-occurrence Vector (CV). Then, Chiranjeevi and Sengupta [89] defined a normalized FCV (NFCV) from which fuzzy statistical texture features are derived. Practically, the background model is initialized with a feature vector, composed of intensity, energy, texture mean and local homogeneity, obtained from the first frame of the video sequence. Then, FST are fused with intensity by using the Choquet integral. This approach is called IFST. For a fair comparison, Chiranjeevi and Sengupta [89] implemented the approaches with Statistical Features (FT), called IST. Experimental results [89] show that IFST handled better dynamic backgrounds than IST, the Choquet integral with (RGB) features [22], and the Choquet integral with (R,G,LBP) features [24]. In a further work, Chiranjeevi and Sengupta [93] used the same FST combined with intensity but aggregated with he intervalued-valued fuzzy Choquet integral instead of the real-valued Choquet integral used in Chiranjeevi and Sengupta [89].

6.5 Others Texture Features

6.5.1 Texture Pattern Flow

Both LBP and SILTP-based approaches do not consider the temporal variations in patterns which contains inherent motion information derived from moving objects in the video. In order to address this problem, Zhang et al. [648] developed a spatial-temporal feature, dubbed as the Texture Pattern Flow (TPF), to compute inherent motion information. TPF encodes texture and motion information in a local region from both spatial and temporal aspects. TPF features are based on the gray-level image in order to increase robustness against complex and dynamic backgrounds. The integral histogram of the TPF is computed within a local region around the pixel and is employed to extract statistically discriminative features. After estimating the TPF integral histogram features, the background model is built and dynamically adapted to complex scenarios as new frames begin to be processed. A similarity measurement between the current image and the background model is made using a Kernel Similarity Modeling (KSM) approach which integrates the TPF integral histogram features. Zhang et al. [650] proposed an improvement to this method using an adaptive threshold.

6.5.2 Textons

Texture Pattern Flow (TPF) [648] is based on the assumption that the movement of the foreground is always towards a certain direction. While this assumption can mostly work with human movements in constrained environments, it may not hold for animals such as a fish as their movement is fairly erratic with frequent direction changes [427][20][19][504]. However, other texture feature called textons have been extensively adopted in texture analysis, and can be defined as sets of patterns shared over an image [266]. Practically, several texton images have been proposed in the literature [342]. For fish detection in underwater scenes, Spampinato [504] adopted 7 textons to capture even the slightest texture variations within any considered rectangular region of size $w \times w$ centered on the considered pixel. As a global texture feature describing each pixel (x, y) and its neighbors in the region, the energy of the texton is computed. Therefore, each pixel is represented with a feature vector which contains the location (x, y), the three color components (RGB) and the energy of the textons. The joint domain-range model consists in the corresponding 6-dimensional space, on which the pdfs of the background and foreground models are built. This is performed by means of KDE. Experimental results [504] on the Fish4Knowledge dataset [502] show that the texton feature allow the KDE model to be more robust than the original KDE with RGB [142], SILTP [332], VKS with rgb [394] and VKS with Lab and SILTP. [394]. In an other work, Panda and Meher [414] used a Texton Co-occurrence Matrix (TCM) feature with the original MOG model [512] to deal with dynamic backgrounds. TCM is widely used in the field of image retrieval. The TCM feature integrates both the color, texture, and shape features and is computed in a neighborhood region of each pixel. Thus, TCM feature implicitly used image

features and the spatial relationship between the pixels. Experimental results [414] on the I2R dataset show that the MOG with the TCM feature show better robustness than original OG with RGB, covariance-based background subtraction [663], moments-based background subtraction [361], the original LBP [207] and intensity and texture-based background subtraction [88] in presence of dynamic backgrounds.

6.5.3 Galaxy Pattern

Liu et al. [344] applied a binary descriptor called galaxy pattern [220] due to its robustness in presence of illumination changes and its efficiency for computation time compared to the state-of-the-art descriptors. Liu et al. [344] used background instances of galaxy patterns computed on non-overlapped regions from observed backgrounds to model the background. Experimental results [344] show that galaxy patterns allow the model to more robust to illumination changes and dynamic backgrounds than MOG [512], Codebook [288] and ViBe [32]. In a further work, Yang et al. [620] improved this approach with a fine level detection method to identify the label of each pixel.

6.6 Discussion on Texture Features

For local binary patterns, BGLBP [111]and XCS-LBP [490] appear to be the best LBP variants for background modeling and foreground detection while ST-SILTP and CS-SILTP seem to be the best LTP variants. SILS [638], LSS [264], LSBP [43] and LRP [641] present a robust alternative to LBP. Furthermore, patterns can be extended to spatio-temporal patterns, statistical patterns and fuzzy textures patterns to provide more robustness in presence of dynamic backgrounds and illumination changes. Moreover, designed texture features are proposed as textons [504] to deal with specific challenges met in aquatic environments.

7 Stereo Features

Stereo features encapsulate *spectral information* in the depth domain, and are obtained through a specific sensor which can provided a quantitative estimate of disparity [237] or depth [202]. Traditional stereo systems can provide both color and depth information [237][202]. However, it requires the setup of two cameras. Practically, stereo depth computation relies on finding small area correspondences between image pairs, and therefore does not produce reliable results in regions with little visual texture and in regions, often near depth discontinuities in the scene, that are visible in one image but not the other. Most stereo depth implementations attempt to deal with such cases of limitations and label them with one or more special values. Furthermore, noise and subtle lighting changes can cause the depth measurement at a pixel

to erroneous, even if nothing actually changes in the scene. In this context, shadows often provide the texture needed to extract valid depth in regions where measurements are usually invalid.

On the other hand, Time of Flight Cameras (ToF) which is a range imaging camera system obtains distance based on the known speed of light and by measuring the time-of-flight of a light signal between the camera and the object for each point of the image. Apart from their advantages of high frame rates and ability to capture the scene all at once, ToF based cameras have generally the disadvantage of low resolution. Several other limitations of ToF cameras are detailed in [317]. Nevertheless, the 2D/3D cameras provide additional depth features when compared to ordinary video, which makes it possible to deal with color camouflage issues during background modeling [494][308][307].

Recently, low cost RGB-D cameras such as the Microsoft's Kinect or the Asus's Xtion Pro are completely transformed several scientific visual signal processing applications. For example, Camplani et al.[66][65][62][63][64] used a Microsoft Kinect for objects detection through foreground/background segmentation. An other background subtraction algorithm based on RGB-D camera was developed by Fernandez-Sanchez et al. [148][149]. RGB-D cameras based on structured light scanner (i.e., Microsoft Kinect) are not suitable for outdoor environments, due to the range limitation and errors introduced by interference with the sunlight. Stereo features provided by RGB-D cameras presents other challenging problems [66][148]:

1. **Non-measured depth (NMD) pixels:** NMD pixels are mainly caused due to *1)* occlusions (typically around object boundaries), *2)* scattering of particular surfaces, *3)* concave surfaces, *4)* multiple reflections, *5)* out-of-range points (very distant points), and *6)* randomly, in homogeneous image regions.

2. **Noisy and irregular object boundaries:** Depth measurements at object boundaries are also heavily affected by noise. Sharp depth transitions produce misleading reflection patterns that result in rough and inaccurate depth measurements that are far from being correctly aligned to the actual object boundaries.

3. **Time dependent measurement noise:** Depth measurements are also affected by instability over time and space. On one side, measurements taken for a static object that correspond to the same image pixel vary with time. On the other side, different depth values are obtained for spatially neighboring pixels that correspond to points situated at the same distance from camera. The impact of this error varies with the distance.

4. **Distance dependent measurement noise:** The theoretical dispersion of depth measurements varies with the distance, following a quadratic law. This variation pattern has been confirmed in the study made by Camplani et al. [65].

5. **Camouflage in depth:** Similar depth between background and foreground lead to the problem of camouflage in depth.

Other limitations of the Microsoft Kinect based depth features are detailed in [173]. The stereo features allows dealing with camouflage in color and hence are generally used in conjunction with color features. However, the impact of using depth features independently from other features has been studied in [45][173]. In the following paragraphs, we review the different exiting approaches and the reader can see how the stereo features are fused with other features in Section 17.

7.1 Disparity

Ivanov et al. [237][238][239] was among the first authors who proposed a background subtraction method based on disparity verification. The use of disparity as a feature within background modeling allows coping to changes in illumination. The set-up for the extraction of the disparity map included three cameras called primary, left and right auxiliary cameras. A basic background disparity verification algorithm is applied on each two cameras views (primary, right) and (primary, left) for each pixel in the primary image:

– Using the disparity map find the current pixel in the auxiliary image, which corresponds to the primary current pixel.
– If the two pixels have the same color, label the current pixel as background.
– If the pixels have different colors, then the current pixel either belongs to the foreground class, or corresponds to occlusion/shadow (a region of the primary image which is not seen in the auxiliary camera view due to the presence of the actual object).

By cross-verifying each pixel across three camera views, Ivanov et al. [237][238][239] can distinguish the foreground object from occlusion/shadows. Practically, this method required the off-line construction of disparity fields mapping the background images that contained no foreground objects. At runtime, foreground detection is made by checking background image to each of the additional auxiliary color intensity values at corresponding pixels. When more than two cameras are available, more robust foreground detection could be achieved as in the case of occlusion/shadows. Because this method only assumed fixed background geometry, illumination variation at runtime can be handled. Since no disparity search was performed, the algorithm could be implemented in real-time on conventional hardware. Experimental results [237][238][239] show that this method extracted robustly silhouettes even under illumination changes in indoor scenes.

Eveland et al. [144] proposed the use of a disparity feature which was coded from 0 to 63, with higher values being brighter and closer to the camera, respectively. A single Gaussian [589] background model is applied on the disparity feature to achieve reliable foreground detection in indoor video scenes. Results have been applied to a real-time tracking system.

Gordon et al. [167] modeled the background using a multidimensional mixture

of Gaussians model with the (R,G,B,D) features. A significant advantage of incorporating both color and depth features within the background model is that, Gordon et al. [167] could correctly estimate depth and color of the background when the background is available in a fewer number of initialization frames. For pixels which possessed significantly invalid range of values in depth, Gordon et al. [167] relied on using only the color features. Finally, Gordon et al. [167] used a disjunction of the results coming from each feature to obtain the final foreground detection. A pixel classified as foreground based on either color or depth is taken to be foreground in the final foreground detection.

7.2 Depth

Depth-based detections result in compact silhouettes, are not often affected by illumination changes or shadows but is likely to show imprecise and noisy contours and, unclassified pixels due to NMD pixels.

7.2.1 Depth from Stereo Systems

Harville et al. [202][201][200] modeled the background using a mixture of Gaussians with the (Y,U,V, D) features. Color and depth features are considered independent and the same updating strategy of the original MOG [512] is used to update the distribution parameters. The matching strategy of the original MOG [512] is adapted to deal with combination of color (Y,U,V) and depth (D) features. At low luminance, the chroma components U and V become unstable and hence chroma is disregarded when comparing current observations. Similarly, when the depth becomes invalid, it is ignored. One the contrary, if in case a reliable distribution match is found for the depth component, the color-based matching criterion is relaxed thus reducing the color camouflage error. Similarly, in case that the stereo matching algorithm becomes unreliable, the color-based matching criterion is set to be harder to avoid problems such as shadows or local illumination changes. As in Gordon et al. [201], a pixel that is detected as a foreground based on either color or depth is classified as foreground in the final foreground detection.

For more reliability when combining color with depth, Song et al. [499], instead of appending depth information into a color vector, designed two probabilistic background models corresponding to color and depth based on MOG [512]. Then, the combined probability and hence the foreground detection, is marginalized as the product of each individual probabilities. In addition, Song et al. [499] also incorporate a scheme that handles noise in depth images to improve the accuracy and robustness of foreground detection.

7.2.2 Depth from ToF Cameras

Silvestre [494] proposed to used both the grey-scale (I) and the depth information provided by a ToF camera (SwissRanger). The background was modeled by the MOG

[512] applied on (I,D) features. Then, the distance between the samples has been considered in a two-dimensional grayscale-depth space. Thus, the depth and the grey-scale information are considered dependent. In the same way, Silvestre [494] adapted the KDE [142]. The probability to belong to the foreground is then given to each pixel according to the difference between the current and background pixel depth and brightness.

Langmann et al. [308][307] modeled the background using a mixture of Gaussians with the (Y,Cr,Cb, D, a) features where a is an amplitude modulation value. Thus, the 2D/3D camera produced a full size color image, low resolution depth and amplitude modulation images which are resized to match to color images by the nearest neighbor method. The matching function assumed that observations in the color, depth and amplitude modulation dimensions are in practice not independent. Indeed, a foreground object has most likely only a different depth, and also at least a slightly different color. Other reasons concern the limitations of ToF cameras as the infrared reflectance of an object has an influence on the depth measurement. Therefore, a linkage between the dimensions reduces the noise level in the foreground mask, the amount of misclassification due to shadows and block artifacts which occur when only depth measurements are inappropriate.

Stormer et al. [514] used also a MoG model [512], where depth and infrared features are combined to detect foreground objects in case of close or overlapping objects. Two independent background models are built. Each pixel is classified as background or foreground only if the two models matching conditions agree. But a failure of one of the models affects the final pixel classification.

Leens et al. [317] combined color and depth features, obtained with a low resolution ToF camera in a multi-camera system. The ViBe algorithm [32] is applied independently to the color and the depth features. Then, the obtained foreground masks are then combined with logical operations and then post processed with morphological operations.

Hu et al. [217] realized the foreground detection by using a weighted average on the probabilities obtained from the MOG model [512]. The different weights are updated adaptively for each output of the classifier by considering foreground detections in the previous frames and the depth feature. Experimental results [217] show that the proposed approach can effectively solve the limitations of color-based or depth-based detection.

7.2.3 Depth from RGB-D Cameras

Camplani et Salgado [66] proposed a combination of classifiers by jointly considering color and depth features. This combination is based on a weighted average that allows to adaptively modifying the support of each classifier in the ensemble by using the foreground detections in the previous frames and the depth and color edges. Thus, Camplani et al. [66] developed a weights selection scheme. For all those pixels for

which the depth measurements is not available, the depth-based classifier weight is set to 0, and the color-based classifier weight is set to 1. So, when depth data is not available in the current frame or in the background model, only the color feature is considered for the final pixel classification. For the pixels that do not belong to the nmd set, the weights are assigned following a function of the depth-image edges as depth data guarantee compact detection of moving object regions except that for the very noisy depth values at object boundaries. To reduce this effect, the influence of the color based classifiers is increased in these regions. For all those pixels that have valid depth data, the weights are assigned following a function of the depth-image edges to limit the effect of noisy depth values at object boundaries by using the color information in these zones. On the contrary, the depth information is more reliable in the regions far from depth-edges, since it guarantees compact foreground and it is resilient to shadows and illumination changes. This method reduced false detections due to challenges such as noise in depth measurements, moved background objects, color and depth camouflage, illumination changes and shadows.

Camplani et al. [65] proposed a foreground detection that combines depth and color information and reduces the noise present in the depth maps to improve their accuracy. Indeed, this method reduced the distance-dependent spatial noise and the nmd pixel effect while accurately preserving object depth boundaries, thanks to an adaptive filtering strategy. Temporal fluctuations are reduced by iteratively building a reliable color/depth model of the static elements in the scene. The parameters of the proposed filtering process and the temporal model are continuously adapted to the distance-dependent noise. The background model is based on two independent MoG models in color and depth, respectively. Binary image operations are used to combine foreground detection results in depth and color to obtain a binary mask which preserves depth-based compactness and color-based accuracy in the final detection. A real-time implementation on GPU architecture was developed by Camplani et al. [64].

Camplani et al. [62] used a multiple region-based classifiers in a mixture of experts fashion to improve the final foreground detection. It is based on multiple background models that provide a description at region and pixel level by considering the color and depth features. As Camplani et Salgado [66], the combination of the four models (pixel-color, region-color, pixel-depth, region-depth) is based on a weighted average to efficiently adapt the contribution of each classifier to the final classification.

Camplani et al. [63] used also a weighted average scheme in a Bayesian framework. Thus, probabilities from color and depth features are combined. The weights are chosen as a function of the each input to increase the support of the most reliable classifier as in [66].

Camplani et al. [45] proposed a combination of two algorithms to obtain a high-quality foreground detection using only depth feature acquired by the first generation of Microsoft Kinect. The first algorithm is the original MOG algorithm [512] adapted

for depth feature. The second algorithm is based on a Bayesian network, which explicitly exploits the spatial characteristic of the depth feature. This Bayesian network is able to accurately predict the foreground/background regions between consecutive frame using two dynamic models, which encode the spatial and depth evolution of the foreground/background regions regions.

Fernandez-Sanchez et al. [148] used the Codebook model [288], which has been extended to integrate depth features. This method is called Depth-Extended Codebook (DECB). To combine depth and color information, depth cues are used to bias the foreground detection based on color. The inclusion of depth information is made in two different ways: *1)* the first one considers depth as the fourth channel of the codebook, which has an independent mechanism from color and brightness, and *2)* the second one biases the distance in chromaticity associated to a pixel according to the depth distance. The first approach called 4D-DECB improved the robustness of the color-based algorithm to sudden illumination changes, highlighted regions and shadows. As depth features are more robust to lighting artifacts and shadows, dependence between RGB and depth have been used in the second approach which can be interpreted in the following way: if an input pixel is considered to be foreground, but it is close enough to the threshold, the classification will take into account the knowledge about the depth value for that pixel. This modification produced less false positive detections than 4D-DECB without biasing the color threshold. In the same idea, Fernandez-Sanchez et al. [149] improved this approach called DECB-LF (DECB-Late Fusion) by refining the foreground mask obtained from DECB (with biasing the color threshold) using the output of the color-based algorithm. Since depth images tend to have more noise than color ones, Fernandez-Sanchez et al. [149] evaluated a fusion method that reduces the impact of that noise in the resultant segmentation without using erosion or small region suppression.

Gallego and Pardas [153] combined color and depth features to perform a more complete Bayesian segmentation between foreground and background classes. For the background, the model consists of two independent Gaussians per pixel, one in the RGB domain and the second one in the depth domain. For the foreground, two parametric region-based foreground models combined color, space and depth domains, called Spatial Color Gaussian Mixture Model (SCGMM) and the Spatial Depth Gaussian Mixture Model (SDGMM), respectively. Then, the method combined the spatial-color and spatial-depth region-based models for the foreground as well as color and depth pixel-wise models for the background in a Logarithmic Opinion Pool fusion framework used to correctly combine the likelihoods of each model. A posterior enhancement step based on a trimap analysis is also proposed in order to correct the precision errors that the depth sensor introduces.

Greff et al. [173] provided a comparison of background models with only the depth feature. These models are the following ones: First frame without foreground objects, single Gaussian [589] and Codebook model [288]. The best performance is obtained by the Codebook model which eliminates the errors of uncertain and alter-

nating regions without missing the true foreground.

Ottonelli et al. [410][411] used a logical operations to fuse the results coming from the color and depth, respectively. For each features, the background model is the original MOG [512]. A depth-based compensation factor is computed using a logical AND applied on the depth mask, and the difference mask between the depth mask and the RGB mask. Then, the final mask is obtained by using a logical OR between the RGB mask and the depth-based compensation factor.

To make the background and foreground models more robust to effects such as camouflage and illumination changes, Spampinato et al. [503] explicitly models the scene's background and foreground with a Kernel Density Estimation approach in a quantized x-y-hue-saturation-depth space after a preprocessing stage for aligning color and depth data and for filtering/filling noisy depth measurements. Experimental results in three different indoor environments, with different lighting conditions, showed that this approach achieves an accuracy in foreground segmentation over 90% that the combination of depth data and illumination-independent color space proved to be very robust against noise and illumination changes.

Song et al. [500] used the original MOG [512] with color and depth information. For combining color and depth information, Song et al. [500] did not add depth to a color vector but designed two probabilistic background models corresponding to color and depth based on MOG and denoised depth image. Thus, they address solving the color camouflage problem and depth denoising. Experimental datasets made on their own dataset [500] show robustness in case of camouflage in color and depth.

Liang et al. [331] developed a refinement framework based on LUV color space and depth. As the foreground detection may be very inaccurate in some cases such as shadowing and color camouflage, Liang et al. [331] refined the inaccurate results by a supervised learning way. Thus, features are re-extracted from the source when it is detected. Since the depth data is not accuracy especially at the edge region, the edge of color map is used to detect the pixel wrongly classified as foreground by depth data. The re-extrated features with the initial detection results are fed to classifiers to obtain a better foreground detection. Experimental results on the RGB-D Object Detection Dataset [66] show that the refinement method outperforms the following color-depth methods: Camplani and Salgado [66], Camplani et al. [62] and Nguyen et al. [401][400] in presence of color camouflage and shadowing. The code is available at Github[4].

Amamra et al. [9] designed on the GPU a GMMM based background subtraction algorithm for joint RGB and depth. This parallel algorithm benefits from asynchronous data exchange between the host (CPU) and the device (GPU). In addition, the data structures are organized in the GPU to permit a higher memory coalescing.

---

[4]https://github.com/leonzfa/RBGS

Practically, this algorithm works at 30 fps.

Liu et al. [343] used only the depth feature to efficiently remove the background from depth data. According to the distance from the camera, the background is classified into far-field background and near-field background. If the distance of the far-field background is beyond the sensing range of the camera, the depth value cannot be obtained. In the depth map provided by the Kinect device, the depth value of the far-field background is set to zero. Then, Liu et al. [343] developed the concept of a probability map used to describe the probability of each pixel position belonging to the far-field background. For each position, the number of times that the depth value is equal to zero is counted. Then, the percentage of zero values is used as the probability. If the probability is large, the location most likely belongs to the far-field background. Note that the non-zero depth values indicate presence of foreground objects. Since the near-field background is within the sensing range of the camera, the depth value of the near-field background is also non-zero. However, the distance between the camera and the near-field background is longer than the distance between the camera and the foreground. Thus, the depth value in any position of the near-field background is larger than those of the foreground objects. Based on this observation, Liu et al. [343] designed a maximum depth map, which is used to record the maximum depth value of each position in a depth sequence. Finally, the far-field and near-field background are combined to obtain one background model used for foreground detection. Experiments [343] show that this method outperforms both the average model and ViBe [32].

7.3 Discussion on Stereo Features

Stereo features can not be used alone and need to be carefully used following their properties as developed in Nghiem and Bremond [398]. There is no study about the influence on how the depth is acquired (Stereo cameras, ToF Cameras, RGB-D Cameras) and the robustness in the foreground detection.

8 Motion Features

The motion features provide ***temporal information*** and are usually obtained via optical flow to deal with irrelevant motion in the background. Theoretically, the most robust optical flow strategy should be employed to obtain the best performance. But most of the optical flow algorithms are computationally slow. Three alternative approaches are then used to introduce temporal attributes: **1)** the ones based only on the difference between consecutive frames. Then, the background model is only computed in stationary regions of the scene, **2)** optical flow (computed on all the pixels) which is used to detect moving areas. As in the previous approaches, the background model is only computed in stationary areas, and **3)** optical flow is only computed on

moving areas after foreground detection. In this case, optical flow allows the algorithm to distinguish the unimportant moving areas from the moving objects. Some fast optical flow algorithms such as the fast nearest neighbor field based optical flow algorithms EPPM [31] run in near real-time on state-of-the-art GPU. We review in the following paragraphs the different exiting approaches and the reader can see how the motion features are fused with other features in Section 17.

Tang et al. [532] used the motion speed value in addition to the intensity in the MOG model [512]. The motion speed value at each pixel is obtained through saliency motion filtering. First, a motion map is obtained by consecutive frame difference of Gaussian from each frame, from which a number of feature points are extracted with Monte Carlo importance sampling. Their corresponding velocities are computed using an optical flow algorithm. The consecutive frame difference allows to detect slow moving objects, and speed up the algorithm by only applying the optical flow to the regions of change which are detected by consecutive frame differenc. In that region, for each pixel, the motion is considered as salient motion if the pixel and its neighborhood move in the same direction in a period of time.

Huang et al. [229][227][226][225] used dense optical flow for describing motion vectors. Regions with coherent motion are then extracted as initial motion markers. Pixels not assigned to any region are labelled uncertain ones. Finally, a watershed algorithm based on motion and color is utilized to associate uncertain pixels to the nearest similar mark. Further, MRFs are used to formulate foreground detection as a labelling problem. The optimization over the MRF model is then performed. The posterior probabilities initialized with the ones computed with the MOG model [512] are maximized to obtain the final classification result. Finally, regions which have the same classification label and similar colors are merged to derive a more consistent foreground mask. Experimental results [227] on gradual illumination changes and shadows demonstrate the robustness of this method, but the computational complexity of the technique has not been mentioned. In similar studies, Huang et al. [230][228][231] used motion information captured through the difference of consecutive frames to model the background in stationary areas.

Zhou and Zhang in [690] used Lucas-Kanade gradient-based method for computing optical flow. As this method can only be applied for small displacements and as the displacements of moving objects between consecutive frames is anticipated to be more than 15 pixels, Zhou and Zhang [690] used a hierarchical coarse-to-fine warping technique based on a Gaussian pyramid decomposition. The original MOG [512] with the color features is used to model the background. Then, in the fusion step, Zhou and Zhang [690] only considered those foreground objects as moving objects, where the amplitude and direction of the optical flow are within the ranges of consideration.

Using EPPM [31], Chen et al. [78] ensured temporally-consistent background subtraction with optical flow estimation by tracking the foreground pixels. Here, motion information is integrated with a temporal M-smoother. A similarity measure-

ment is obtained directly from optical flow estimation with the assumption that the background estimate for the same object appearing in the difference video frames should be identical. As the direct implementation of EPPM [31] is extremely slow as optical flow estimation is required between any two video frames, Chen et al. [690] developed a recursive implementation so that optical flow estimation is required only between every two successive frames. As described in previous approaches, the background model is initially obtained using the MOG model [512]. Then, a spatial and a temporal M-smoother are employed to obtain a spatially-temporally-consistent foreground mask. Experimental results [78] on the ChangeDetection.net dataset [169] and SABS dataset [61] show that this algorithm outperforms most of state- of-the-art algorithms.

In an other approach, Dou and Li [133] proposed a moving object detection method based on SIFT flow [341]. SIFT flow addressed the problem of image registration by aligning a query image to a target image at the scene level by spatially warping the query image to match the target image. This alignment is achieved by using dense pixel-wise SIFT descriptors. One aim of this algorithm is to counteract object motion between two scenes. This allows SIFT flow to be well suited to correcting object plane motion. Thus, Dou and Li [133] proposed to tp model each pixel with a pixel-wise SIFT descriptor and facilitate dynamically updating the model. The background model is a K histogram of the SIFT flow. Experimental results on the I2R dataset [326] demonstrate that the proposed approach provides an effective and efficient way for background modeling and foreground detection.

Using multiple features, Zhong et al. [682] proposed to fuse texture (LBP[207]) and motion patterns. A temporal operator to obtain the motion pattern is formulated as follows:

$$T_{motion}(x,y) = \sum_{i=0}^{7} b_{t_i}(x,y)2^i \qquad (8)$$

where the function $b_{t_i}(x,y)$ keeps the sign of the difference between the central pixel at location (x,y) attime t and its ith neighboring pixel in previous $(t-1)^{th}$ frame as follows:

$$b_{t_i}(x,y) = 1 \ \text{ if } \ I_{t-1}(x_i,y_u i) > I_t(x,y) b_{t_i}(x,y) = 0 \ \text{ otherwise} \qquad (9)$$

where $I_t(x,y)$ is the intensity value at pixel location (x,y) at time t. For each pixel, its probability to be either a background or foreground is computed from the histogram of each feature. Then, the results are combined using a weighted average mechanism. Experimental results [682] show that the combination of LBP and motion pattern outperforms the original LBP [207] in presence of dynamic backgrounds.

Martins et al. [362] used a bio-inspired feature called Magno channel in addition with color features. Thus, this approach merges information from two inherently different methods: **(1)** bio-inspired motion detection method using the Magno channel, and **(2)** a background subtraction algorithm based on pixel color information. Thus,

the background subtraction can be any one existing algorithms such as MOG [512], Flux Tensor with Split Gaussian model (FTSG) [571], KDE [142], AMBER [561] and SuBSENSE [508]. The foreground detection is obtained by merging the detection of the two methods. Experimental results [362] on the CD.net 2014 dataset show that the hybrid approach always substantially improves the performance of the original background subtraction methods.

9 Local Histogram Features

9.1 Local Histogram of Color

Local histograms on intensity or color provides color information about the neighborhood of the pixel that the one color value do not give. In addition to their invariance to image rotation and translation, histograms are easy to compute. However, without data quantization to reduce size, histograms need more time to compute. Furthermore, histograms are sensitive to quantization errors and require a more complex similarity measure for the comparison. The different local histograms of color developed in the literature can be classified as crisp and fuzzy local histograms.

9.1.1 Crisp Local Color Histograms

– **Local Color/Edge Histograms:** First, Mason and Duric [363] used local color histograms in the RGB color space with local edge features. First, a depth reduction formula is applied to transform 24-bit color to 12-bit color to reduce computation time, and to reduce the complexity because 24-bit histograms are more hard to compare since 1-bit change color value places the corresponding pixel into a different histogram bin. Furthermore, the lower four bits obtained by low-cost cameras that capture 24-bit video are often very noisy, and 12-bit pixel representations only require 4096 bins for each histogram. Thus, Mason and Duric [363] obtained smaller histograms which are much easier to build and compare. Then, edge histograms are composed of 36 bins. For each edge pixel, the bin index is computed using the edge orientation (10^0 per bin). When the bin index is dertermined, the bin is incremented with the dege magnitude. Finally, the histograms in color are compared by using an intersection measure while the histograms in edge are computed with a chi-squared measure. Experimental results [363] show that the Local Gradient Histogram (LGH) with edge allow the method a better outline of the objects than the Local Color Histograms (LCH).

– **Local Kernel Color Histograms:** Noriega et al. [406] proposed Local Kernel Color Histograms (LK-CH) to keep the advantages of histograms avoiding their drawbacks. First, each image is segmented into overlapped local squares with a histogram for each one to collect spatial information. To reduce the amount of data, the color space is quantized according to the most representative colors extracted from the scene. Second, Noriega et al. [406] used a double Gaussian

kernel, that are one in the image space and one in the color space, to be robust against noise. Then, local kernel histograms are computed from image overlapped regions using the two Gaussian kernels. Finally, a pixel scale probability map is obtained by using the Bhattacharyya distance. Experimental results [406] on the Wallflower dataset show that LK-CH is less affected by noise, camera vibrations and swaying trees than the mean, the median and the LCH [363]. But, local kernel histograms with color features cannot handle illumination changes.

- **Estimated Local Kernel Histograms:** Li et al. [318] developed a nonparametric Estimated Local Kernel Histogram (ELKH) for moving objects detection in presence of dynamic backgrounds. First, Local Kernel Histogram (LKH) are built by using the correlation and texture of spatially proximal pixels. Then, the probability distribution of LKH is estimated with a nonparametric technique. The Bhattacharyya distance is used to measure the similarity of LKH between the estimated background model and the current frame. Practically, Li et al. [318] used the Estimated LKH which is a block-wise version of the LKH, and in which there are only several nonzero bins, and the others are all zero. That means intensity of color is congregative in a pixel block. The EKLH contains some texture information, and it is more robust to disturbing noise than LKH. Experimental results [318] show that ELKH reduced false detections due to dynamic texture robustly, but also allow to detect the small moving objects.

- **Local Color Difference Histograms:** Li [324] used histograms in the YUV color space to obtain difference image of color distance. Histograms are obtained by statistics of the difference image. According to the mono-modal feature of histogram of the difference image, Ji [324] employed an adaptive clustering method, and removed noise with morphological filtering. Finally, an updating scheme is used to adapt the model to the illumination changes and environmental conditions. Experimental results [324] show that LCH with the YUV color space offer robustnees in presence of illumination changes.

- **Local Dependency Histograms:** Zhang et al. [660] [661] proposed the Local Dependency Histograms (LDH) to model the spatial dependencies between a pixel and its neighboring pixels for dynamic background subtraction. LDH is computed using the direct neighbors and the indirect neighbors. The direct dependencies can be along any directions. The indirect spatial dependencies are confined to be only along the horizontal or vertical direction. Based on LDH, Zhang et al. [660] [661] developed a dynamic background subtraction in which each pixel is modeled as a group of weighted LDHs. Foreground detection is obtained by comparing the LDH computed in current frame against its model LDHs with a histogram intersection measure and an adaptive thresholding method. Then, The model LDHs are adaptively updated by the new LDH. Experimental results [660] [661] on the I2R dataset [326] and CMU dataset [480] show that LDH is more robust in presence of dynamic backgrounds and camera jitter than the original MOG [512].

– **Spatiotemporal Condition Information:** First, Wang et al. [562] built a spatiotemporal neighborhood based on the center surround visual saliency model. To reduce the false detection in the foreground mask, a neighborhood weighted spatiotemporal condition information (NWSCI) is used to classify pixel based on the similarity of neighborhood pixels. Second, a joint cascade and hierarchical framework reduced computational cost by rejecting the unchanged regions before foreground detection. Experimental results [562] show that NWSCI based background subtraction effectively outperforms the original MOG [512], the original KDE [142] and ViBe [32], and gives similar performance than [297] and [662].

9.1.2 Fuzzy Local Color Histograms

– **Local Fuzzy Color Histograms:** Kim and Kim [290] adopted a clustering-based feature called Fuzzy Color Histogram (FCH) (FCH) to attenuate color variations due to background motions and still highlight moving objects. FCH [198] is a fuzzy version of the conventional color histogram. In CCH, the quantized color feature is assumed to be into exactly one color bin and it often lead to abrupt changes even though color variations are actually small. On the other hand, FCH uses the fuzzy membership [135] to relax this crisp condition. Kim and Kim [290] computed FCH on the CIELab color space arguing that CIELab color correctly quantifies the perceptual color similarity. Then, the colors in the CIELab color space are classified into clusters using the FCM clustering technique [38]. The foreground detection is made by thresholding the difference between the local FCH in the background and the background. The similarity measure used is the normalized histogram intersection for simple computation. Then, the background model is updated an online procedure. Experimental results [290] on the I2R dataset [326] show that the Local FCH (LFCH) outperforms the original MOG [512], the generalized MoG (g-MoG) [110], STLBP [659] and the local CCH. In other works, Kanna and Murthy [272], and Gutti and Shankar [190] employed the local FCH too while Kumar et al. [299] used it in an automated surveillance system. In an other approach, Yang et al. [624] added to local FCH spatial coherence and temporal consistency in a Markov random field statistical (MRF) framework.

– **Local Fuzzy Color Difference Histograms:** First, Panda et al. [416] presented a background subtraction algorithm based on color difference histogram (CDH) by measuring the color difference between a pixel and its neighbors in a small local neighborhood. CDH reduced false detections due to the non-stationary background, illumination variations and camouflage. Secondly, the color difference is fuzzified with a Gaussian membership function. Finally, Panda et al. [416] proposed a Fuzzy Color Difference Histogram (FCDH) by using fuzzy c-means (FCM) clustering [38] and exploiting the CDH. FCM clustering algorithm applied to CDH allow to reduce the large dimensionality of the histogram bins in the computation and also decrease the effect of intensity variation due to the unimportant motion or illumination changes. Experimental results [416] on the I2R dataset [326] show that FCDH is more robust than the original MOG [512], the origi-

nal LBP, STLBP [659] LIBS [203], FBS [22], FST [89], MKFC [91], and LFCH [290].

9.2 Local Histogram of Gradient

– **Local Gradient Histograms:** Please see the item "Local Color/Edge Histograms" in Section 9.1.

– **Local Histogram of Oriented Gradient:** Fabian [145] presented a MOG based background subtraction which exploits Histograms of Oriented Gradients (HoG) instead of intensity to deal with camera jitter, automatic iris adjustment and exposure control. Thus, Fabian [145] avoided false detections of foreground mask. Practically, the brightness value is replaced by a local image gradient because image gradient and its orientation are invariant against changes in brightness. This scheme applied only to a certain extent given by limited range of pixel values. In practice, this assumption is quite suitable and allow the method to reduce the effects of automatic control. Furthermore, to deal with camera jitter, Fabian [145] handled small image movements. For this, the gradient is measured for squared area of size 8×8 called cell. Finally, an on-line spatial rearrangement of cells minimized the variance of dominant gradient for every cell. Practically, the most significant orientation bin for every cell in the image is extracted and the mean HOG (MHOG) is computed to provide the reference values for minimizing the variance of gradients and especially resulting bins. To measure the distance between two different bins, Fabian [145] defined a metric in a discrete metric space to obtain difference between two different bins, and such that a decision can be made when a same bin doesnt exist. Then, the HOG feature ares used in the MOG model. Experimental results [145] on traffic surveillance videos show the pertinence of the HOG against the intensity. In an other approach, Mukherjee et al. [386] used both RGB and HOG in the MOG model. First, a modified distance based on support weight is developed to compare RGB features, and a HOG distance is presented to compare HOG features in order to provide distinct cluster values. Second, a multi-layers model is employed to obtain the foreground mask. Experimental results [386] on five datasets show that this method is more robust in presence of illumination changes and dynamic backgrounds than the original MOG [512], CRF-based MOG [580], self-adaptive MOG [82], the Type-2 Fuzzy MOG (T2-FMOG) [25], T2-FMOG with MRF [679] and SOBS [356]. Javed et al. [252]. In an other work, Panda et al. [415] combined HOG and LBP for complex dynamic scenes. Experimental results [415] on the I2R dataset show that LBP-HOG is more robust than LBP, STLBP [659], the Choquet integral with RGB features [22] MKFC [91] in presence of dynamic backgrounds.

– **Local Adaptive HOG:** Hu et al. [218] proposed to use HOG in a coarse-to-fine strategy. First, a HOG-based background model is constructed with a group of adaptive HOG. Then, the foreground detection is achieved by comparing HOG of each pixel in the current frame and the background model. To deal with some

missing foreground regions, the foreground mask is is improved by using the pixel-wise detection provided by MOG algorithm and morphological operations. In the refinement step, the foreground maskis refined based on the distinction in color feature to eliminate errors such as noises and shadows. In the experimental results [218] on the Wallflower dataset [218], this method outperforms the original MOG [512] and the codebook model [287].

– **Local Orientation Histograms:** Orientation histograms were applied with success for visual tracking, and keep the probability density function of local gradients which is robust against illumination changes and easy to compute. In this context, Jang et al. [247] developed Local Orientation Histograms (LOH) with Gaussian kernel, thus obtaining 1D orientation histograms to reduce the quantization error. LOHs allow to compared the background and the foreground by dividing each frame into small local regions. Each local region has the foreground probability given by comparing LOH between background and foreground image. In a multi-scaled approach, Jang et al. [247] used an foreground detection algorithm that dynamically partition and compare regions with a recursive partitioning algorithm. Because it requires multiple extractions of histograms from multiple rectangular cells, Jang et al. [247] used the integral histogram [434] to provide fast extraction of histogram over the multiple overlapped cells. Experimental results [247] show that LOH suppress local false detection in presence of illumination changes.

– **Local Kernel Histograms of Oriented Gradient:** Noriega and Bernier [407] proposed Local Kernel Histograms of Oriented Gradient (LK-HOG). First, contour-based features are extracted for local kernel histograms. Experimental results [407] on the Wallflower dataset show that LK-HOG is more robust to illumination changes than the mean, the median, the LCH [363], LK-CH [406], LGH [363], and local HOG. In all the videos, Gaussian kernels improved the quantization error rate to reduce both false positives and negatives.

9.3 Local Histograms of Figure/Ground

Zhong et al. [687] represented each pixel as a local histogram of figure/ground segmentations, which combines several prospective solutions that are generated with simple background algorithms to get a more reliable and robust feature for background subtraction. The background model of each pixel is constructed as a group of weighted adaptive local histograms of figure/ground segmentations, which describe the structure properties of the surrounding region. The goal of using LH-FGs is to make up for deficiencies each individual algorithm, thus achieving a better overall performance than each single algorithm. First, a local histogram of figure/ground segmentations is computed over a squared fixed-size $N \times N$ neighborhood for each algorithm with integral histogram [434]. Then, LH-FGs of each algorithm are concatenated together to provide the final representation of each pixel. The feature ex-

traction step thus yields to S (2-bins) histograms where S denote the number of foreground detection maps. Finally, the S histograms are concatenated together to give a final $2S$-bins histogram, which is normalized to sum to one, so that it is also a probability distribution. The similarity between two histograms is computed with the Bhattacharya distance.

10 Local Histon Features

Histon and its associated measure Histon Roughness Index (HRI) were applied to still image segmentation with good performance. In this context, Chiranjeevi and Sengupta [90] introduced histon and rough set theory for foreground detection, extended the histon concept to a 3D histon, and incorporated fuzziness into the 3D HRI measure, and thus obtained 3D Fuzzy Histon. Then, the labeling decision is based on Bhattacharyya distance between the model HRI and the corresponding measure in the current frame. Practically, histon and these variants are defined as follows:

- **Histon:** Histon allows to visualize the color information for the evaluation of similar colored regions. For each intensity value in the histogram, the number of pixels which are in the similar color sphere is computed, and this value is added to the histogram value, to obtain the histon value of the corresponding intensity. Histogram and histon distributions give color information and spatial information. Identical distributions of both the histogram and the histon in a region indicate that the region present lacks in its spatial homogeneity or its spatial similarity. So, Chiranjeevi and Sengupta [90] used this property to model the pixel in the center of the region which is formulated by Histon Roughness Index (HRI), after correlating the histon concepts with the rough set theory. Finally, HRI incorporates both the color and the spatial information, and it is used for comparison between the background model and the current frame.

- **3D Histon:** In the original formulation of histon, each color channel is considered separately, rather independently, to obtain the histon for each color channel. But, exploiting the three color component values of each pixel with a 3D spatial distribution is better to get more information than exploiting the spatial distribution of the three independent color planes. So, Chiranjeevi and Sengupta [90] proposed an integrated 3D histon, where the histon distribution is computed by using the color value on three channels jointly. Then, the 3D HRI distribution for a region, centered at a pixel, is calculated using the 3D histon and the 3D histogram. Experimental results [90] show the effectiveness of 3D histon compared to the basic histon.

- **3D Fuzzy Histon:** In 3D histon, whether a pixel is similar to its neighbors or not, is determined crisply. By determining the extent of similarity using Gaussian membership function, Chiranjeevi and Sengupta [90] proposed the 3D fuzzy histon as an extension of the basic fuzzy histon. 3D Fuzzy histon is subsequently

used to compute 3D Fuzzy Histon Roughness Index (3D FHRI). Experimental results [90] on the I2R dataset [326] show the effectiveness of 3D Fuzzy histon compared to the 3D histon. In a further work, Giveki et al.[164] employed Atanassovs Intuitionistic Fuzzy Sets (A-IFS) theory to the concept of histon or constructing Atanassovs Intuitionistic 3D Fuzzy Histon Roughness Index.

11 Local Correlogram Features

Correlogram is an image of correlation statistics, and it can also be used as feature as follows:

- **Correlogram:** Correlogram captures inter-pixel relationships in a block or a region, and alleviates the drawbacks of histogram, which only considers the pixel intensities for calculating the distribution [676]. So, it is suitable for modeling dynamic backgrounds as developed in Chiranjeevi and Sengupta [87]. But, computation of correlogram for RGB color components involves huge computations as the correlogram size is $256^3 \times 256^3$. Even with single color channel, it requires significant computations as the correlogram size then becomes 256×256. Hence, the single channel is quantized to a finite number of levels l. Due to this, the correlogram size is further reduced $l \times l$ with $l \ll l$.

- **Fuzzy correlogram:** Crisp assignment of quantized intensity pair to a particular correlogram bin is sensitive to quantization noise [87]. To reduce computational time and to address the crisp assignment limitations, Chiranjeevi and Sengupta [87] developed fuzzy correlogram, composed by applying fuzzy c-means (FCM) algorithm [38] on correlogram. Thus, each intensity pair is related to all the bins by their respective membership values. The membership matrix is obtained by using fuzzy c-mean algorithm where the Euclidean distance is used as a measure between the cluster centers and the data points. Fuzzy correlogram greatly reduces the number of correlogram bins and the quantization noise. Experimental results [87] on the I2R dataset [326] show that fuzzy correlogram can handle multi-modal distributions better that approaches which used multiple model features.

- **Multi-channel fuzzy correlogram:** In Chiranjeevi and Sengupta [87], correlogram was obtained from a single color channel, and thus the color information is loosed. The dependency across the color components was also ignored. Because there is more information in the color distribution than a single monochrome distribution, Chiranjeevi and Sengupta [91] proposed a inter-channel correlogram, which captures both the color information and the dependency across the color planes. However, inter-channel correlograms exclude the inter-pixel relations within the same color plane. To use this information too, Chiranjeevi and Sengupta [91] combined both inter-channel and intra-channel correlograms, called multi-channel correlogram which exploits both the color information by taking

into account the color dependencies and the inter-pixel relations across and within the color planes, unlike the correlogram proposed in Chiranjeevi and Sengupta [87] that captures only the inter-pixel relations on a single color plane. Then, the correlograms are mapped to a space of reduced dimensionality by using a transformation based on a fuzzy membership matrix, whose elements indicate the belongingness of each intensity pair to the new bins. Instead of fuzzy correlogram which used FCM to computed membership values with the disavandages that Euclidian distance is only suitable for spherical clusters and that FCM is sensitive to the outliers, Chiranjeevi and Sengupta [91] obtained the membership values by using Kernel Fuzzy C-Mean (KFCM) algorithm [651]. The membership values applied over multi-channel correlogram results in a feature called Multi-channel Kernel Fuzzy Correlogram (MKFC). KFCM uses a kernel which is based on a metric distance in place of Euclidean distance, which made it more robust than FCM. Furthermore, MKFC involves less number of bins ($k \ll l^2$) than the original correlogram of dimensionality l^2 and this significantly reduces the computational complexity in distance computation for foreground detection.

11.1 Discussion on Histogram/Correlogram Features

Local histograms/correlograms on color or edge gives color or edge information about the neighborhood of the pixel that the one color or edge value do not give. Thus, these features allow the background model to be more robust in presence of dynamic backgrounds and illumination changes, and their main drawback is the number of bins. Furthermore, several approaches were developed to address this problem. Furthermore, fuzzy Histogram/Correlograms allow to avoid crisp assignments in the clustering step. Practically, LFCH [290] and FCDH [416] provide good performances in the case of histograms while MKFC [91] appears as the best descriptor in the case of correlograms.

12 Haar-like Features

Haar-like features [557] have been used in different approaches in background modeling and foreground detection as follows:

- **Haar-like Features in hierarchical approaches:** Zhao et al. [675] used Haar-like features in a hierarchical codebook approach. In the block-based level, four Haar-like features and a block average value are used to represent a block. These features are computed using integral image to reduce computation time and are less sensitive to dynamic backgrounds. Thus, most of the background blocks are removed at this step without reducing the true positive rate. n the block that has been detected as foreground, a pixel-based codebook is applied to increase the precision. Experiment results show [675] on the I2R dataset show that the Haar-like feature allow to be more robust in presence of dynamic backgrounds.

– **Haar-like Features in multi-features approaches:** In an other approach, Klare [293] and Klare and Sarkar [294] used the Haar-like features in an extended feature set. In a similar way, Han and Davis [195] employed Haar-like features in a multi-feature approach with density based background subtraction. In an other way, Lopez-Rubio and Lopez-Rubio [349] employed Haar-like features in a method based on stochastic optimization [348]. Please see Section 17 for details.

– **Haar-like Features in feature selection approaches:** Grabner and Bischof [170], Grabner et al. [171] and Lee et al. [316] used Haar-like features in an online boosting algorithms which select the best combination of features. Please see Section 18 for details.

13 Haralick Features

Subudhi et al. [516] used Haralick features [199].

14 Location Features

Pixel-wise models ignore the dependencies between proximal pixels and it is asserted here that these dependencies are important. Thus, several authors proposed to use the location (x, y) of the pixel to add to the ***spatial information***. This location information is used directly in a pixel manner [481][479] or indirectly via invariant moments in a region manner [361]:

– **Location (x,y):** First, Sheikh and Shah [481][479] used the location (x,y) in addition to the color features (Normalized RGB) in a joint representation called "domain-range representation" that provides a direct means to model and exploit the dependency between the pixels. Thus, a kernel density estimation (KDE) model represents the background and foreground processes by combining the three color dimensions and two spatial dimensions into a five-dimensional joint space. But, this method was found to be dependent on the size of the image. Indeed, the classification criterion, based on the ratio of likelihoods in this five-dimensional space, has an undesirable dependence on the size of the image. Similar to the work of Sheikh and Shah [481][479], Narayana et al. [394] modeled the foreground and background likelihoods with a KDE using pixel samples from previous video frames but by building an explicit model of the prior probability of the background and the foreground at each pixel. Moreover, Narayana et al. [394] modeled the processes using a three-dimensional color distribution at each pixel. In addition, Narayana et al. [394] incorporated spatial priors for the background and foreground processes by using classification labels from the previous frame. Large spatial covariance allows neighboring pixels to contribute more to the score at a given pixel location. Color covariance allows for some color appearance changes at a given pixel location. In this framework, the distributions

are conditioned on spatial location, rather than being joint distributions over location and color. So, the model of Narayana et al. [394] avoids the dependence on the image size and yields better results. Furthermore, Narayana et al. [394] used Lab and SILTP [332] as color and spatial features respectively. Furthermore, instead of an uniform kernel [481][479], an adaptive kernel scheme is used and is nearly as accurate as the full procedure but runs much faster. Narayana et al. [395][396] improved this scheme by using a separation of the foreground process into "previously seen" and "previously unseen" foreground processes, and by using explicit spatial priors for the three processes - background, previously seen foreground, and previously unseen foreground. The probabilistic formulation with likelihoods and a spatially dependent prior for each process leads to a posterior distribution over the processes. Instead of a constant kernel bandwidth as in Sheikh and Shah [481][479], Antic and Crnojevic [13] adapted the kernel bandwidth according to the local image structure. Thus, image gradient is used to adaptively change the orientation and dimensions of the kernel at the borders of the region. This adaptive scheme provides more accurate modeling of non-stationary background containing regions with different texture and illumination and suppress structural artifacts present in detection results when the kernel density estimation with constant bandwidth is used. In an other work, Zhou et al. [691] used the normalized RGB and the intensity with the location features to be robust against illumination changes and shadows.

– **Moments:** Marie et al. [361] used invariant moments based on the Hu set moments [219]). Each pixel is modeled as a set of moments computed from its neighborhood and stored using a codebook model [287]. Hu [219] defined 7 invariants, but their complexity increases dramatically. Moreover, the results obtained using the three first are very similar than the one with all the moments. Practically, Marie et al. [361] retained only the moment named "I1", which is computationally the fastest. Experimental results [361] on the Sheikh sequence [480] demonstrate that the codebook with moment outperforms the original codebook [287], the original MOG [512], and approximated median [364].

15 Tranform Domain Features

15.1 Features from Frequency Domain Transform

Transform domain features obtained through frequency analysis can provide valuable *spectral and spatial information* that can in-turn lead to better background modeling and foreground detection. Some of the frequency domain features include:

– **Fourier Transform Features:** Fourier transform features encapsulate *spectral information* which are suitable for scenes that contain periodic patterns. That is, scene where is significant correlation between structures in the scene and the observations across time. For example, a tree swaying in the wind or a wave lapping on a beach is not just a collection of randomly shuffled appearances, but a physical

system that has characteristic frequency responses associated with its dynamics. In this context, Wren and Porikli [591] estimated the background model that captures spectral signatures of multi-modal backgrounds using Fast Fourier Transform (FFT) features through a method called Waviz. Here, FFT features are then used to detect changes in the scene that are inconsistent over time. Experimental results [591] show robustness to low-contrast foreground objects in dynamic scenes. In an other work, Tsai and Chiu [547] presented a background subtraction method using two-dimensional (2D) discrete Fourier transform (DFT). This 2D-DFT based method first converts input frames to gradient images, and then Tsai and Chiu [547] applied the 2D-DFT on each spatial-temporal slice of the gradient image sequence and removed the vertical line pattern of the static backgrounds. In this way, this 2D-DFT based method can detect foreground objects without a training phase. However, the foreground masks contain only boundaries of moving objects and also present ringing around object boundaries.

– **Discrete Cosine Transform (DCT) Features:** Porikli and Wren [437] developed an algorithm called Wave-Back that generated a representation of the background using the frequency decompositions of pixel history. The Discrete Cosine Transform (DCT) coefficients are used as features are computed for the background and the current images. Then, the coefficients of the current image are compared to the background coefficients to obtain a distance map for the image. Then, the distance maps are fused in the same temporal window of the DCT to improve the robustness against noise. Finally, the distance maps are thresholded to achieve foreground detection. This algorithm is efficient in the presence of waving trees. An other approach developed by Zhu et al. [693] used a set of DCT-based features to exploit spatial and temporal correlation using a single Gaussian model. Thus, Zhu et al. [693] used two features: the DC and the low frequency AC parameters. Each of which focuses on intensity and texture information, respectively. The AC feature parameter consists of the sum of the low frequency coefficients. The high frequency coefficients are not used because they are more sensitive to noise and they concern fine details that are more susceptible to small illumination changes. The use of these two features is equivalent to using both intensity and texture information, and produces more robust and reliable foreground detection masks. In an other approach, Wang et al. [574] used only the information from DCT coefficients at block level to construct background models at pixel level. Wang et al. [574] reports the implementation of running average, median and 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. In an other work, Amith et al. [11] proposed an algorithm based on the combination of DCT and PCA. The elementary frequency components are obtained from separable property of DCT and the dimensionality of these components is usually high. In order to reduce dimensionality and to extract effective features of the elementary frequencies, PCA approach is used. Experimental results [11] on the PETS dataset and other real time video sequences show that the relevance of the approach.

– **Wavelet Transform Features:** First, Huang and Hsieh [221][223] proposed the use of Discrete Wavelet Transform (DWT) to obtain features that are used in a change detection based method for interframe-difference but DWT is not suitable for video applications as the use of DWT makes the method shift sensitive [278]. In an other work, Gao et al. [158][159][160][161] proposed a Marr wavelet kernel and a background subtraction technique based on Binary Discrete Wavelet Transforms (BDWT). Thus, the background and current frames are transformed in the binary discrete wavelet domain, and background subtraction is performed in each sub-band. Experiments results [158] on several traffic video sequences show that this BDWT method produces better results with much lower computational complexity than the original MOG. To detect the moving objects more accurately, Hsia and Guo [215] proposed to use a modified directional lifting-based 9/7 discrete wavelet transform (MDLDWT), which is based on the coefficient of lifting-based 9/7 discrete wavelet transform (LDWT). Furthermore, to overcome that clear shape information of moving objects may not be available from multiple-level decomposition image such as LL3, Hsia and Guo [215] preserved the shape of objects in the low resolution image. Thus, the MDLDWT detects foreground moving objects in spatial domain. DLDWT not only retains the features of the flexibilities for multiresolution, but also achieves low computing cost when it is applied for LL-band images. MDLDWT based method presents the advantages of low critical path and fast computational speed. Moreover, the LL3-band is used solely to reduce the image transform computing cost and to remove noise. Experimental results [215] show that MDLDWT based method better retains slow motion of objects than DWT based method [221][223]. In an other approach, Gao et al. [156] used orthogonal non-separable wavelet transformation for background modeling, and extracted the approximate information to reconstruct information frames. If the background present gradual changes, weighted superposition of multi background modeling images with time is applied to update the background. If the background presents sudden changes, the background is remodeled from this frame. In an other approach, Guan et al. [175][177][176][179][178] used wavelet multi-scale transform for detecting foreground moving objects and suppressing shadows. Practically, 2-D dyadic WT coefficients are obtained. Experimental results [175] show that the 2-D dyadic WT approach gives less false detections than the original MOG. Although, these wavelet based methods shown promising results [453][445], they are not adaptive in nature and tested against simple scenarios. Indeed, it addressed moving objects detection in presence of static backgrounds, but not effectively in presence of dynamic background changes. Practically, these previous wavelet based methods present various problems such as ghost like appearance, object shadows, and noise. Also discrete real wavelet transform (DWT) presents shift-sensitivity [245]. To alleviate these weaknesses, Jalal and Singh [245][244][246] used Daubechies complex wavelet transform which is approximately shift-invariant and presents better directionality information compared with DWT. Thus, the noise resilience nature of wavelet domain is addressed, as the lower frequency sub-band of the wavelet transform presents the ability of a low-pass filter. So, Jalal and Singh [245][244][246] developed a background sub-

traction based on the low frequency sub-band characteristics of the object image, and exploited the local spatial coherence of the foreground objects to achieve a more robust foreground detection in presence camera jitter and illumination changes. In a similar way, Khare et al. [277] used a single change detection method based on Daubechies complex wavelet while Kushwaha and Srivastava [305] used a median filter to model the background. In a further work, Khare et al. [276] used a double change detection method based on Daubechies complex wavelet coefficients of three consecutive frames. In an othe work, Kushwah and Srivastava [303][302][304] proposed a framework for dynamic background modeling based on Daubechies complex wavelet domain. First, wavelet decomposition of frame is obtained using the Daubechies complex wavelet transform. The change detection is then achieved by using detail coefficients (LH, HL, and HH), and the dynamic background model is obtained by an improved Gaussian mixture-based model applied on the approximate coefficient (LL subband). Experimental results [304] show that the Daubechies complex wavelet based method outperforms in presence of dynamic backgrounds the original codebook model [287], the single change detection method based on Daubechies complex wavelet [277] and MDLDWT [215].In an other approach, Mendizabal and Salgado [365] proposed to model the background at the region-levelin a wavelet based multi-resolution framework. Practically, the background model is obtained for each region independently as a mixture of K Gaussian modes, by considering both the model of the approximation coefficients and the model of the detail coefficients at the different decomposition levels. Experimental results [365] show the robustness of this approach in presence of sudden illumination changes and strong shadows. All of the previous wavelet based methods are based on the same well-established moving object detection framework in which foreground objects are detected according to the differences of features between adjacent frames or between the current frame and background models. Furthermore, the previous works only use various two-dimensional (2D) WTs to extract approximate coefficients and wavelet coefficients as features, in order to calculate differences between adjacent frames [15][104][14][215][276][305][277] or between the current frame and background models [158][159][160][161] [175][245][244][246][365]. In this context, Han et al. [197] proposed a completely different method from the above mentioned framework of the other works and developed a background subtraction based on Three-Dimensional Discrete Wavelet Transform (3D-DWT). After analyzing frequency domain characteristics of the intensity temporal consistency of static backgrounds, the 3D-DWT based method decomposes the data cube built with a set of consecutive frames into multiple 3D DWT sub-bands, and then the relationship between static backgrounds and certain 3D-DWT sub-bands is established. Thus, the background and the foreground are separated in different 3D-DWT sub-bands. Practically, the 3D-DWT based method directly removes the background and retains the foreground by discarding sub-bands corresponding to the backgrounds. 3D-DWT presents the advantage of the frequency domain characteristics of intensity temporal consistency. Experimental results [197] show that the 3D-DWT based method rapidly produces accurate foreground masks in challenging situations lacking training opportunities and outperforms ViBe [32],

2D-DFT [547] and 2D-UWT [15].

- **Curvelet Transform Features:** Because the wavelet transform can not describe curve discontinuities and can present inaccurate segmentation of moving object due to non-removal of noise in consecutive frames, Khare et al. [278] proposed to use curvelet transform. Thus, moving objects are detected with a change detection applied on curvelet coefficients of two consecutive frames. Experimental results [278] show that the change detection based on curvelet transform is more robust to noise than the change detection with Discrete Wavelet Transform (DWT) [221].

- **Walsh Transform Features:** Tezuka and Nishitani [534][536][535] modelled the background using the original MOG [512] applied on multiple block sizes obtained through the Walsh transform (WT). Walsh spectrum feature parameters are determined by using a set of coefficients from vertical, horizontal and diagonal directions, exhibiting strong spatial correlation among them. By using WT of the luminance component, four features are computed. The spectral nature of WT also reduces the computational steps required in feature extraction. Furthermore, Tezuka and Nishitani [534][536][535] developed a Selective Fast Walsh Transform (SFWT) with WT parameters consisting of only the low frequency coefficients. Experimental results [534] show that the MOG with the Walsh Transform Features outperforms the MOG with RGB features [512] and the single Gaussian with DCT features [693].

- **Hadamard Transform Features:** Baltieri et al. [30] proposed a fast background initialization method designed at the block-level in a non-recursive manner to obtain the best background model using the least 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 using the Hadamard Transform which is faster than DCT. Experimental results [30] demonstrate that this method outperforms its DCT counterpart [447].

- **Slant Transform Features:** Haberdar and Shah [193][192] proposed a framework for detecting relevant changes in dynamic background scenes. Practically, the changes are classified into two main classes called ordinary changes and relevant changes. Thus, this framework is based on a set of orthogonal linear transforms which allow to capture spatio-temporal features of local ordinary change patterns and subsequently employ them in the detection of relevant changes. Three orthogonal linear transforms as the base transforms are used: **1)** discrete cosine transform (DCT), **2)** Walsh-Hadamard transform (WHT) and **3)** Slant transform (ST). Because DCT, WHT, and ST provide complementary basis vectors, their combination allows to capture different types of ordinary change patterns. DCT is a sinusoidal transform that is widely used to obtain compact representations. WHT is a non-sinusoidal transform providing basis vectors that are rectangular or square wave. WHT can represent patterns with sharp discontinuities

more accurately with fewer values than DCT. ST provided basis vectors derived from sawtooth wave-forms and are a good complement to WHT. Depending on the ordinary change patterns in each set, elements of the data sequence are assigned to one of these three base transforms. Experimental results [193][192] on the ChangeDetection.net 2012 dataset that this approach is more robust in the dynamic background category than DPGMM [194], Spectral-360 [470], and [380].

– **Gabor Features:** In the works of Xue et al. [607][609], the input image is first convolved with local Gabor filters so that each pixel has a group of features containing multiple amplitudes and corresponding phase values. Then, the features with the most effective phase information is selected according to the criteria that higher amplitude value in the feature group means more accurate local structure information has been captured, and its corresponding phase information is more representative. This phase feature is then defined as the sum of the selected phase values. This phase feature presents several advantages for background modeling and foreground detection. First, it is insensitive to illumination changes as it is an inherent property of phase information. Second, the feature is relatively stable. Although noise exists in real videos, such feature change very little. Third, the wide value range makes it more suitable for background modeling and foreground detection. Finally, the feature is discriminative. When the true foreground appears in the scene, its value changes rapidly and drastically. For the baground model, Xue et al. [607][609] used the original MOG [512] and refined the foreground mask with blob aggregation using the Euclidean distance transform. Experimental results [607][609] on I2R dataset [326] that the MOG with the phase feature is more robust in presence of dynamic backgrounds and illumination changes than the original MOG [512], KDE [142] and LBP [207]. In an other approach, considering the advantages of Gabor filters which include the robustness to illumination changes, Wei et al. [588] adopted Gabor filters at multi-scale and multi-orientation to decompose an input video for sequential spatial feature extraction. Thus, a spatial feature vector (SFV) is computed for each pixel. Then, Wei et al. [588] used the MOG model [512] on the SFVs. Experimental results [588] on the Wallflower dataset [545] show that this feature is more robust than the RGB one in the presence of illumination changes.

15.2 Features from Video Domain Transform

The primary purpose of pixel-level background modeling is to detect moving objects. However, in video coding and transmission applications, its use is to compress video data without degrading image quality and to transmit as less data as possible (only the data that have changed [425][71][72][665][673][666]). Although features from the pixel domain are not effective for compression, several features from the video compressed domain are efficient for background/foreground separation. Furthermore, motion information is directly available without incurring cost of estimating a motion

field. This set of features include:

- **Features in MPEG Domain:** Babu et al. [18] proposed to extract moving objects
from MPEG compressed video by using motion vectors (MVs) which are sparse
in MPEG. Thus, MVs are used for automatically estimating the number of objects
and extracting independently moving objects. Thus, MVs are accumulated over
few frames to enhance the motion information, and are further spatially interpo-
lated to get dense motion vectors. The foreground mask is obtained by using an
expectation maximization (EM) algorithm. To determine the number of motion
models used in the EM step, Babu et al. [18] used a block-based affine clustering
method, and the segmented objects are temporally tracked. Finally, precise object
boundaries are obtained with an edge refinement. In an other approach, Zeng et
al. [645] proposed a moving object extraction which discriminated background
and moving objects by means of the higher-order statistics applied on the inter-
frame differences of Discrete Cosinus Transform (DCT) image. Thus, the DCT
image is partially decoded from the compressed video for a rapid reconstruction
of image data. The background is detected by the moment preserving thresh-
olding technique. Based on the background statistic, an optimal threshold based
on the background variance allows to extract the final object mask by compari-
son the fourth moment measure and the variance. In a further work, Wang et al.
[436] presented three algorithms (running average, median, MOG) which model
the background directly from compressed video. Each algorithm used DCT co-
efficients including AC coefficients at block level to represent background, and
adapted the background by updating DCT coefficients. The foreground mask is
obtained with pixel accuracy through two-stage approach. Thus, the block regions
fully or partially occupied by moving objects are identified in the DCT domain,
and then pixels from these foreground blocks are further classified in the spatial
domain. Experimental results [436] show the three algorithms achieved compa-
rable accuracy to their counterparts in the spatial domain. Furthermore, the com-
putational cost of the proposed median and MoG algorithms are only 40.4% and
20.6% of their counterparts in the spatial domain. In an other work, Porikli et al.
[435][436] exploited the macro-block structure to decrease the spatial resolution
of the processed data, which exponentially reduces the computational time. Fur-
thermore, Porikli et al. [435][436] used temporal grouping of the intra-coded and
estimated frames into a single feature layer. To achieve foreground detection, the
DCT coefficients for I-frames and block motion vectors for P-frames are com-
bined and a frequency-temporal data structure is constructed. From the blocks
where the AC-coefficient energy and local inter-block DC-coefficient variance
is small, the homogeneous volumes are enlarged by evaluating the distance of
candidate vectors to the volume characteristics. Affine motion models are fit to
volumes. Finally, the foreground mask is generated with a hierarchical clustering
stage which iteratively merges the most similar parts.

- **Features in H.264/AVC Domain:** Dey and Kundu [124] used the temporal statis-
tics of feature vectors, describing macroblock (MB) units in each frame. Thus,

feature vectors are used to select potential candidates containing moving objects. From the candidate macroblocks, foreground pixels are determined by comparing the colors of corresponding pixels pair-wise with a background model. This approach allows each macroblock to have a different quantization parameter, satisfying the requirements of both variable and fixed bit-rate applications. Additionally, a low-complexity technique for color comparison is used to obtain pixel-wise precision at a negligible computational cost as compared to classical approaches. A similar approach can be found in Pope et al. [433]. But, these methods [124][433] depend only on the number of encoding bits and fail to detect motion/activity when the bitrate is severely constrained and rate-distortion optimization is enabled. To solve this problem, Dey and Kundu [125] proposed enhanced MB features which use both encoding bits as well as the quantization step-sizes of individual coefficients in a MB. This is particularly important as the quantization parameters vary widely between simple and complex sections of an encoded image at lower bitrates. Then, the method is a two-steps hybrid foreground detection where MBs covering prospective foreground regions are first identified. An adaptive thresholding technique selected the candidate MBs corresponding to foreground objects. In the second step, pixels constituting the selected MBs are classified into background or foreground. Because the presence of shadows usually causes false classification, Dey and Kundu [125] estimated pixel differences using the Luv color instead of using the YCbCr color space [124]. Experimental results [125] shows that the enhanced MB approach with Luv color space outperforms both enhanced MB approach with YCrCb and the classical MB approach [124]. In an other approach, Tong et al. [541] first constructed the background model using the average MB RDCost over the initial N successive frames. After the MB mode selection during encoding, each MB RDCost which is available is compared to the background model. If an MB has RDCost larger than a pre-defined threshold, it is regarded as motion MB. But, holes may appear inside large moving objects because MBs with less change in large moving objects have small RDCost, making those MBs to be misclassified as background. To solve this problem, Tong et al. [541] used spatial refinement. If more than half of the eight adjacent MBs of a background MB are moving, then the MB under consideration is regarded to be in motion. Further, moving objects are subjected to temporal refinement.

– **Features in HEVC Domain:** For surveillance video coding, the rate-distortion analysis shows that a larger Lagrange multiplier should be used if the background in a coding unit took a larger proportion. An adaptive Lagrange multiplier is better for rate distortion optimization. Thus, Zhao et al. [674] developed a background proportion adaptive Lagrange multiplier selection method based on HEVC. From the analysis of the relationship between the Lagrange multiplier and the background proportion, a Lagrange multiplier selection model for surveillance video is used. However, there is a large amount of static background regions which provide motion vectors equal to zero. Because motion search is very time-consuming in the process of video coding, Zhao et al. [673] proposed a background-foreground division based search algorithm (BFDS) to accelerate

the motion search in surveillance video coding by utilizing the background and foreground information of coding units. The idea is to classify a predicting unit into a background predicting unit or a foreground predicting unit and then adopt different search strategies for each unit. In an other work, Chakraborty et al. [73] proposed an adaptive weighted non-parametric background model for efficient video coding in the HEVC domain.

15.3 Features from Compressive Cameras

While the previous works performed background subtraction on compressed images, they do no addressed compressive features from cameras that record MPEG video directly. Thus, several works have been developed for Compressive Sensing (CS) imaging, and not compressed video files. Compressive features are then obtained by using a basis which provides a K-sparse representation. Practically, many different basis can provide sparse approximations of images such as wavelets, curvelets and Gabor transformations. So, an image does not result in an exactly K-sparse representation and its transform coefficients decay exponentially to zero. In the literature, several approaches based on compressive features have been used and differ from the basis, the optimization method and the background model.

First, Cevher et al. [70] used an orthonormal basis and a Basis Pursuit Denoising (BDP) algorithm for the minimization problem while Cevher et al. [69] used Lattice Matching Pursuit (LaMP) [69] in a further work. This algorithm needs a large amount of storage and computation for training the object silhouette, which is not suitable for real-time background/foreground separation. In an other work, Mota et al. [383][382] proposed a l_1-l_1 minimization. Needell and Tropp [397] developed an algorithm called (CoSaMP) while He et al. [204] obtained compressive features through linear compressive measurements and used an improved CoSAMP algorithm called CoSaMP-Subspace. In an other work, Warnell et al. [585][584] proposed two methods: one based on cross-validation measurements with an algorithm named (ARCS-CV), and a second one based low-resolution measurements with an algorithm named (ARCS-LR). In an other approach, Li et al. [325] employed an orthonormal wavelet basis, a linear programming technique for the minimization problem and the running average for the background model. Wang et al. [578] developed a Gradient projection for sparse reconstruction (GPSR). Xu and Lu [602] proposed two approaches: one based on canonical sparsity basis, and the second based on wavelet sparsity basis with a K-cluster-valued CoSaMP algorithm. For linear compressive measurements, Davies et al. [114] compared the Basis Pursuit (BP) and Orthogonal Matching Pursuit (OMP) algorithms. Wang et al. [581] used wavelets transform to obtain a sparse representation. Shah et al. [474] used an algorithm called Convex Lattice Matching Pursuit (CoLaMP). For embedded camera networks, Shen et al. [482] developed two algorithms in the case of random projections basis called CS-MoG and Colour Space Compressed Sampling (CoSCS)-MoG with YCrCb components,

respectively. Liu and Pados [345] used Walsh-Hadamard measurements (l_1-PCA).

In the Robust Principal Component Analysis (RPCA) framework [67][57], several works used compressive sensing features too. Waters et al. [586][587] proposed to recover the low-rank and sparse matrices from compressive measurements in method called SpaRCS for SPArse and low Rank decomposition via Compressive Sensing. In a further work, Kyrillidis and Cevher [587] provided a real-time implementation for SpaRCS while Ramesh and Shah [441] developed a Regularized version of the SpaRCS algorithm called R-SpaRCS. In an other approach, Zonoobi and Kassi [696] proposed a modification to the SpaRCS algorithm in order to incorporate priori-knowledge of the sparse component. In an other approach, Jiang et al. [260][261] proposed a background subtraction based on low-rank and sparse decomposition by using the compressed measurement too. Although this model is suitable to the limited bandwidth of multimedia sensor networks, it is not enough robust to the movement turbulences and sudden illumination changes because the wavelet transform coefficients are not sparse in the case of turbulences. In a further work, Yang et al. [615] developed an adaptive CS-based algorithm which can exactly and simultaneously reconstruct the video foreground and background by using only 10% of sampled measurements. However, it still uses the wavelet transform as Jiang et al. [260][261]. This causes false detection in presence of turbulences and sudden illumination changes. In an other work, Li and Qi [328] proposed a recursive Low-rank and Sparse estimation through Douglas-Rachford splitting (rLSDR) algorithm by recursively estimating low-rank and sparse components in the reconstructed video frames from CS measurements.

In a low-rank minimization (LRM) framework, Shu et al. [487][488] proposed a three-dimensional compressive sampling (3DCS) approach to decrease the required sampling rate of the CI camera. So, a generic three-dimensional sparsity measure(3DSM) is decoded a video from incomplete samples by exploiting its 3D piecewise smoothness and temporal low-rank property. Furthermore, a decoding algorithm is used for this 3DSM with guaranteed convergence. Experimental results [487][488] show that the 3DCS based method requires a much lower sampling rate than the existing CS methods with the same accuracy. In an other work, Kang et al. [271][270] used a three-dimensional circulant compressive sampling method to obtain sampled measurement, based on which the video foreground and background are reconstructed by solving an optimization problem. Experimental results [271][270] on the I2R dataset demonstrate that this method outperforms both the original RPCA [67] and the original MOG [512] in presence of dynamic backgrounds.

16 Multiple Characteristics

The use of multiple characteristics consists to use additional values of the directed value of the concerned feature such as the mean, the median or high-order statistics. For example, Thongkamwitoon et al. [539][540] and Amnuaykanjanasin et al. [12] used four tuples which contains the expected color vector, the color covariance

matrix, the brightness distortion and the color distortion. An other representative approach is the codebook model in which minimum brightness, maximum brightness, frequency with which the code word has occurred and maximum negative run-length (MNRL) defined as the longest interval during the training period that the codeword has not recurred, and the first and last access times, respectively, that the codeword has occurred. So, the codebook model contains color (RGB [287], YUV [123] and HSV [108] in the cylinder color model, YUV [216] in the spherical color model, and HSV in the hybrid cone cylinder color model [131][132]), brightness and temporal information even if it uses color features.

17 Multiple Features

The use of multiple features (also called Bag-of-Features (BoF) [632][327][517]) for background modeling has become a promising solution to improving robustness in real applications. The fundamental idea is to add spatial and/or temporal dimensions to the already existing spectral information available from the visual scene. The different features can be obtained from the same sensor (that is one camera) or from different sensors such as IR cameras or RGB-D cameras. Table 9, Table 10 and Table 11 summarize different strategies based on multiple features in terms of number of features, fusion operators and background models.

17.1 Two features

One of the popular choice of multiple features has been the combination of an alternative spatial feature to the already existing color features. Some common extensions of the color features to multiple features include:

- **Color-Gradient:** As the gradients of an image are relatively less sensitive to changes in illumination, some previous works have added it to the color feature to obtain quasi illumination invariant foreground detection. Jabri et al. [240] were the first authors who combined the intensity with the gradient features. The foreground detection mask was obtained by an union operator and the features were considered as independent. Holtzhausen et al. [210] used the same fusion scheme but with RGB features instead of intensity features. In an other way, Javed et al. [249] used the RGB features with the gradient but with the AND operator. To take into account that the dependence within the features, Zang and Xu [654] used Ohta and gradient with the Sugeno integral, and Ding et al. [128] YCrCb and gradient with the Choquet integral which is more suitable than the Sugeno integral for the application of foreground detection. Approaches with fuzzy integrals are more robust than approaches with AND or OR operators as can be seen in [654][128].

- **Color-Texture:** An other and one of the more common approaches of extending the feature space beyond color is the addition of texture features. It has been

shown that the addition of texture to the color feature could improve robustness towards illumination changes and shadows as in [95]. For texture features, most of the authors used 1) the LBP [95][306] or one of its variants (ULBP [636],DLBP [600]), 2) the LTP [670][669] or one of its variants (SILTP [394][258]), or 3) statistical feature [92][162] or fuzzy statistical features [89][93]. In literature, different fusions scheme were used such as AND operator [333], weighted average [95][259][636][669] [258], and fuzzy integrals (Sugeno integral [655], Choquet integral [21][23][352][29][92][351][162][89], Interval valued Choquet Integral [93]). Even if there is no rigorous study of the best combination and fusion scheme, the YCrCb color space and LBP features aggregated with the Choquet integral [21], and intensity and FST features aggregated with the Interval valued Choquet Integral [93] seem to be the more suitable and robust solutions for foreground detection in presence of illumination changes and dynamic backgrounds.

– **Color-Depth:** With the increased use of RGB-D types of sensor, the combination of color together with depth information has emerged as a popular strategy for improved foreground detection. This combination of color and depth features has proven to deal with the camouflage in color and in some specific studies stereo features (disparity, depth) have also been used. Because depth features are different of color features as its distribution is different, the fusion scheme needs to be suitably chosen as the independence aspect [398]. The most representative works in color-depth are the ones of Camplani et al. [66][65][62][63], Fernandez-Sanchez [148][149], and Gallego and Pardas [545]. More generally, different fusions scheme were used in literature as AND or OR operator [148][149], weighted average [66][62][63] and logarithmic opinion pools [153]. Most of the time, the color space is the RGB one [217][66][65][62][63][148][149][153][410] but HSV and LUV color spaces can be found in [503] and [331], respectively.

– **Color-Motion:** In addition to color information, the use of motion can allow the foreground detection to deal with unimportant movements such as in dynamic backgrounds. Zhou and Zhang [690] combined intensity with optical flow information. Gong and Chen [165] used both RGB color components and 2D motion vector obtained by optical flow in a 1-SVMs background model. The GPU implementation processes QVGA-sized video sequences at 39.3 fps on a laptop. Lin et al. [335][336] used the optical flow value and the mean of inter-frame image difference as features in a probabilistic SVM approach for background initialization. Martins et al.[362] used a bio-inspired hybrid segmentation which merges information from two inherently different methods: (1) bio-inspired motion detection method using a feature called Magno channel based from the modeling of the human visual system, and a background subtraction algorithm based on pixel color information. The foreground detection is obtained by merging with a logical AND the detection of the two methods.

– **Color-Location:** Sheikh and Shah [481] proposed to use the location in addition to normalized RGB to exploit the dependency between pixel and thus spatial coherence can be taken into account with the KDE model. In an other work, Dickin-

son et al. [126] used an adaptive MOG in 5-dimensional feature space (2D image coordinates, and RGB color coordinates).

– **Texture-Motion:** Zhong et al. [682] used LBP and motion information obtained by a temporal operator (See Section 8).

17.2 More than two features

The combination of multiple features, beyond two, seems useful particularly if the combination exploits spectral, spatial and temporal information. However, such larger combination of features requires careful consideration to avoid redundancy that could impact computational demand and also preserve discriminative ability of features to ensure wide range of adaptation capabilities. In literature, there are different combination of three features: *1)* color-edge-texture [17], *2)* color-edge location [632], *3)* color-texture-motion [327], and *4)* color-texture-location [504]. Thus, Azab et al. [17] aggregated three features, i.e color (RGB), edge (obtained with the Sobel operator) and texture (LBP) using the Choquet integral to deal with illumination changes and dynamic backgrounds. The Choquet integral is used in aggregating color, edge and texture confidence maps. Experimental results of [17] on the PETS 2006 dataset showed that the addition of the edge feature increases the performance of the similar method [21] which used only the color and the texture features.

More than three features can be found in ensemble of features based approaches [294][195][349], feature selection schemes (See Section 18) or bag-of-features approaches [632][327][517]. For example, Klare and Sarkar [294] proposed an algorithm that incorporates multiple instantiations of the MOG algorithm with 13 features including: *1)* color features (RGB), and *2)* edge features which are the gradient and magnitude obtained with a Canny edge detector, eight texture feature (Haar features). The fusion method used is the average rule. Similarly, Han and Davis [195] performed background subtraction using a Support Vector Machine over background likelihood vectors for a set of features which consist of 11 features: color featurees (RGB), gradient (horizontal and vertical) and six Haar-like features. Then, Han and Davis [195] used a SVM classifier over the background probability vectors for the feature set. The aim of integrating the classifier for foreground/background segmentation is the selection of the discriminative features. Moreover, it also reduced the feature dependency problem. Otherwise, highly correlated non-discriminative features may dominate the classification process regardless of the states of other features. A radial basis function kernel is used to deal with non-linear input data. As Han and Davis [195] trained the classifier based on probability vectors rather than feature vectors directly, a universal SVM is used for all sequences and not a separate SVM for each pixel nor for each sequence.

Lopez-Rubio and Lopez-Rubio [349] developed a method based on stochastic optimization [348] which overcomes the limitations of some features along with a set

of relevant features that yields adequate results. Thus, a probabilistic model, in addition to handling any number of pixel features, could also account for the correlations among the features, so that a more realistic model is obtained. This method is called Multiple Feature Background Model (MFBM). Since the number of features N_f is not restricted, a fast and numerically stable implementation is of great importance to the practical applicability of MFBM. The slowest part of MFBM is the computation of the Gaussian probability density. It is $O(N_f)$ due to the inversion of the covariance matrix, and the computation of its determinant [349]. The other equations are $O(N_f)$ or lower. The following 24 features are used: *1)* Normalized RGB are used due to their robustness with respect to illumination changes, *2)* six Haar-like features which are robust features that convey texture information, *3)* previous features 1 to 9 processed with the bidimensional median filter of window size 5×5 pixels, *4)* Gradient features estimated with Sobel operator are less affected by illumination changes while they provide local texture information, *5)* two features which consider both color information and a pixel adjacent to the pixel at hand, as a local indication of color texture, and *6)* two small filters of size 3×3 pixels to extract the texture information. Experimental results [349] on several large scale dataset (I2R dataset [326], ChangeDetection.net [169], BMC 2012 dataset [552]) tested all possible combinations of pairs of features and triples of the 24 features. The conclusion is that Haar-like features are not well suited for MFBM which needs few features with the best possible discriminative power. The performance of MFBM increases as more relevant features are added, but the improvement decreases until a point where inserting more features is useless ($N_f = 5$). Normalized color channels and median filtered feature present particularly good results. This is due to their robustness against illumination changes for normalized features and their ability to filter the background noise for median filtered features. State-of-the-art background modeling approaches are outperformed by MFBM method but it seems that is due to the number and kind of features and not due to the model itself. In a further work, Molina-Cabello et al. [377] used the same set of features with a background subtraction method called Features based FSOM (FFSOM) which is an extension of Foreground Self-Organizing Maps (FSOM) [347].

In Subudhi et al. [517], six local features are taken into consideration: *1)* Three existing features: brightness (Bs), inverse contrast ratio (ICR) and average sharpness (AS)), and *2)* three proposed features: absolute relative height (ARH), local weighted variance (LWV) and integrated modal variability (IMV)). As most of the background subtraction methods consider non bi-unique model (which is the fact that local changes corresponding to the moving objects are obtained by making a combination of multiple features rather than combination of decision on individual features), Subudhi et al. [517] suggested the use of a bi-unique model as in Bovolo et al. [58] where individual spectral properties are combined or results of the spectral channels are fused to obtain better results. Experimental results on the ChangeDetection.net dataset [169] show the robustness of this method in the categories "Dynamic Backgrounds", "Camera Jitter" and "Thermal".

Wang and Wan [579] used 10 features: three in color (RGB components), six in gradient computed via Prewitt operator, and one in Gabor filter [268]. Then, Wang and Wan [579] proposed a Multi-Task Robust Principal Component Analysis (MTR-PCA) model which integrate multi-feature jointly into the RPCA framework [67]. Experimental results on the ChangeDetection.net dataset [169] show that the MTR-PCA with all the features outperforms MTRPCA with color and gradient, RPCA with color, and RPCA with gradient. Gan et al. [154] proposed a similar method called Multi-feature Robust Principal Component Analysis (MFRPCA).

Exploiting multiple cues, Huerta et al. [234] first combined both intensity and color (chromatic and brightness distortion) features in order to solve some of the color motion segmentation problems such as saturation or the lack of the color when the background model is constructed. Nonetheless, some colors problems are still unsolved such as dark and light camouflage. Then, to solve these problems, Huerta et al. [234] used edge features obtained with the Sobel operator. Finally, intensity, color and edge cues are combined using the logical operators AND and OR. In a further work, Huerta et al. [232] improved these multiple cues approach by using a chromatic-invariant cone model for the color features.

Features	Categories (number of papers)	Background Model	Fusion (number of papers)	Dependence-independence	Authors – Dates
Two Features	**Color-Edge (12)**				
	Intensity-Gradient	-	Union (Foreground, Boundary)(1)	Independent	Jabri et al. (2000) [240]
	rgs-Gradient	KDE [142]	(1)	Independent	Li et al. (2009) [330]
	rgb-Gradient	KDE [105]	(1)	Independent	Cuevas and Garcia (2010) [105]
	RGB-Gradient	-	AND (Foreground, Boundary) (1)	Independent	Javed et al. (2002) [249]
	RGB-Gradient	-	Union (Foreground, Boundary)(1)	Independent	Holtzhausen et al. (2012) [210]
	IHLS-Gradient (Multi-cues)	-	- (1)	Independent	Unzueta et al. (2011) [551]
	Normalized RGB-HSV-Gradient	MOG [512]	- (1)	Independent	Bin and Liu (2010) [44]
	RGB-HOG	MOG [386]	AND, OR	Independent	Mukherjee et al. (2010) [386]
	HSV-Gradient	KDE [142]	(1)	Independent	Park and Lee (2011) [420]
	RGB-LDN	ViBe [32]	Weighted scheme (1)	Dependent	Roy et al. (2017) [452]
	Ohta-Gradient	-	Sugeno Integral (1)	Dependent	Zhang and Xu (2006) [654]
	YCrCb-Gradient	-	Choquet Integral (1)	Dependent	Ding et al. (2009) [128]
	Color-Texture (35)				
	LCP-LBP	-	Weighted Average (2)	Independent	Chua et al. (2012) [95]
	Intensity-SKR Texture	-	Boundary Curvature (1)	Independent	Guo and Yu (2012) [181]
	RGB-BM	-	Conditions (Foreground) (1)	Independent	Lai et al. (2013) [306]
	Photometric Invariant Color-LBP	-	Boundary Texture (1)	Independent	Shang et al. (2012) [476]
	RGB-LBP	-	AND (Foreground) (1)	Independent	Lin et al. (2010) [333]
	RGB-LBP	K Multi-layers [259]	Weighted Average (1)	Independent	Jian and Odobez (2007) [259]
	Intensity-SCS-LTP	MOG [512], Pattern Adaptive KDE [670]	Product Likelihoods Foreground (1)	Independent	Zhang et al. (2011) [670]
	Intensity-SCS-LTP	MOG [512]	Weighted Average (1)	Independent	Zhang et al. (2012) [669]
	HSV-ULBP	K Histograms [207]	Weighted Average (1)	Independent	Yuan et al. (2011) [636]
	Color-DLBP	-	-	Independent	Xu et al. (2009) [600]
	RGB-ST-SILTP	K Histograms	Weighted Average (1)	Independent	Ji and Wang (2014) [258]
	Lab-SILTP	KDE [394]	- (3)	Independent	Narayama et al. (2012) [394]
	Normalized RGB-SILTP	MOG [512]	- (1)	Independent	Shi and Liu (2016) [483]
	Brightness-Chromaticity-Texture Variations	Conditional Probability Densities [567]	- (1)	Independent	Wang et al. (2016) [567]
	Ohta-LBP	-	Sugeno Integral (1)	Dependent	Zhang and Xu (2006) [655]
	YCrCb-LBP	-	Choquet Integral (3)	Dependent	El Baf et al. (2008) [21]
	IR-LBP	-	Choquet Integral (1)	Dependent	El Baf et al. (2008) [23]
	HSI-LBP	-	Choquet Integral (4)	Dependent	Ding et al. (2009) [127]
	RGB-LBP	-	Choquet Integral (1)	Dependent	Lu et al. (2012) [352]
	CrCb-LBP	-	Choquet Integral (1)	Dependent	Balcilar et al. (2013) [29]
	YCrCb-ULBP	-	Choquet Integral (1)	Dependent	Lu et al. (2014) [351]
	Intensity-ST	-	Choquet Integral (1)	Dependent	Chiranjeevi et al. (2014) [92]
	Intensity-ST	-	Choquet Integral (1)	Dependent	Gayathri et al. (2014) [162]
	Intensity-FST	-	Choquet Integral (1)	Dependent	Chiranjeevi and Sengupta (2012) [89]
	Intensity-FST	-	Interval valued Choquet Integral (1)	Dependent	Chiranjeevi and Sengupta (2016) [93]
	Color-Stereo Features (Part 1) (30)				
	RGB-Disparity	Fist Image	AND (Background), AND (Foreground) (3)	Independent	Ivanov et al. (1998) [237]
	RGB-Disparity	MOG [512]	OR (Foreground) (1)	Independent	Gordon et al. (1999)[167]
	YUV-Depth	MOG [512]	OR (Foreground)(3)	Independent	Harville et al. (2001) [202]
	Intensity-Depth	MOG [512]	OR (Foreground) (1)	Dependent	Song et al. (2014)[499]
	Intensity-Depth	MOG [512]	AND (Foreground) + Conditions (1)	Dependent	Silvestre (2007) [494]
	Intensity-Depth	KDE [142]	AND (Foreground) (1)	Dependent	Silvestre (2007) [494]
	YCrCb-Depth-Modulation	MOG [512]	OR (Foreground) (2)	Dependent	Langmann et al. (2010) [308]
	IR-Depth	MOG [512]	Boundary Depth (1)	Independent	Stormer et al. (2010) [514]
	Intensity-Depth	ViBe [32]	OR (Foreground) (1)	Independent	Leens et al. (2009) [317]

Table 9 Multiple Features: An Overview (Part 1).

Features	Categories	Background Model	Fusion (number of papers)	Dependence-independence	Authors – Dates
Two features	**Color-Stereo Features (Part 2)**				
	RGB-Depth	-	Weighted average (1)	Independent	Hu et al. (2014) [217]
	RGB-Depth	MOG [512]	Weighted average (1)	Independent	Camplani and Salgado (2013) [66]
	RGB-Depth	MOG [512]	Combination (1)	Independent	Camplani et al. (2013) [65]
	RGB-Depth	MOG [512]	Weighted average (1)	Independent	Camplani et al. (2013) [62]
	RGB-Depth	2 MOG Temporal (Background) 2 MOG Spatial (Foreground)	Weighted average (1)	Independent	Camplani et al. (2013) [63]
	RGB-Depth (Approach 1)	Codebook [288]	AND (Background) (1)	Independent	Fernandez-Sanchez et al. (2013) [148]
	RGB-Depth (Approach 2)	Codebook [288]	AND (Background) (1)	Dependent	Fernandez-Sanchez et al. (2013) [148]
	RGB-Depth	2 MOG Temporal (Background)	AND (Background) (1)	Dependent	Fernandez-Sanchez et al. (2014) [149]
	RGB-Depth	2 MOG Spatial (Foreground)	Logarithmic Opinion Pool (1)	Independent	Gallego and Pardas (2013) [153]
	RGB-Depth	MOG [695]	Combination (2)	Independent	Ottonelli et al. (2013) [410]
	xy-HSV-Depth	KDE [503]		Independent	Spampinato et al. (2014) [503]
	Color-Depth	MOG [512]	Two probabilistic models (1)	Independent	Song et al. (2014) [500]
	LUV-Depth	MOG [512]	-	Independent	Liang et al. (2016) [331]
	LUV-Depth	FuzzyAdaptiveSOM [357]	-	Independent	Liang et al. (2016) [331]
	RGB-Depth	KDE [142]	- (1)	Independent	Moya-Alcover et al. (2016) [384]
	Colorimetric Invariant-Depth	-	- (1)	Independent	Murgia et al. (2014) [389]
	RGB-Depth	GMM [512]	- (1)	Independent	Amamra et al. (2014) [9]
	RGB-Depth	Fast KDE	- (1)	Independent	Trabelsi et al. (2017) [546]
	RGB-Depth	ViBe	Weighted fusion rule (1)	Independent	Huang et al. (2017) [224]
	Color-Motion (5)				
	Intensity-Optical Flow	MOG [512]	Conditions (1)	Independent	Zhou and Zhang [690]
	Normalized RGB-Optical Flow	KDE [373]	(1)	Independent	Mittal et al. [373]
	RGB-2D motion vector	1-SVM	Weighted average (1)	Dependent	Gong and Cheng (2011) [165]
	Interframe Difference-Optical Flow	SVM	-(2)	Dependent	Lin et al. (2002) [165]
	Color-Magno Channel	MOG [512]	AND (1)	Independent	Matins et al. (2016) [362]
	Color-Location (7)				
	Normalized RGB-(x,y)	KDE [481]	- (2)	Independent	Sheikh and Shah [481]
	Normalized RGB-(x,y)	KDE [394]	- (3)	Independent	Narayana et al. [394]
	Intensity-(x,y)	KDE [142]	(1)	Independent	Antic and Crnojevic [13]
	rgl-(x,y)	KDE [142]	(1)	Independent	Zhou et al. [691]
	RGB-(x,y)	MOG [512]	- (1)	Independent	Dickinson et al. [126]
	Color-Intensity (2)				
	RGB-Intensity (Multi-cues)		AND-OR (1)		Huerta et al. (2007) [233]
	Color invariant H-Intensity (Multimodal)		- (1)		Cocorullo et al. (2016) [97]
	Intensity-IR (11)				
	Intensity-IR (Multi-sensors)		Contour Saliency (4)		Davis and Sharma (2005) [118]
	Intensity-IR (Multi-sensors)		Physics-based fusion (1)		Nadimi and Bhanu (2003) [392]
	Intensity-IR (Multi-sensors)		Evolutionary-based fusion (2)		Nadimi and Bhanu (2004) [393]
	Intensity-IR (Multi-sensors)		Transferable Belief Model (3)		Conaire et al. (2006) [100]
	Intensity-IR (Multi-sensors)		Joint Sample Consensus Model (1)		Han et al. (2013) [196]
	Color-IR (2)				
	RGB-IR (Multi-sensors)		- (1)		Pejcic et al. (2009) [426]
	RGB-IR (Multi-sensors)		AND/OR (1)		Yan et al. (2017) [614]
	Texture-Motion (1)				
	LBP-Temporal Operator		Weighted average (1)		Zhong et al. (2008) [682]
	Edge-Texture (1)				
	Gradient-LPTP [442]		- (1)	Dependent	Kim et al. (2015) [285]
	Texture-Location (1)				
	LBP-(x,y)	-	Weighted average likelihood (1)		Rodrigues et al. [451]
	Texture-Local Gradient Histograms (1)				
	LBP-HOG	MOG [512]	- (1)		Panda and Meher [415]

Table 10 Multiple Features: An Overview (Part 2).

Features	Categories	Background Model	Fusion (number of papers)	Dependence-independence	Authors – Dates
Three Features	**Color-Intensity-Edge**				
	Color-Intensity-Edge		OR-AND (1)	Independent	Huerta et al. [234]
	Chromatic-invariant cone model-Intensity-Edge		OR-AND (1)	Independent	Huerta et al. [232]
	Color-Intensity-Texture				
	Color-Intensity-LRP	Codebook [287]		Independent	Zaharescu and Jamieson [641]
	Color-Edge-Texture				
	RGB-Gradient-LBP	SG [589]	Choquet Integral (1)	Dependent	Azab et al. [17]
	RGB-Gradient-Gabor filter	MFRPCA [154]	-	Independent	Gan et al. (2013) [154]
	RGB-Gradient-Gabor filter	MTRPCA [579]	-	Independent	Wang and Wan (2014) [579]
	Intensity-Gradient-LBP	Covariance [550]	-	Dependent	Zhang et al. (2008) [663]
	RGB-Gradient- LBP	Codebook [287]		Independent	Lin et al. (2016) [334]
	Color-Edge-Location (1)				
	RGB-Gradient-(x,y)	KDE [632]		Independent	Yoo and Kim (2013)[632]
	Color-Texture-Motion (1)				
	RGB-LBP-Optical Flow	Nested model [327]		Independent	Li et al. (2014) [327]
	Color-Texture-Location (2)				
	RGB-Textons-(x,y)	KDE [481]		Independent	Spampinato et al. (2014) [504]
	Lab-U-LBP+(x,y)	KDE [481]		Independent	Yan et al. (2016) [613]
	Other features (1)				
	Object size-Direction-Speed			Independent	Jodoin et al. (2012) [265]
More than Three Features	**Ensembled based approaches (5)**				
	YUV-Edge Deviation-Vector Deviation-magnitude		OR (1)	Independent	Kamkar-Parsi et al. (2005) [269]
13 features	Color-Edge-Haar Features	MOG [512]	Weighted average (1)	Independent	Klare and Sarkar (2009) [294]
11 features	Color-Gradient-Haar Features	KDA [195]	SVM Classifier (1)	Independent	Han and Davis (2011) [195]
24 features	Color-Edge-Texture-Haar Features	Stochastic Approximation [348]		Independent	Lopez-Rubio and Lopez-Rubio (2014) [349]
24 features	Color-Edge-Texture-Haar Features	FSOM [347]		Independent	Molina-Cabello et al. (2016) [377]
5 features	Intensity-Color-Edge-Texture-Laplacian Features	OR-PCA with MRF [251]	Weighted similarity measure (1)	Independent	Javed et al. (2015) [251]
	Feature selection approaches (9)				
11 features	RRGB-Gradients,Color co-occurrences	Bayesian framework (1)		Independent	Li et al. (2004) [326]
9 features	RGB-Gradients (Horizontal/Vertical)	KDE (1)		Independent	Parag et al. (2006) [417]
9 features	RGB-HSV-YCbCr	ViBe (1)		Independent	Braham et al. (2015) [59]
9 features	HOG-LBP-Haar Features	online boosting (2)		Independent	Grabner and Bischof (2006) [170]
9 features	HOG-LBP-Haar Features	Time dependent online boosting (1)		Independent	Grabner et al. (2008) [171]
9 features	HOG-OCLBP[316]-Haar Features	Hierarchical online boosting (1)		Independent	Lee et al. (2011) [316]
7 features	Color-Gradient-HOG-LBP[207]	OR-PCA (1)		Independent	Javed et al. (2015) [254]
19 features	R,G,B, H,S,V, gray-scale, XCS-LBP, OC-LBP	SVM (1)		Independent	
	gradient orientation and magnitude, optical flow			Independent	
	7 spectral narrow bands			Independent	
	Bag of features (1)				
6 features	Bs-ICR-AS-ARH-LWV-IMV	ViBe [32]		Independent	Subudhi et al. (2016) [517]

Table 11 Multiple Features: An Overview (Part 3).

18 Feature Selection

Commonly, the background subtraction methods do not take into account the properties of each features, and the same feature is used globally for the whole scene. Indeed, all existing background subtraction methods operate with a uniform feature map in the sense that the features used to model the background and perform the detection are the same for all pixels of the image, thus ignoring the non-uniformity of the spatial distribution of background properties and neglecting the foreground properties. But in practice for complex scene comprising of several elements such as waving trees, sky, soil and car, the most discriminant features for these elements are probably different, and therefore a single-feature background subtraction algorithm may not be appropriate.

Despite the choice of the best features for each region is not an easy task as it requires a deep knowledge of the scene, it is possible to automatically select the most relevant feature to improve the foreground segmentation in complex scenes thanks to their capability to select a subset of highly discriminant features removing irrelevant and redundant ones. One can find ensemble-based approaches, minimizing the use of traditional feature selection methods (filters, wrappers and embedded) [511]. The objective of such approaches is to generate multiple feature selectors and then aggregating their outputs, producing a more robust classification. The main known re-sampling ensemble methods that generate and combine a diversity of learners are: bagging, boosting and random subspace [409]. The boosting approach and its variants for feature selection has been used in [170][172][171][316][417]. However, usually only exemplars of one-class elements are available (i.e. the background component is always present), whereas the other classes are unknown (i.e. foreground objects can appear/disappear several times in the scene). This is known as the one-class classification (OCC) problem, which specific nature is not taken into account in boosting approaches. The OCC approach has been used by Silva et al. [491][492]. Practically, the different feature selection scheme applied to background modeling and foreground detection can be classified as follows:

– **Predetermined feature selection:** Li et al. [326] were the first to express the need for modeling distinct part of the image with different features, and described the background image as consisting of two pixel categories, static pixels and dynamic pixels. Unfortunately, this method lacks of generality as *1)* the cardinality of the feature set, i.e. the number of candidate features or group of features that are predetermined, is limited to two (one group of features for static pixels and one other for dynamic pixels), and *2)* the choice of features is not made dynamically but only the type of pixels.

– **Feature selection via AdaBoost:** Grabner and Bischof [170][172] introduced an on-line boosting based feature selection framework using the Adaboost algorithm[150]. This method was enabled to tackle a large variety of challenges and used large feature pools at reduced computational costs. This gives the feasibility to achieve real-time computational complexity. The feature set contained *1)*

Fusion Scheme	Categories	Authors
Classical Operators	AND	*Foreground:* Javed et al. [249], Lin et al. [333]
		Ivanov et al. [237], Ivanov et al. [238], Ivanov et al. [239]
		Distances: Silvestre [494]
		Background: Fernandez-Sanchez et al. [148], Fernandez-Sanchez et al. [149]
	OR	*Foreground:* Gordon et al. [167], Harville et al. [202]
		Harville et al. [201], Harville [200]
		Langmann et al. [308], Leens et al [317]
	Combination	*OR-AND-Morphological Operators:* Camplani et al. [65]
		Ottonelli et al. [410], Ottonelli et al. [411]
	Conditions	*Boundaries:* (Edge) Javed et al. [249], (Edge) Guo and Yu [181]
		(Texture) Shang et al. [476], (Depth) Stormer et al. [514]
		Weak/Strong Pixels: Lai et al. [306]
		Weak/Strong Motion Vectors: Zhou and Zhang [690]
	Weigthed Average	*Similarities:* Chua et al. [95], Chua et al. [96]
		Ji and Wang [258]
		Distances: Zhang et al. [669], Yuan et al. [636]
		Jian and Odobez (2007) [259]
		Gong and Cheng (2011) [165]
Statistical Operators	Product of the likelihoods	Zhang et al. [670]
	Weighted average of the likelihoods	Hu et al. [217], Camplani and Salgado [66]
		Camplani et al. [62], Camplani et al. [63]
		Klare and Sarkar [294]
		Zhong et al. [682]
	Logarithm Opinion Pool	Gallego and Pardas [153]
	SVM Classifier	Han and Davis [195]
Fuzzy Operators	Sugeno Integral	Zhang and Xu [654], Zhang and Xu (2006) [654]
	Choquet Integral	El Baf et al. [21], El Baf et al. [24]
		El Baf et al. [23], El Baf et al. [22]
		Ding et al. [130], Ding et al. [127]
		Ding et al. [128], Li et al. [329]
		Ding et al. [129], Azab et al. [17]
		Lu et al. [352], Lu et al. [351]
		Balcilar et al. [29], Chiranjeevi and Sengupta [92]
		Gayathri and Srinivasan [162], Chiranjeevi and Sengupta [89]
	Interval valued Choquet integral	Chiranjeevi and Sengupta [93]

Table 12 Features Fusion: An Overview.

the Haar-like features [557], *2)* orientation histograms with 16 bins similar to Dalal and Triggs [109], and *3)* a simple version with 4th neighborhood of LBP [207]. Grabner and Bischof [170][172] computed integral images and integral histograms [434] as efficient data structures for fast calculation of all features. The current image is divided into overlapping patches, and the patches are used as training data for foreground/background classifiers. As only background parts of the current image can be used as training data, Grabner and Bischof [170][172] generated an arbitrary image patch, of which the mean pixel value is 128 and the variance is $256^2/12$, as a foreground patch for initial training. As a new current image is given to the classifiers, the image is divided into patches and each patch is classified as either foreground or background using the trained classifiers. After classification, the patches classified as background with high confidence are used as training data to update the classifiers. The main advantage of this scheme is that it is robust to illumination changes and dynamic backgrounds since the classifiers are consistently updated. However, as it is an off-line learning algorithm, all labeled training samples are required to be available a-priori. In most real ap-

plications, training data should not always be sampled from one fixed distribution as it may not encapsulate all the complexities in real-scenes. However, since the method is based on self-learning (i.e. classifier predictions feed model updates), the background model can end up in catastrophic state, as mentioned by Grabner and Bischof [171]. Therefore, Grabner and Bischof [171] proposed a time-dependent on-line boosting algorithm using exponential forgetting of the samples over time. Nevertheless, these two algorithms naturally lose the advantage of color information and still suffers from large computational complexity [316]. Thus, Lee et al. [316] proposed a hierarchical on-line boosting which uses block based Opponent Color Local Binary Pattern (OC-LBP). Thus, Lee et al. [316] divided a patch into three layers of R,G, and B, and then each layer into 3 blocks. 9 different block based OC-LBP are generated by placing different colors to the center block and the surrounding blocks. For each block based OC-LBP, 8 classifiers are constructed. Then, Lee et al. [316] considered 9 classifier pools, each of which includes 8 classifiers. While the original on-line boosting [170][172][171] should generate arbitrary foreground patches, the hierarchical on-line boosting [316] only needs background patches. The computational time is shortened significantly by making the process of training and classification simple. As pointed out by Braham et al. [59], the drawbacks of these approaches are unrealistic assumptions about the statistical distributions of foreground features that are used, i.e. a uniform distribution is assumed for the gray value of foreground objects and serves as a basis for computing other feature distributions.

– **Feature selection via RealBoost:** Parag et al. [417] proposed a generic model that is capable of automatically selecting the features that obtain the best invariance to the background changes while maintaining a high detection rate for the foreground detection. In this study, Parag et al. [417] proposed the use of a Realboost algorithm. Unlike Adaboost algorithm which combines weak hypotheses having outputs in $\{-1, +1\}$, RealBoost algorithm computes real-valued weak classifiers given real numbered feature values, and generates a linear combination of these weak classifiers that minimizes the training error. To generate the background model, Elgammal et al. [417] used the Kernel Density estimation (KDE). Using the density estimates, the Realboost algorithm [467] is able to select the features most appropriate for any pixel. In the implementation phase, Parag et al. [417] used 9 types of features, namely three color values R,G,B and spatial derivatives for each of these color channels in both x and y directions for each pixel of a color image. For intensity images, 3 features are used, that are the pixel intensity value and its spatial gradients in horizontal and vertical directions. Experimental results [417] show that this approach outperforms the original KDE [142] by a slight margin. While this framework can be used with an arbitrary number of candidate features, it has a serious drawback as pointed out by Braham et al. [59]. Indeed, the synthetic foreground examples used for boosting are generated randomly from a uniform distribution. Indeed, all candidate features are assumed to be uniformly distributed for foreground objects. This assumption is not valid because of the wide variety of foreground statistical distributions among different features [59]. For example, gradient has a foreground probability density func-

tion concentrated around low values while RGB color components have a wider foreground probability density function. This may explain why the classification performance of this framework decreases when gradient features are added to the feature set. This limitation can be addressed by means of global statistical foreground models.

– **Dynamic feature selection strategy:** Javed et al. [254] used a dynamic feature selection strategy with Online Robust Principal Component Analysis (OR-PCA) framework. Feature statistics, in terms of means and variances, are used as a selection criterion. Unlike aforementioned purely spatial or spatio-temporal selection approaches, this method is exclusively temporal-based, which means that all pixels use the same features for the foreground segmentation. The non-uniformity of the spatial distribution of background properties is thus ignored.

– **Feature selection via a generic method:** Like in Parag et al. [417], Braham et al. [59] proposed that the feature selection process only occurs during the training phase, to avoid extra computations during normal background subtraction operations. Thus, the training phase is divided into three parts: *1)* The first part consists to accumulate images free of foreground objects which are further processed to build local background statistical models, *2)* The second part uses an other sequence of images, which include foreground objects and use to build a global foreground statistical model for each candidate feature of the feature set, and *3)* the third part selects the best feature/threshold combination which is a local process, meaning that it is performed for each pixel individually. The goal of the selection process consists to detect which feature is most appropriate to discriminate between the local background and the foreground at the frame level. Unlike previous approaches, the foreground feature statistical distributions are estimated at the frame level. Features are selected locally depending on their capability to discriminate between local background samples and global foreground samples for a specific background model and for a chosen performance metric. Experiments results [59] show that this feature selection scheme significantly improves the performance of ViBe [32].

– **Feature selection via One-class SVM:** Silva et al. [491] used an Incremental Weighted One-Class Support Vector Machine (IWOC-SVM) which select suitable features for each pixel to distinguish the foreground objects from the background. In addition, an Online and Weighted version of the Random Subspace (OW-RS) [491] is used to assign a degree of importance to each feature set, and these weights are used directly in the training step of IWOC-SVM. Moreover, a heuristic approach is used to reduce the complexity of the background model maintenance while maintaining the robustness of the background model. Practically, the features were chosen to have at least one feature in the five type of features: color feature (R,G,B, H,S,V and gray-scale), texture feature (XCS-LBP [490]), color-texture (OC-LBP [366]), edge feature (gradient orientation and magnitude), and motion feature (optical flow). In addition, multispectral bands (a total of 7 spectral narrow bands) are also used. Experimental results [491] on mul-

Algorithms	Background Model	Number	Features	Authors-Dates
Off-line feature selection				
Predetermined	Bayesian framework [326]	5	Static Background Pixels: RGB-Gradients	Li et al. (2004) [326]
		6	Dynamic Background Pixels: Color co-occurrences	
Feature selection in the training step				
Realboost [467]	KDE [142]	3	Intensity-Gradients (Horizontal/Vertical)	Parag et al. (2006) [417]
Realboost [467]	KDE [142]	9	RGB-Gradients (Horizontal/Vertical)	Parag et al. (2006) [417]
Generic method	ViBe [32]	9	RGB-HSV-YCbCr	Braham et al. (2015) [59]
On-line feature selection				
Adaboost [150]	Single Gaussian [589]	9	Haar Features, HoG [109], LBP [207]	Grabner and Bischof (2006) [170]
Adaboost [150]	Single Gaussian [589]	9	Haar Features, HoG [109], LBP [207]	Grabner et al. (2006) [172]
Time dependent Adaboost [150]	Single Gaussian [589]	9	Haar Features, HoG [109], LBP [207]	Grabner et al. (2008) [171]
Hierarchical Adaboost [150]	Single Gaussian [589]	9	Haar Features, HoG [109], OCLBP [366]	Lee et al. (2011) [316]
Selection criterion	OR-PCA [254]	7	Color-Gradient-HOG-LBP [207]	Javed et al. (2015) [254]
IWOC-SVM [491]	SVM	19	R,G,B, H,S,V, gray-scale, XCS-LBP, OC-LBP gradient orientation and magnitude, optical flow 7 spectral narrow bands	Silva et al. (2016) [491]
Superpixel-OWAOC [492]	SVM	19	R,G,B, H,S,V, gray-scale, XCS-LBP, OC-LBP gradient orientation and magnitude, optical flow 7 spectral narrow bands	Silva et al. (2017) [492]

Table 13 Features Selection: An Overview.

tispectral video sequences from the MSVS dataset [35] show the pertinence of this selection scheme. In a further work, Silva et al. [492] improved IWOC-SVM by using a best methodology to select the best features based on wagging, by using a superpixel segmentation strategy, and by adding Ensemble Pruning (EP) to suitably update the importance of each feature. This method is named Superpixel-based Online WAgging One-Class Ensemble(Superpixel-OWAOC). Experimental results [492] on the MSVS dataset [35] and the RGB-D dataset [66] show that Superpixel-OWAOC outperforms IWOC-SVM (also called OWOC-RS).

In these different methods, the feature selection is made at a different step: *1)* in off-line step in a predetermined way (as in Li et al. [326]) which is less adaptive over time because the set of features for each category of pixel is fixed over time and it only requires additional computation for the detection of the category of each pixel , *2)* in the training step only (as in Parag et al. [417] and in Braham et al. [59]) which is more robust than the off-line way with a set of selected features more flexible and thus more diverse for each pixel, or *3)* in an online way all over the process (as in Grabner and Bischof [170], Javed et al. [254], Silva et al. [491], and Silva et al. [492]) which is the most adaptive scheme but it requires additional computation time.

19 Resources, datasets and codes

19.1 Features Website

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 type of features as in this paper. An overview of the content of the Feature Website is given at the home page. In addition to the sections which concern the type of features, this website gives references and links to available implementations, datasets, conferences, workshops and websites.

19.2 Datasets

Several datasets available to evaluate and compare background subtraction algorithms have been developed in the last decade. We classified them in terms of IR datasets, color datasets and RGB-D datasets. IR and color datasets provide videos from one camera in IR or color space. On the other hand, RGB-D datasets provide color+depth videos obtained by stereo cameras or the Microsoft Kinect RGB-D camera or the Asus Xtion Pro Live camera. All these datasets are publicly available and their links are provided on the Features Web Site in the section Available Datasets.

19.2.1 IR dataset

The OTCBVS 2006 Dataset is related to the conference "Object Tracking and Classification in and Beyond the Visible Spectrum" (OTCVBVS[28]) 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.

19.2.2 Color datasets

Color datasets are the majority in the available datasets and the reader can find the full list in Bouwmans [51]. Here, we focus on the recent large-scale datasets and the datasets in aquatic environments.

1. **Large-scale datasets:** These datasets are realistic large-scale datasets with accurate ground-truth providing a balanced coverage of the range of challenges present in the real world. Three large-scale datasets are available and very interesting:

   – **ChangeDetection.net Dataset:** The CDW[5] 2012 dataset [169] 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. This dataset was extended in 2014 to the CDW[6] 2014 dataset which contains all the 2012 videos plus additional ones with the following challenges: challenging weather, low frame-rate, acquisition at night, PTZ capture and air turbulence.

– **BMC 2012 Dataset:** The BMC[7] (Background Models Challenge) [552][497] 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 SABS[8] (Stuttgart Artificial Background Subtraction) dataset [61] 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.

2. **Datasets in aquatic environments** These datasets are very interesting to test and evaluate illumination invariant color features and texture features due to the dynamic background changes and illumination changes particular to aquatic environments. Three datasets are publicly available:

– **Fish4knowledge Dataset:** The Fish4knowledge [9] dataset [274][502][501] 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.

– **Aqu@theque Dataset:** The Aqu@theque[10] dataset [427][20][19] contains four different image sequences from the "Aqu@thque" project [68], 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 boot-

strapping, occlusion, camouflage, light changes and dynamic background.

– **MAR Dataset:** Maritime Activity Recognition (MAR[11]) [46] 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.

19.2.3 RGB-D datasets

Three RGB-D datasets are publicly available and can be described as follows:

– **RGB-D Object Detection Dataset:** The RGB-D[12] dataset [66][65] provides five sequences of indoor environments, acquired with the *Microsoft Kinect RGB-D camera*. Each sequence contains different challenges such as cast shadows, color and depth camouflage. For each sequence a hand-labeled ground truth is provided.

– **CITIC RGB-D Dataset:** The CITIC[13] dataset [148][149] contains sequences recorded with rectified *stereo cameras*, and some frames have been hand-segmented to provide ground-truth information.

– **RGB-D Rigid Multi-Body Dataset:** The RGB-D Rigid Multi-Body [14] dataset [515] consists of 3 RGB-D videos of objects with different sizes (chairs, box/watering can, small box/teacan). The datasets have been recorded using an *Asus Xtion Pro Live camera* in a resolution of 640x480 at 30 Hz frame rate. Ground truth for the camera pose has been obtained with an OptiTrack Motion Capture system. The moving objects are also manually annotated in frames at every 5 seconds.

19.3 Libraries

Several codes for texture features are provided in Matlab by the Center for Machine Vision (CMV[15]) at University of Oulu: as follows: LBP, LBP Histogram Fourier Fea-

---

[5] http://www.changedetection.net/

[6] http://www.changedetection.net/

[7] http://bmc.univ-bpclermont.fr

[8] http://www.vis.uni-stuttgart.de/index.php?id=sabs

[9] http://www.fish4knowledge.eu

[10] http://sites.google.com/site/thierrybouwmans/recherche—aqu-theque-dataset

[11] http://labrococo.dis.uniroma1.it/MAR/

[12] http://www.gti.ssr.upm.es/ mac/

[13] http://atcproyectos.ugr.es/mvision/

[14] http://www.ais.uni-bonn.de/download/rigidmultibody/

tures [8], spatio-temporal LBP [672], discriminative features of LBP variants [189] and Local Configuration Pattern [188]. These codes do not specifically address the application of background modeling and foreground detection. In an other way, the LBP Library provided by Caroline Silva, Cristina Lazar and Andrews Sobral is a collection of eleven Local Binary Patterns (LBP) algorithms developed for background subtraction problem. The algorithms were implemented in C++ based on OpenCV. A CMake file is provided and the library is complatible with Windows, Linux and Mac OS X. The library was tested successfully with OpenCV 2.4.10. The local texture pattern available are the following ones: original LBP [207], CS-LBP [523], SCS-LBP [608], ϵ-LBP [568], HCS-LBP [606], OCLBP [316], VARLBP [408], BG-LBP [111], XCS-LBP [490], SILTP [332] and CS-SILTP [592].

20 Features and Evaluation

Fig. 5 shows the relation between the type of features used in each method, and the evaluation metrics that are used to evaluate each of them according to the reviewed papers. Here we can see that for each category there are no dominant evaluation metrics. Also, it is observed that a decent number of papers are using qualitative evaluation only, without providing quantitative results related to their accuracy and performance, especially when compared between older and more recent papers. Thus, it is a need to develop a rigorous quantitive evaluation of influence of each feature with the same measure metrics.

21 Conclusion

In conclusion, this review on the role and the importance of features for background modeling and foreground detection highlights the following points:

– Features can be classified following their size, their type in a specific domain, their intrinsic properties and their mathematical concepts. Each type of features presents different robustness against challenges met in videos taken by a fixed cameras. For the color feature, YCrCb color space seems to be the more suitable feature [24][350]. For the texture feature, Silva et al. [491] provided a study on the LBP and its variants that show that XCS-LBP is the best LBP feature for this application in presence of illumination changes and dynamic backgrounds. Although this study covered texture features, it is restricted to LBP features and then there is not a full study on the different texture features. For the depth feature, it needs to carefully used them following their properties as developed in Nghiem and Bremond [398]. Features in a domain transform are very useful to reduce computation times as in the case of compressive sensing features.

---

[15]http://www.cse.oulu.fi/CMV/Downloads/LBPMatlab
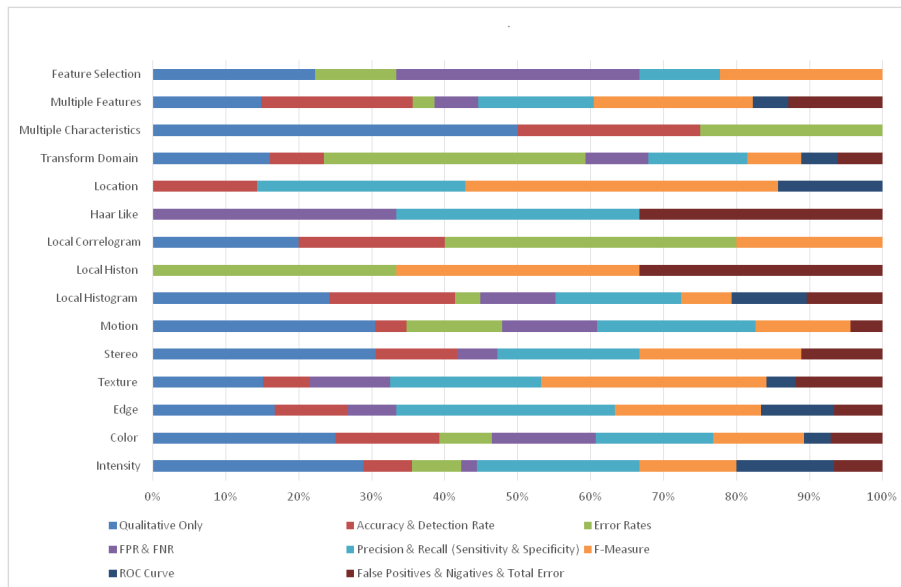
Fig. 5 Metrics used to evaluate the performance of background modeling and foreground detection following the type of features used.

– Several features have been used in other applications and none in background modeling and foreground detection such several variants of LBP (Multi-scale Region Perpendicular LBP (MRP-LBP) [399], Scale-and Orientation Adaptive LBP (SOA-LBP) [205]). Furthermore, statistical or fuzzy version of crisp feature could be investigated such as histograms of fuzzy oriented gradients [455]. It would be interesting to evaluate them for this application.

– Because each feature has its strengths and weaknesses against each challenge, multiple features schemes are used to combine the advantages of their different robustness. Most of the time, gradient, texture, motion and stereo features are used in addition to the color feature to deal with camouflage in color, illumination changes, dynamic backgrounds and shadows. Different fusion operators can be used to combine these different features but fuzzy integrals such as the Choquet integral [21] and interval-valued Choquet [93] seem the best way to aggregate different features because dependency between features can be taken into account.

– Because there is not a unique feature that performs better than any other feature independently of the background and foreground properties, feature selection allows to use the best feature or the best combination of features. Experimental results provided by the existing approaches show the pertinence of feature selection in background modeling and foreground detection. However, basic algorithms such as Adaboost and Realboost have been used most of the time. The

most advanced scheme is the IWOC-SVM algorithm developed by Silva et al. [491], but more advanced selection schemes can be used such as statistical or fuzzy feature selection.

To summarize, the most interesting approach seems to fuse multiple features with the intervalued fuzzy Choquet integral. The best set of features seems to beillumination invariant color features combined with spatio-temporal texture features and depth features. Future research should concern **1)** a full evaluation of texture features, **2)** a full comparison of feature fusion schemes, **3)** feature selection schemes and **4)** reliability of features because it has been less investigated. Finally, features learned by deep learning methods such as Stacked Denoising Auto-Encoder (SDAE) [667] and Convolutional Neural Networks (CNN) [60][582]. are surely the features that will outperforms all the other features because deep learning methods have the sole ability of learning features that best fit a given set of data. Furthermore, unlike conventional hand-crafted features, learned features come from multiple layers which focus on various level of details in the video. Thus, learned feature representation allows to well capture the intrinsic structural properties of a scene and adaptively discover a set of filter patterns that are robust to complicated factors such as noise and illumination variation.

References

1. T. Aach and A. Condurache. Transformation of adaptive thresholds by significance invariance for change detection. *International Workshop on Statistical Signal Processing*, July 2005.
2. T. Aach, L. Dumbgen, R. Mester, and D. Toth. Bayesian illumination-invariant motion detection. *IEEE International Conference on Image Processing, ICIP 2001*, 3:640–643, October 2001.
3. T. Aach and A. Kaup. Bayesian algorithms for adaptive change detection in image sequences using Markov random fields. *Signal Processing Image Communication*, 7:147–160, August 1995.
4. T. Aach, A. Kaup, and R. Mester. Change detection in image sequences using Gibbs random fields: a Bayesian approach. *IEEE Workshop Intelligent Signal Processing and Communications Systems*, October 1993.
5. T. Aach, A. Kaup, and R. Mester. Statistical model-based change detection in moving video. *Signal Processing,*, pages 165–180, 1993.
6. T. Aach, D. Toth, and R. Mester. Motion estimation in varying illumination using a total least squares distance measure. *Picture Coding Symposium, PCS 2003*, pages 145–148, April 2003.
7. A. Adam, I. Shimshoni, and E. Rivlin. Aggregated dynamic background modeling. *International Conference on Image Processing, ICIP 2006*, 2006.
8. T. Ahonen, J. Matas, C. He, and M. Pietikainen. Rotation invariant image description with local binary pattern histogram Fourier features. *Scandinavian Conference on Image Analysis, SCIA 2009*, 2009.
9. A. Amamra, T. Mouats, and N. Aouf. GPU based GMM segmentation of Kinect data. *ELMAR 2014*, 2014.
10. A. Amato, M. Mozerov, F. Roca, and J. Gonzalez. Robust real-time background subtraction based on local neighborhood patterns. *EURASIP Journal on Advances in Signal Processing*, 2010, 2010.
11. R. Amith, V. Aradhya, and S. Niranjan. The application of cosine transform and principal components for foreground detection in video. *IEEE International Conference on Green Computing and Internet of Things, ICGCIoT 2015*, October 2015.
12. P. Amnuaykanjanasin, T. Thongkamwitoon, N. Srisawaiwilai, S. Aramvit, and T. Chalidabhongse. Adaptive parametric statistical background subtraction for video segmentation. *ACM International Workshop on Video Surveillance and Sensor Networks, VSNN 2005*, November 2005.
13. B. Antic and V. Crnojevic. Joint domain-range modeling of dynamic scenes with adaptive kernel bandwidth. *Advanced Concepts for Intelligent Vision Systems, ACVIS 2007*, pages 777–788, 2007.

14. B. Antic, V. Crnojevic, and D. Culibrk. Efficient wavelet based detection of moving objects. *International Conference on Digital Signal Processing, DSP 2009*, July 2009.

15. B. Antic, J. Oswaldo, N. Castaneda, D. Culibrk, A. Pizurica, and V. Crnojevic W. Philips. Robust detection and tracking of moving objects in traffic video surveillance. *ACVIS 2009*, October 2009.

16. N. Armanfard, M. Komeili, M. Valizade, and E. Kabir. Effective hierarchical background modeling and foreground detection in surveillance systems. *International CSI Computer Conference, CSICC 2009*, October 2009.

17. M. Azab, H. Shedeed, and A. Hussein. A new technique for background modeling and subtraction for motion detection in real-time videos. *International Conference on Image Processing, ICIP 2010*, pages 3453–3456, September 2010.

18. R. Babu, K. Ramakrishnan, and S. Srinivasan. Video object segmentation: a compressed domain approach. *IEEE Transactions on Circuits and Systems for Video Technology*, 14:452–474, 2004.

19. F. El Baf, T. Bouwmans, and B. Vachon. Comparison of background subtraction methods for a multimedia application. *International Conference on systems, Signals and Image Processing, IWSSIP 2007*, pages 385–388, June 2007.

20. F. El Baf, T. Bouwmans, and B. Vachon. Comparison of background subtraction methods for a multimedia learning space. *International Conference on Signal Processing and Multimedia, SIGMAP 2007*, July 2007.

21. F. El Baf, T. Bouwmans, and B. Vachon. Foreground detection using the Choquet integral. *International Workshop on Image Analysis for Multimedia Interactive Integral, WIAMIS 2008*, pages 187–190, May 2008.

22. F. El Baf, T. Bouwmans, and B. Vachon. A fuzzy approach for background subtraction. *IEEE International Conference on Image Processing, ICIP 2008*, pages 2648–2651, October 2008.

23. F. El Baf, T. Bouwmans, and B. Vachon. Fuzzy foreground detection for infrared videos. *IEEE International Conference on Computer Vision and Pattern Recognition, CVPR-Workshop OTCBVS 2008*, pages 1–6, June 2008.

24. F. El Baf, T. Bouwmans, and B. Vachon. Fuzzy integral for moving object detection. *IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2008*, pages 1729–1736, June 2008.

25. F. El Baf, T. Bouwmans, and B. Vachon. Type-2 fuzzy mixture of Gaussians model: Application to background modeling. *International Symposium on Visual Computing, ISVC 2008*, pages 772–781, December 2008.

26. M. Balcilar, M. Amasyali, and A. Sonmez. Moving object detection using Lab2000HL color space with spatial and temporal smoothing. *International Journal of Applied Mathematics and Information Science*, pages 1755–1766, 2014.

27. M. Balcilar, F. Karabiber, and A. Sonmez. Performance analysis of Lab2000HL color space for background subtraction. *IEEE International Symposium on Innovations in Intelligent Systems and Applications, INISTA 2013*, June 2013.

28. M. Balcilar and A. Sonmez. The effect of color space and block size on foreground detection. *Signal Processing and Communications Applications Conference, SIU 2013*, pages 1–4, April 2013.

29. M. Balcilar and A. Sonmez. Region based fuzzy background subtraction using Choquet integral. *International Conference on Adaptive and Natural Computing Algorithms, ICANNGA 2013*, pages 298–296, April 2013.

30. D. Baltieri, R. Cucchiara, and R. Vezzani. Fast background initialization with recursive Hadamard transform. *International Conference on Advanced Video and Signal Based Surveillance, AVSS 2010*, 2010.

31. L. Bao, Q. Yang, and H. Jin. Fast edge-preserving patch match for large displacement optical flow. *International Conference on Computer Vision and Pattern Recogntion, CVPR 2014*, 2014.

32. O. Barnich and M. Van Droogenbroeck. ViBe: a powerful random technique to estimate the background in video sequences. *International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2009*, pages 945–948, April 2009.

33. C. Bautista, C. Dy, M. Manalac, R. Orbe, and M. Cordel. Convolutional neural network for vehicle detection in low resolution traffic videos. *TENSYMP 2016*, 2016.

34. S. Becker, N. Scherer-Negenborn, P. Thakkar, W. Hubner, and M. Arens. The effects of camera jitter for background subtraction algorithms on fused infrared-visible video streams. *SPIE 9995*, October 2016.

35. Y. Benezeth, D. Sidibe, and J. Thomas. Background subtraction with multispectral video sequences. *IEEE International Conference on Robotics and Automation, ICRA 2014*, 2014.

36. A. Bevilacqua, G. Capelli, L. Di Stefano, and A. Lanza. A novel approach to change detection based on a coarse-to-fine strategy. *IEEE International Conference on Image Processing, ICIP 2005*, pages 434–437, 2005.

37. A. Bevilacqua, L. Di Stefano, and A. Lanza. Coarse-to-fine strategy for robust and efficient change detectors. *IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2005*, pages 87–92, September 2005.

38. J. Bezdek. Pattern recognition with fuzzy objective function algorithms. *Kluwer Academic Publishers Norwell*, 1981.

39. B. Bhanu and S. Nadimi. Evolutionary sensor fusion for security. *Optical Imaging Sensors and Systems for Homeland Security Applications*, 2:245–269, 2006.

40. H. Bhaskar, L. Mihaylova, and A. Achim. Object detection based on adaptive background subtraction using alpha stable distribution. *Institution of Engineering and Technology Conference on Target Tracking and Data Fusion*, April 2008.

41. H. Bhaskar, L. Mihaylova, and A. Achim. Video foreground detection based on symmetric alpha-stable mixture models. *IEEE Transactions on Circuits, Systems and Video Technology*, March 2010.

42. H. Bhaskar, L. Mihaylova, and S. Maskell. Background modeling using adaptive cluster density estimation. *German Workshop on Sensor Data Fusion: Trends, Solutions, Applications*, September 2007.

43. G. Bilodeau, J. Jodoin, and N. Saunier. Change detection in feature space using local binary similarity patterns. *Conference on Computer and Robot Vision, CRV 2013*, pages 106–112, May 2013.

44. Z. Bin and Y. Liu. Robust moving object detection and shadow removing based on improved Gaussian model and gradient information. *International Conference on Multimedia Technology, ICMT 2010*, October 2010.

45. C. Blanco, T. Mantecon, M. Camplani, F. Jaureguizar, L. Salgado, and N. Garca. Foreground segmentation in depth imagery using depth and spatial dynamic models for video surveillance applications. *Sensors 2014*, 2014.

46. D. Bloisi, A. Pennisi, and L. Iocchi. Background modeling in the maritime domain. *Machine Vision and Applications, Special Issue on Background Modeling*, 2013.

47. C. Blum, J. Jacques, G. Cavalheiro, H. Geraldo, C. Jung, and S. Musse. An improved background subtraction algorithm and concurrent implementations. *Parallel Processing Letters*, 20(1):71–89, 2010.

48. T. Bouwmans. Subspace Learning for Background Modeling: A Survey. *Recent Patents on Computer Science, RPCS 2009*, 2(3):223–234, November 2009.

49. T. Bouwmans. Recent Advanced Statistical Background Modeling for Foreground Detection: A Systematic Survey. *Recent Patents on Computer Science, RPCS 2011*, 4(3):147–176, September 2011.

50. T. Bouwmans. Background Subtraction For Visual Surveillance: A Fuzzy Approach. *Chapter 5, Handbook on Soft Computing for Video Surveillance, Taylor and Francis Group, S.K. Pal, A. Petrosino, L. Maddalena*, pages 103–139, March 2012.

51. T. Bouwmans. Traditional and recent approaches in background modeling for foreground detection: An overview. *Computer Science Review*, 11:31–66, May 2014.

52. T. Bouwmans, F. El Baf, and B. Vachon. Background Modeling using Mixture of Gaussians for Foreground Detection - A Survey. *Recent Patents on Computer Science, RPCS 2008*, 1(3):219–237, November 2008.

53. T. Bouwmans, F. El Baf, and B. Vachon. Statistical Background Modeling for Foreground Detection: A Survey. *Part 2, Chapter 3, Handbook of Pattern Recognition and Computer Vision, World Scientific Publishing, Prof C.H. Chen*, 4:181–199, January 2010.

54. T. Bouwmans, J. Gonzalez, C. Shan, M. Piccardi, and L. Davis. Special issue on background modeling for foreground detection in real-world dynamic scenes. *Special Issue in Machine Vision and Applications*, July 2014.

55. T. Bouwmans, F. Porikli, B. Hoferlin, and A. Vacavant. Handbook on background modeling and foreground detection for video surveillance. *Chapman and Hall/CRC*, July 2014.

56. T. Bouwmans, A. Sobral, S. Javed, S. Jung, and E. Zahzah. Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset. *Computer Science Review*, 2016.

57. T. Bouwmans and E. Zahzah. Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance. *Special Isssue on Background Models Challenge, Computer Vision and Image Understanding, CVIU 2014*, 122:22–34, May 2014.

58. F. Bovolo, G. Camps-Valls, and L. Bruzzone. An unsupervised support vector method for change detection. *SPIE*, 6748:110, 2007.

59. M. Braham and M. Van Droogenbroeck. A generic feature selection method for background subtraction using global foreground models. *Advanced Concepts for Intelligent Vision Systems, ACIVS 2015*, 2015.

60. M. Braham and M. Van Droogenbroeck. Deep background subtraction with scene-specific convolutional neural networks. *International Conference on Systems, Signals and Image Processing, IWSSIP 2016*, May 2016.

61. S. Brutzer, B. Höferlin, and G. Heidemann. Evaluation of background subtraction techniques for video surveillance. *International Conference on Computer Vision and Pattern Recognition, CVPR 2011*, pages 1937–1944, June 2011.

62. M. Camplani, C. Blanco, L. Salgado, F. Jaureguizar, and N. Garca. Multi-sensor background subtraction by fusing multiple region-based probabilistic classifiers. *Pattern Recognition Letters*, 2013.

63. M. Camplani, C. Blanco, L. Salgado, F. Jaureguizar, and N. Garca. Advanced background modeling with RGB-D sensors through classifiers combination and inter-frame foreground prediction. *Machine Vision and Applications*, 2014.

64. M. Camplani, A. Blasco, D. Berjon, L. Salgado, and F. Moran. Real-time RGB-D data processing on GPU architecture. *IEEE International on Design and Architectures for Signal and Image Processing, DASIP 2013*, October 2013.

65. M. Camplani, T. Mantecon, and L. Salgado. Depth-color fusion strategy for 3D scene modeling with kinect. *IEEE Transaction on Cybernetics*, 2013.

66. M. Camplani and L. Salgado. Background foreground segmentation with RGB-D Kinect data: an efficient combination of classifiers. *Journal on Visual Communication and Image Representation*, 2013.

67. E. Candes, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *International Journal of ACM*, 58(3), May 2011.

68. J. Carranza, C. Theobalt, M. Magnor, and H. Seidel. Free-viewpoint video of human actors. *ACM Transactions on Graphics*, 22(3):569–577, 2003.

69. V. Cevher, M. Duarte, C. Hedge, and R. Baraniuk. Sparse signal recovery using markov random fields. *Annual Conference on Neural Information Processing Systems, NIPS 2008*, 2008.

70. V. Cevher, D. Reddy, M. Duarte, A. Sankaranarayanan, R. Chellappa, and R. Baraniuk. Compressive sensing for background subtraction. *European Conference on Computer Vision, ECCV 2008*, October 2008.

71. S. Chakraborty, M. Paul, M. Murshed, and M. Ali. An efficient video coding technique using a novel non-parametric background model. *IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2014*, pages 1–6, July 2014.

72. S. Chakraborty, M. Paul, M. Murshed, and M. Ali. A novel video coding scheme using a scene adaptive non-parametric background model. *IEEE International Conference on Multimedia Signal Processing*, pages 1–6, September 2014.

73. S. Chakraborty, M. Paul, M. Murshed, and M. Ali. Adaptive weighted non-parametric background model for efficient video coding. *Neuro Computing*, 2016.

74. A. Chan, V. Mahadevan, and N. Vasconcelos. Generalized Stauffer-Grimson background subtraction for dynamic scenes. *Machine Vision and Applications*, 22(5):751–766, September 2011.

75. K. Chan. Background modeling using perception-based local pattern. *Conference on Computer Graphics, Visualization and Computer Vision, WSCG 2016*, 2016.

76. M. Chang and Y. Cheng. Motion detection by using entropy image and adaptive state-labeling technique. *IEEE International Symposium on Circuits and Systems, ISCAS 2007*, pages 3667–3670, 2007.

77. J. Chen, S. Shan, C. He, M. Pietikainen, X. Chen, and W. Gao. WLD: A Robust Local Image Descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1705–1720, September 2009.

78. M. Chen, Q. Yang, Q. Li, G. Wang, and M. Yang. Spatiotemporal background subtraction using minimum spanning tree and optical flow. *European Conference on Computer Vision, ECCV 2014*, 2014.

79. M. Chen, Q. Yang, Q. Li, G. Wang, and M. Yang. Spatiotemporal background subtraction. *Submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.

80. S. Chen, T. Xu, D. Li, J. Zhang, and S. Jiang. Moving object detection using scanning camera on a high-precision intelligent holder. *MDPI Sensors 2016*, October 2016.

81. Y. Chen, C. Chen, C. Huang, and Y. Hung. Efficient hierarchical method for background subtraction. *Pattern Recognition*, 41(10):2706–2715, October 2007.

82. Z. Chen and T. Ellis. Self-adaptive Gaussian mixture model for urban traffic monitoring system,. *IEEE International Conference on Computer Vision Workshops,ICCV 2011*, pages 1769–1776, 2011.

83. F. Cheng, B. Chen, and S. Huang. A background model re-initialization method based on sudden luminance change detection. *Engineering Applications of Artificial Intelligence*, 38:138–146, 2015.

84. H. Cheng. Efficient sparse representation and modeling. *Sparse Representation, Modeling and Learning in Visual Recognition*, pages 117–151, 2015.

85. S. Cheng, X. Luo, and S. Bhandarkar. A multiscale parametric background model for stationary foreground object detection. *International Workshop on Motion and Video Computing, WMCV 2007*, 2007.

86. S. Cheung and C. Kamath. Robust background subtraction with foreground validation for urban traffic video. *Journal of Applied Signal Processing, EURASIP*, 2005.

87. P. Chiranjeevi and S. Sengupta. Detection of moving objects using fuzzy correlogram based background subtraction. *International Conference on Signal and Image Processing Applications, ICSIPA 2011*, pages 255–259, November 2011.

88. P. Chiranjeevi and S. Sengupta. Moving object detection in the presence of dynamic backgrounds using intensity and textural features. *SPIE Journal of Electronic Imaging*, 20(4):111, 2011.

89. P. Chiranjeevi and S. Sengupta. New fuzzy texture features for robust detection of moving objects. *IEEE Signal Processing Letters*, 19(10):603–606, October 2012.

90. P. Chiranjeevi and S. Sengupta. Robust detection of moving objects in video sequences through rough set theory framework. *Image and Vision Computing, IVC 2012*, 30(11):829–842, 2012.

91. P. Chiranjeevi and S. Sengupta. Detection of moving objects using multi-channel kernel fuzzy correlogram based background subtraction. *IEEE Transactions on Systems, Man, and Cybernetics*, pages 870–881, October 2013.

92. P. Chiranjeevi and S. Sengupta. Neighborhood supported model level fuzzy aggregation for moving object segmentation. *IEEE Transactions on Image Processing*, 23(2):645–657, February 2014.

93. P. Chiranjeevi and S. Sengupta. Interval-valued model level fuzzy aggregation-based background subtraction. *IEEE Transactions on Cybernetics*, 2016.

94. S. Choi and J. Han. Fine-level moving object detection using region-based background/foreground model. *International Conference on ICT Convergence*, pages 761–762, October 2012.

95. T. Chua, K. Leman, and Y. Wang. Adaptive texture-color based background subtraction for video surveillance. *International Conference on Image Processing, ICIP 2012*, pages 49–52, October 2012.

96. T. Chua, K. Leman, and Y. Wang. Fuzzy rule-based system for dynamic texture and color based background subtraction. *IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2012*, pages 1–7, June 2012.

97. G. Cocorullo, P. Corsonello, F. Frustaci, and S. Perri. Multimodal background subtraction for high-performance embedded systems. *Journal of Real-Time Image Processing*, pages 1–17, 2016.

98. C. Conaire, E. Cooke, N. O'Connor, N. Murphy, and A. Smeaton. Background modeling in infrared and visible spectrum video for people tracking. *International Conference on Computer Vision and Pattern Recognition, CVPR 2005*, June 2005.

99. C. Conaire, E. Cooke, N. O'Connor, N. Murphy, and A. Smeaton. Fusion of infrared and visible spectrum video for indoor surveillance. *International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2005*, April 2005.

100. C. Conaire, N. O'Connor, E. Cooke, and A. Smeaton. Multispectral object segmentation and retrieval in surveillance video. *IEEE International Conference on Image Processing, ICIP 2006*, pages 2381–2384, 2006.

101. M. Cristani, M. Bicego, and V. Murino. Integrated region- and pixel-based approach to background modelling. *IEEE Workshop. on Motion and Video Computing*, December 2002.

102. M. Cristani and V. Murino. A spatial sampling mechanism for effective background subtraction. *International Conference on Computer Vision Theory and Applications, VISAPP 2007*, 2:403–410, March 2007.

103. M. Cristani and V. Murino. Background subtraction with adaptive spatio-temporal neighborhood analysis. *International Conference on Computer Vision Theory and Applications, VISAPP 2008*, January 2008.

104. V. Crnojevic, B. Antic, and D. Culibrk. Optimal wavelet differencing method for robust motion detection. *International Conference on Image Processing, ICIP 2009*, November 2009.

105. C. Cuevas and N. Garcia. Tracking-based non-parametric background-foreground classification in a chromaticity-gradient space. *International Conference on Image Processing, ICIP 2010*, September 2010.

106. D. Culibrk, V. Crnojevic, and B. Antic. Multiscale background modelling and segmentation. *International conference on Digital Signal Processing, DSP 2009*, pages 922–927, 2009.

107. B. Cyganek and M. Wozniak. Efficient real-time background detection based on the pca subspace decomposition. *International Conference on Artificial Intelligence and Soft Computing, ICAISC 2017*, June 2017.

108. K. Dai, Q. Li, and Z. Zou. Color comparison in codebook model for moving objects detection. *International Conference Frontier of Computer Science and Technology, FCST 2010*, pages 518–121, August 2010.

109. N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *International Conference on Computer Vision and Pattern Recognition, CVPR 2005*, 1:886–893, 2005.

110. G. Dalley, J. Migdal, and W. Grimson. Background subtraction for temporally irregular dynamic textures. *IEEE Workshop on Applications of Computer Vision, WACV 2008*, pages 1–7, January 2008.

111. S. Davarpanah, F. Khalid, L. Abdullah, and M. Golchin. A texture descriptor: Background local binary pattern (BGLBP). *Multimed Tools Applications*, 2015.

112. S. Davarpanah, F. Khalid, and M. Golchin. A block-based multi-scale background extraction algorithm. *Journal of Computer Science*, pages 1445–1451, 2010.

113. S. Davarpanah, F. Khalid, N. Lili, S. Puteri, and M. Golchin. Using multi-scale filtering to initialize a background extraction model. *Journal of Computer Science*, pages 1077–1084, 2012.

114. R. Davies, L. Mihaylova, N. Pavlidis, and I. Eckley. The effect of recovery algorithms on compressive sensing background subtraction. *Workshop Sensor Data Fusion: Trends, Solutions, and Applications*, 2013.

115. J. Davis and M. Keck. A two-stage template approach to person detection in thermal imagery. *Workshop on Applications of Computer Vision*, January 2005.

116. J. Davis and V. Sharma. Robust background-subtraction for person detection in thermal imagery. *IEEE International Workshop on Object Tracking and Classification Beyond the Visible Spectrum*, July 2004.

117. J. Davis and V. Sharma. Robust detection of people in thermal imagery. *International Conference on Pattern Recognition*, pages 713–716, August 2004.

118. J. Davis and V. Sharma. Fusion-based background-subtraction using contour saliency. *IEEE Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum*, June 2005.

119. J. Davis and V. Sharma. Background-subtraction in thermal imagery using contour saliency. *International Journal of Computer Vision*, 2006.

120. J. Davis and V. Sharma. Background-subtraction in thermal imagery using contour saliency. *Journal of Computer Vision and Image Understanding*, 71(2):161–181, 2007.

121. J. Davis and V. Sharma. Background-subtraction using contour-based fusion of thermal and visible imagery. *Journal of Computer Vision and Image Understanding*, 2007.

122. K. Delibasis, T. Goudas, and I. Maglogiannis. A novel robust approach for handling illumination changes in video segmentation. *Engineering Applications of Artificial Intelligence*, 49:43–60, 2016.

123. X. Deng, J. Bu, Z. Yang, C. Chen, and Y. Liu. A block-based background model for video surveillance. *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2008*, pages 1013–1016, March 2008.

124. B. Dey and M. Kundu. Robust background subtraction for network surveillance in H.264 streaming video. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(10):1695–1703, October 2013.

125. B. Dey and M. Kundu. Enhanced macroblock features for dynamic background modeling in H.264/AVC video encoded at low-bitrate. *IEEE Transactions on Circuits and Systems for Video Technology*, 2016.

126. P. Dickinson, A. Hunter, and K. Appiah. Segmenting video foreground using a multi-class MRF. *International Conference on Pattern Recognition, ICPR 2010*, pages 1848–1851, August 2010.

127. Y. Ding, W. Li, J. Fan, and H. Yang. A fuzzy background model for moving object detection. *IEEE International Conference on Computer-Aided Design and Computer Graphics, CAD/Graphics 2009*, pages 610–613, 2009.

128. Y. Ding, W. Li, J. Fan, and H. Yang. Robust motion detection with color-texture feature fuzzy approach. *International Conference of COMPUTING in Engineering, Science and Informatics, ICC 2009*, April 2009.

129. Y. Ding, W. Li, J. Fan, and H. Yang. A moving object detection algorithm based on Choquet integral. *Chinese Journal of Electronics*, 38(2):263–268, February 2010.

130. Y. Ding, W. Li, T. Fan, and H. Yang. Robust moving object detection under complex background. *Computer Science and Information Systems*, 7(1), February 2010.

131. A. Doshi and M. Trivedi. Hybrid cone-cylinder codebook model for foreground detection with shadow and highlight suppression. *AVSS 2006*, November 2006.

132. A. Doshi and M. Trivedi. Satellite imagery based adaptive background models and shadow suppression. *Journal Signal, Image and Video Processing, SIVP 2007*, pages 119–132, June 2007.

133. J. Dou and J. Li. Modeling the background and detecting moving objects based on SIFT flow. *Optik - International Journal for Light and Electron Optics*, 125(1):435–440, January 2014.

134. X. Du and G. Qin. Foreground detection in surveillance videos via a hybrid local texture based method. *International journal on smart sensing and intelligentsystems*, 9(4), December 2016.

135. R. Duda, P. Hart, and D. Stork. Pattern classification. *New York: Wiley-Interscience*, 2001.

136. E. Durucan and T. Ebrahimi. Robust and illumination invariant change detection based on linear dependence. *European Signal Processing Conference, EUSIPCO 2000*, pages 1141–1144, September 2000.

137. E. Durucan and T. Ebrahimi. Change detection and background extraction by linear algebra. *Special Issue of IEEE on Video Communications and Processing for Third Generation Surveillance*, October 2001.

138. E. Durucan and T. Ebrahimi. Change detection by nonlinear gramian. *Annual Conference on Neural Information Processing Systems, NISP 2001*, June 2001.

139. E. Durucan and T. Ebrahimi. Moving object detection between multiple and color images. *IEEE Conference on Advanced Video and Signal Based Surveillance, AVSS 2003*, July 2003.

140. S. Erfanian Ebadi and E. Izquierdo. Foreground detection with dynamic tree-structured sparse RPCA. *submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.

141. S. Erfanian Ebadi and E. Izquierdo. Foreground segmentation via dynamic tree-structured sparse RPCA. *European Conference on Computer Vision, ECCV 2016*, 2016.

142. A. Elgammal and L. Davis. Non-parametric model for background subtraction. *European Conference on Computer Vision, ECCV 2000*, pages 751–767, June 2000.

143. O. Elharrouss, H. Tairi, and D. Moujahid. Motion detection based on the combining of the background subtraction and the structure-texture decomposition. *Optik - International Journal for Light and Electron Optics*, 2015.

144. C. Eveland, K. Konolige, and R. Bolles. Background modeling for segmentation of video-rate stereo sequences. *IEEE Computer Vision and Pattern Recognition, CVPR 1998*, pages 266–271, June 1998.

145. T. Fabian. Mixture of gaussians exploiting histograms of oriented gradients for background subtraction. *International Symposium on Visual Computing, ISVC 2010*, November 2010.

146. H. Fan, J. Guo, and J. Li. Patchnet-based background subtraction algorithm for dynamic scenes video. *ICIC Express Letters*, 9(4):1101–1107, January 2015.

147. X. Fang, W. Xiong, B. Hu, and L. Wang. A moving object detection algorithm based on color information. *International Symposium on Instrumentation Science and Technology*, 48:384–387, 2006.

148. E. Fernandez-Sanchez, J. Diaz, and E. Ros. Background subtraction based on color and depth using active sensors. *Sensors 2013*, 13:8895–8915, 2013.

149. E. Fernandez-Sanchez, L. Rubio, J. Diaz, and E. Ros. Background subtraction model based on color and depth cues. *Machine Vision and Applications*, 2014.

150. Y. Freund and R. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, January 1997.

151. S. Fukui, T. Ishikawa, Y. Iwahori, and H. Itou. Extraction of moving objects by estimating background brightness. *IAPR Workshop on Machine Vision Applications*, December 2002.

152. S. Fukui, T. Ishikawa, Y. Iwahori, and H. Itou. Extraction of moving objects by estimating background brightness. *Journal of the Institue of Image Electronics Engineers of Japan*, 33(3):350–357, 2004.

153. J. Gallego and M. Pardas. Region based foreground segmentation combining color and depth sensors via logarithmic opinion pool decisions. *Journal of Visual Communication and Image Representation*, April 2013.

154. C. Gan, Y. Wang, and X. Wang. Multi-feature robust principal component analysis for video moving object segmentation. *Journal of Image and Graphics*, 18(9), 2013.

155. B. Gao, T. Liu, Q. Cheng, and W. Ma. A linear approximation based method for noise-robust and illumination-invariant image change detection. *Pacific Rim conference on Advances in Multimedia Information Processing, PCM 2004*, pages 95–102, 2004.

156. D. Gao, Z. Jiang, and M. Ye. A new approach of dynamic background modeling for surveillance information. *International Conference on Computer Science and Software Engineering, CSSE 2008*, 1, 2008.

157. L. Gao, Y. Fan, N. Chen, Y. Li, and X. Li. Moving objects detection using adaptive region-based background model in dynamic scenes. *Foundations of Intelligent Systems, Advances in Intelligent and Soft Computing*, 122:641–651, 2012.

158. T. Gao, Z. Liu, W. Gao, and J. Zhang. A robust technique for background subtraction in traffic video. *International Conference on Neural Information Processing, ICONIP 2008*, pages 736–744, November 2008.

159. T. Gao, Z. Liu, W. Gao, and J. Zhang. A robust technique for background subtraction and shadow elimination in traffic video sequence. *Next-Generation Applied Intelligence*, pages 311–322, 2009.

160. T. Gao, Z. Liu, S. Yue, J. Mei, and J. Zhang. Traffic video-based moving vehicle detection and tracking in the complex environment. *Journal Cybernetics and Systems*, 40(7):569–588, October 2009.

161. T. Gao, Z. Liu, S. Yue, J. Zhang, J. Mei, and W. Gao. Robust background subtraction in traffic video sequence. *Journal of Central South University of Technology*, 17, February 2010.

162. A. Gayathri and A. Srinivasan. Moving object detection by fuzzy aggregation using low rank weightage representation. *International Conference on Frontiers of Intelligent Computing: Theory and Applications, FICTA 2014*, pages 335–342, 2014.

163. L. Geng and Z. Xiao. Real time foreground-background segmentation using two-layer codebook model. *International Conference on Control, Automation and Systems Engineering, CASE 2011*, pages 1–5, 2011.

164. D. Giveki, G. Montazer, and M. Soltanshahi. Atanassovs intuitionistic fuzzy histon for robust moving object detection. *International Journal of Approximate Reasoning*, September 2017.

165. M. Gong and L. Cheng. Incorporating estimated motion in real-time background subtraction. *IEEE International Conference on Image Processing, ICIP 2011*, pages 3265–3268, September 2011.

166. G. Gonzalez and R. Woods. Digital image processing. *Prentice Hall, 2nd edition*, 2002.

167. G. Gordon, T. Darrell, M. Harville, and J. Woodfill. Background estimation and removal based on range and color. *International Conference on Computer Vision and Pattern Recognition, CVPR 1999*, pages 459–464, June 1999.

168. K. Goyal and J. Singhai. Texture-based self-adaptive moving object detection technique for complex scenes. *Computers and Electrical Engineering*, pages 1–9, 2016.

169. N. Goyette, P. Jodoin, F. Porikli, J. Konrad, and P. Ishwar. changedetection.net: A new change detection benchmark dataset. *IEEE Workshop on Change Detection, CDW 2012 at CVPR 2012*, June 2012.

170. H. Grabner and H. Bischof. On-line boosting and vision. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2006*, 2006.

171. H. Grabner, C. Leistner, and H. Bischof. Time dependent on-line boosting for robust background modeling. *International Conference on Computer Vision Theory and Applications, VISAPP 2008*, January 2008.

172. H. Grabner, P. Roth, M. Grabner, and H. Bischof. Autonomous learning a robust background model for change detection. *PETS 2006*, 2006.

173. K. Greff, A. Brandao, S. Krauss, D. Stricker, and E. Clua. A comparison between background subtraction algorithms using a consumer depth camera. *International Conference on Computer Vision Theory and Applications*, 2012.

174. J. Gu, Z. Liu, and Z. Zhang. Novel moving object segmentation algorithm using kernel density estimation and edge information. *Journal of Computer-Aided Design and Computer Graphics*, 21(2):223–228, February 2009.

175. Y. Guan. Wavelet multi-scale transform based foreground segmentation and shadow elimination. *Open Signal Processing Journal*, 1:1–6, 2008.

176. Y. Guan. Wavelet multi-scale transform based foreground segmentation and shadow elimination. *Open Signal Processing Journal*, 1:1–6, 2008.

177. Y. Guan. Adaptive motion segmentation for changing background. *Journal of Software Engineering and Applications*, 2:96–102, July 2009.

178. Y. Guan. Motion objects segmentation and shadow suppressing without background learning. *Journal of Engineering*, 2014, 2014.

179. Y. Guan, X. Cheng, and X. Jia. Motion foreground detection based on wavelet transformation and color ratio difference. *International Congress on Image and Signal Processing, CISP 2010*, pages 1423–1426, October 2010.

180. P. Guha, D. Palai, K. Venkatesh, and A. Mukerjee. A multiscale co-linearity statistic based approach to robust background modeling. *Asian Conference on Computer Vision, ACCV 2006*, pages 297–306, 2006.

181. C. Guo and J. Yu. Steering kernel-based video moving objects detection with local background texture dictionaries. *Computers and Electrical Engineering*, 2012.

182. J. Guo, C. Hsia, M. Shih, Y. Liu, and J. Wu. High speed multi-layer background subtraction. *International Symposium on Intelligent Signal Processing and Communications Systems, ISPACS 2012*, pages 74–79, 2012.

183. J. Guo and C. Hsu. Cascaded background subtraction using block-based and pixel-based codebooks. *International Conference on Pattern Recognition, ICPR 2010*, August 2010.

184. J. Guo and C. Hsu. Hierarchical method for foreground detection using codebook model. *IPPR Conference on Computer Vision, Graphics and Image Processing, CVGIP 2010*, August 2010.

185. J. Guo and C. Hsu. Hierarchical method for foreground detection using codebook model. *International Conference on Image Processing, ICIP 2010*, pages 3441–3444, September 2010.

186. J. Guo, Y. Liu, C. Hsia, M. Shih, and C. Hsu. Efficient hierarchical method for background subtraction. *IEEE Transactions on Circuits and Systems for Video Technology*, 2011.

187. L. Guo, D. Xu, and Z. Qiang. Background Subtraction using Local SVD Binary Pattern. *International Conference on Computer Vision and Pattern Recognition, CVPR 2016*, June 2016.

188. Y. Guo, G. Zhao, and M. Pietikainen. Texture classification using a linear configuration model based descriptor. *British Machine Vision Conference, BMVC 2011*, pages 1–10, 2011.

189. Y. Guo, G. Zhao, and M. Pietikainen. Discriminative features for texture description. *Pattern Recognition*, 45(10):3834–3843, 2012.

190. V. Gutti and C. Shankar. A novel approach to background subtraction using fuzzy color histogram. *Journal of Advanced Engineering Technologies*, 3(3):231–239, 2014.

191. C. Guyon, T. Bouwmans, and E. Zahzah. Robust Principal Component Analysis for Background Subtraction: Systematic Evaluation and Comparative Analysis. *Book 1, Chapter 12, Principal Component Analysis, INTECH*, pages 223–238, March 2012.

192. H. Haberdar. Video-based change detection. *PhD Thesis, Department of Computer Science, University of Houston*, 2013.

193. H. Haberdar and S. Shah. Change detection in dynamic scenes using local adaptive transform. *British Machine Vision Conference, BMVC 2013*, 2013.

194. T. Haines and T. Xiang. Background subtraction with Dirichlet processes. *European Conference on Computer Vision, ECCV 2012*, pages 99–113, September 2012.

195. B. Han and L. Davis. Density-based multi-feature background subtraction with support vector machine. *PAMI 2011*, 2011.

196. G. Han, X. Cai, and J. Wang. Object detection based on combination of visible and thermal videos using a joint sample consensus background model. *Journal of software*, pages 987–334, April 2013.

197. G. Han, J. Wang, and X. Cai. Background subtraction based on three-dimensional discrete wavelet transform. *Sensors 2016*, 16(4), 2016.

198. J. Han and K. Ma. Fuzzy color histogram and its use in color image retrieval. *IEEE Transactions on Image Processing*, 11(8):944–952, November 2002.

199. R. Haralick, B. Shanmugam, and I. Dinstein. Textural features for image classifications. *IEEE Transactions on Systems, Man and Cybernetics*, 23(3):610–622, 1973.

200. M. Harville. A framework for high-level feedback to adaptive, per-pixel, mixture-of-gaussian background models. *European Conference on Computer Vision, ECCV 2002*, May 2002.

201. M. Harville, G. Gordon, and J. Woodfill. Adaptive background subtraction using color and depth. *IEEE International Conference on Image Processing, ICIP 2001*, October 2001.

202. M. Harville, G. Gordon, and J. Woodfill. Foreground segmentation using adaptive mixture models in color and depth. *International Workshop on Detection and Recognition of Events in Video*, July 2001.

203. K. Hati, P. Sa, and B. Majh. Intensity range based background subtraction for effective object detection. *IEEE Signal Processing Letters*, 2013.

204. J. He, M. Gao, L. Zhang, and H. Wu. Sparse signal recovery from fixed low-rank subspace via compressive measurement. *Algorithms 2013*, 6(4):871–882, June 2008.

205. S. Hegenbart and A. Uhl. A scale- and orientation-adaptive extension of local binary patterns for texture classification. *Pattern Recognition*, page 26332644, 2015.

206. M. Heikkila and M. Pietikainen. A texture-based method for modeling the background and detecting moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI 2006*, 28(4):657–62, 2006.

207. M. Heikkila, M. Pietikainen, and J. Heikkila. A texture-based method for detecting moving objects. *BMVC 2004*, September 2004.

208. M. Heikkila, M. Pietikainen, and C. Schmid. Description of interest regions with center-symmetric local binary patterns. *Conference on Computer Vision, Graphics and Image Processing*, pages 58–69, 2006.

209. M. Hofmann, P. Tiefenbacher, and G. Rigoll. Background segmentation with feedback: The pixel-based adaptive segmenter. *IEEE Workshop on Change Detection, CVPR 2012*, June 2012.

210. P. Holtzhausen, V. Crnojevic, and B. Herbst. An illumination invariant framework for real-time foreground detection. *Journal of Real Time Image Processing*, November 2012.

211. S. Hong, J. Kim, K. Hee, A. Ramirez-Rivera, G. Song, K. Hee, O. Chae, and K. Hee. Edge shape pattern for background modeling based on hybrid local codes. *International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2016*, 2016.

212. T. Horprasert, D. Harwood, and L. Davis. A statistical approach for real-time robust background subtraction and shadow detection. *IEEE International Conference on Computer Vision, FRAME-RATE Workshop*, September 1999.

213. T. Horprasert, D. Harwood, and L. Davis. A robust background subtraction and shadow detection. *Asian Conference on Computer Vision, ACCV 2000*, January 2000.

214. M. Hossain, M. Dewan, and O. Chae. Moving object detection for real time video surveillance: An edge based approach. *IEICE Transactions*, 2007.

215. C. Hsia and J. Guo. Efficient modified directional lifting-based discrete wavelet transform for moving object detection. *Signal Processing*, 96:138–152, 2014.

216. H. Hu, L. Xu, and H. Zhao. A spherical codebook in YUV color space for moving object detection. *Sensor Letters*, 10(1):177–189, January 2012.

217. L. Hu, L. Duan, X. Zhang, and J. Yang. Moving object detection based on the fusion of color and depth information. *Journal of Electronics and Information Technology*, 36(9):2047–2051, September 2014.

218. L. Hu, W. Liu, B. Li, and W. Xing. Robust motion detection using histogram of oriented gradients for illumination variations. *International Conference on Industrial Mechatronics and Automation, ICIMA 2010*, 2:443–447, May 2010.

219. M. Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, pages 179–187, 1962.

220. G. Huang and C. Huang. Binary invariant cross color descriptor using galaxy sampling. *International Conference on Pattern Recognition, ICPR 2012*, pages 2610–2613, November 2012.

221. J. Huang and W. Hsieh. Wavelet based moving object segmentation. *IET Electronics Letters*, 39(19):1380–1382, 2003.

222. J. Huang, W. Jin, D. Zhao, N. Qin, and Q. Li. Double-trapezium cylinder codebook model based on yuv color model for foreground detection with shadow and highlight suppression. *Journal of Signal Processing Systems*, pages 1–13, 2015.

223. J. Huang, T. Su, L. Wang, and W. Hsieh. Double-change-detection method for wavelet-based moving-object segmentation. *IET Electronics Letters*, 40(13):798–199, 2004.

224. J. Huang, H. Wu, Y. Gong, and D. Gao. Random Sampling-based Background Subtraction with Adaptive Multi-cue Fusion in RGBD Videos. *International Congress on Image and Signal Processing, CISP 2016*, 2016.

225. S. Huang, L. Fu, and P. Hsiao. A region-level motion-based background modeling and subtraction using MRFs. *IEEE International Conference on Robotics and Automation, ICRA 2005*, April 2005.

226. S. Huang, L. Fu, and P. Hsiao. Region-level motion-based foreground detection with shadow removal using MRFs. *IPPR Conference on Computer Vision, Graphics and Image Processing, CVGIP 2005*, 2005.

227. S. Huang, L. Fu, and P. Hsiao. Region-level motion-based foreground detection with shadow removal using MRFs. *Asian Conference on Computer Vision, ACCV 2006*, 2006.

228. S. Huang, L. Fu, and P. Hsiao. A Bayesian network for foreground segmentation in region level. *Asian Conference on Computer Vision, ACCV 2007*, pages 124–133, 2007.

229. S. Huang, L. Fu, and P. Hsiao. Region-level motion-based background modelling and subtraction using MRFs. *IEEE Transactions on Image Processing*, 16(5):1446–1456, May 2007.

230. S. Huang, L. Fu, and P. Hsiao. Region-level motion-based foreground segmentation under a Bayesian network. *IEEE Transactions on Circuits and Systems for Video Technology*, 19(4):522–532, April 2009.

231. S. Huang, P. Hsiao, and L. Fu. A Bayesian framework for foreground segmentation. *IEEE International Conference on Systems, Man and Cybernetics*, 2006.

232. I. Huerta, A. Amato, X. Roca, and J. Gonzalez. Exploiting multiple cues in motion segmentation based on background subtraction. *Neurocomputing, Special issue: Behaviours in video*, pages 183–196, January 2013.

233. I. Huerta, D. Rowe, M. Mozerov, and J. Gonzalez. Improving background subtraction based on a casuistry of colour-motion segmentation problems. *Iberian Conference on Pattern Recognition and Image Analysis, IbPRIA 2007*, 2007.

234. I. Huerta, D. Rowe, M. Vinas, M. Mozerov, and J. Gonzalez. Background subtraction fusing colour intensity and edge cues. *Conference on Articulated Motion and Deformable Objects, AMDO 2008*, pages 279–288, 2008.

235. M. Itoh, M. Kazui, and H. Fujii. Robust object detection based on radial reach correlation and adaptive background estimation for real-time video surveillance systems. *Real Time image Processing, SPIE 2008*, 2008.

236. M. Itoh, M. Kazui, and H. Fujii. Robust object detection and adaptive background estimation based on radial reach correlation. *Korea-Japan joint workshop on Frontiers of Computer Vision, FCV 2015*, January 2015.

237. Y. Ivanov, A. Bobick, and J. Liu. Fast lighting independent background subtraction. *Technical Report*, 1997.

238. Y. Ivanov, A. Bobick, and J. Liu. Fast lighting independent background subtraction. *IEEE Workshop on Visual Surveillance, ICCV 1998*, January 1998.

239. Y. Ivanov, A. Bobick, and J. Liu. Fast lighting independent background subtraction. *International Journal of Computer Vision,*, pages 199–207, June 2000.

240. S. Jabri, Z. Duric, and H. Wechsler. Detection and location of people in video images using adaptive fusion of color and edge information. *ICPR 2000*, 2000.

241. N. Jacobs and R. Pless. Shape background modeling : The shape of things that came. *IEEE Workshop on Motion and Video Computing, WMVC 2007*, pages 1–7, 2007.

242. J. Jacques, C. Jung, and S. Musse. A background subtraction model adapted to illumination changes. *International Conference on Image Processing, ICIP 2006*, October 2006.

243. V. Jain, B. Kimia, and J. Mundy. Background modelling based on subpixel edges. *International Conference on Image Processing, ICIP 2007*, pages 321–324, September 2007.

244. A. Jalal and V. Singh. A robust background subtraction approach based on daubechies complex wavelet transform. *Advances in Computing and Communication, ACC 2011*, pages 516–524, 2011.

245. A. Jalal and V. Singh. Robust foreground detection in video using complex wavelet transform. *International Conference on Advances in Information Technology and Mobile Communication, AIM 2011*, April 2011.

246. A. Jalal and V. Singh. A framework for background modelling and shadow suppression for moving object detection in complex wavelet domain. *Multimedia Tools Applications*, 2013.

247. D. Jang, X. Jin, Y. Choi, and T. Kim. Background subtraction based on local orientation histogram. *Asia-Pacific conference on Computer-Human Interaction, APCHI 2008*, pages 222–231, 2008.

248. L. Jang. Interval-valued Choquet integrals and their applications. *Journal of Applied Mathematics and Computing*, 16(1):429–443, 2004.

249. O. Javed, K. Shafique, and M. Shah. A hierarchical approach to robust background subtraction using color and gradient information. *IEEE Workshop on Motion and Video Computing, WMVC 2002*, December 2002.

250. S. Javed, T. Bouwmans, and S. Jung. Depth Extended Online RPCA with Spatiotemporal Constraints for Robust Background Subtraction. *Korea-Japan Workshop on Frontiers of Computer Vision, FCV 2015*, January 2015.

251. S. Javed, S. Oh, T. Bouwmans, and S. Jung. Robust Background Subtraction to Global Illumination Changes via Multiple Features based OR-PCA with MRF. *SPIE Journal of Electronic Imaging*, 2015.

252. S. Javed, S. Oh, and S. Jung. Foreground object detection via background modeling using histograms of oriented gradient. *International Conference On Human Computer Interaction, HCI-2013*, January 2013.

253. S. Javed, S. Oh, A. Sobral, T. Bouwmans, and S. Jung. Background subtraction via superpixel-based online matrix decomposition with structured foreground constraints. *Workshop on Robust Subspace Learning and Computer Vision, ICCV 2015*, December 2015.

254. S. Javed, A. Sobral, T. Bouwmans, and S. Jung. OR-PCA with dynamic feature selection for robust background subtraction. *ACM Symposium On Applied Computing, SAC 2015*, April 2015.

255. S. Javed, A. Sobral, S. Oh, T. Bouwmans, and S. Jung. OR-PCA with MRF for Robust Foreground Detection in Highly Dynamic Backgrounds. *Asian Conference on Computer Vision, ACCV 2014*, November 2014.

256. D. Jeyabharathi and Dejey. New feature descriptor: Extended symmetrical-diagonal hexadecimal pattern for efficient background subtraction and object tracking. *Computers and Electrical Engineering*, 2017.

257. D. Jeyabharathi and D. Dejey. A novel Rotational Symmetry Dynamic Texture (RSDT) based sub space construction and SCD (Similar-Congruent-Dissimilar) based scoring model for background subtraction in real time videos. *Multimedia Tools and Applications*, pages 1–29, 2016.

258. Z. Ji and W. Wang. Detect foreground objects via adaptive fusing model in a hybrid feature space. *Pattern Recognition*, March 2014.

259. Y. Jian and J. Odobez. Multi-layer background subtraction based on color and texture. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2007*, pages 1–8, June 2007.

260. H. Jiang, W. Deng, and Z. Shen. Surveillance video processing using compressive sensing. *Inverse Problems and Imaging*, 6(2):201–214, 2012.

261. H. Jiang, S. Zhao, Z. Shen, W. Deng, P. Wilford, and R. Cohen. Surveillance video analysis using compressive sensing with low latency. *Bell Labs Technical Journal*, 18(4):63–74, March 2014.

262. G. Jing, C. Siong, and D. Rajan. Foreground motion detection by difference-based spatial temporal entropy image. *TENCON 2004*, 1:379–382, November 2004.

263. M. Jmal, W. Souidene, and R. Attia. Spatio-temporal center-symmetric local derivative patterns for objects detection in video surveillance. *International Conference on Computer Vision Theory and Applications, VISAPP 2016*, 2016.

264. J. Jodoin, G. Bilodeau, and N. Saunier. Background subtraction based on local shape. *CoRR 2012*, 2012.

265. P. Jodoin, V. Saligrama, and J. Konrad. Behavior subtraction. *IEEE Transactions on Image Processing*, 2011.

266. B. Julesz. Textons: The elements of texture perception, and their interactions. *Nature*, 290:91–97, March 1991.

267. C. Jung. Efficient background subtraction and shadow removal for monochromatic video sequences. *IEEE Transactions on Multimedia*, 11(3):571–577, April 2009.

268. J. Kamarainen, V. Kyrki, and H. Kalviainen. Invariance properties of Gabor filter-based features: overview and applications. *IEEE Transactions on Image Processing*, 15(5):1088–1099, 2006.

269. A. Kamkar-Parsi, R. Laganier, and M. Bouchard. Multi-criteria model for robust foreground extraction. *ACM International Workshop on Video Surveillance and Sensor Networks, VSSN 2005*, pages 67–70, November 2005.

270. B. Kang and W. Zhu. Robust moving object detection using compressed sensing. *IET Image Processing*, 2015.

271. B. Kang, W. Zhu, and J. Yan. Object detection oriented video reconstruction using compressed sensing. *EURASIP Journal on Advances in Signal Processing Sample*, February 2015.

272. S. Kanna and N. Murthy. Background Subtraction for Dynamic Texture Scenes using Local FCH features With Adaptive Updating Procedure. *Indian Streams Research Journal*, 3(11):1–13, December 2013.

273. S. Kanprachar and S. Tangkawanit. Performance of RGB and HSV color systems in object detection applications under different illumination intensities. *International Multi Conference of Engineers and Computer Scientists*, 2:1943–1948, March 2007.

274. I. Kavasidis, S. Palazzo, and C. Spampinato. An innovative web-based collaborative platform for video annotation. *Multimedia Tools and Applications*, pages 1–20, 2013.

275. C. Kertesz. Texture-based foreground detection. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 4(4), December 2011.

276. M. Khare, R. Srivastava, and A. Khare. Moving object segmentation in daubechies complex wavelet domain. *Signal, Image and Video Processing*, May 2013.

277. M. Khare, R. Srivastava, and A. Khare. Single change detection-based moving object segmentation by using daubechies complex wavelet transform. *IET Image Processing*, 8(6):334–344, June 2014.

278. M. Khare, R. Srivastava, A. Khare, and M. Jeoni. Curvelet transform based moving object segmentation. *IEEE International Conference on Image Processing, ICIP 2013*, 2013.

279. C. Kim and J. Hwang. Fast and automatic video object segmentation and tracking for content-based applications. *IEEE Transactions on Circuits Systems and Video Technologies*, 2002.

280. H. Kim. Robust change detection by global-illumination-change compensation and noise-adaptive thresholding. *Humantech Thesis*, 2004.

281. H. Kim, S. Kim, and S. Lee. Robust change detection by global-illumination-change compensation and noise-adaptive thresholding. *Optical Engineering*, 43(3):580–590, March 2004.

282. J. Kim, M. Murshed, A. Ramirez-Rivera, and O. Chae. Background modelling using edge-segment distributions. *International Journal of Advanced Robotic Systems*, October 2012.

283. J. Kim, A. Ramirez-Rivera, M. Park, and O. Chae. Scene modeling using edge segment distributions. *International Conference on Image Processing, Computer Vision, and Pattern Recognition, IPCV 2012*, July 2012.

284. J. Kim, A. Ramirez-Rivera, G. Song, B. Ryu, and O. Chae. Edge-segment-based background modeling: Non-parametric online background update. *International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2013*, 2013.

285. J. Kim, A. Rivera, B. Ryu, and O. Chae. Simultaneous foreground detection and classification with hybrid features. *International Conference on Computer Vision, ICCV 2015*, 2015.

286. K. Kim. Moving object detection using spatial coherence. *Applications of Digital Image Processing, ADIP 2009*, 2009.

287. K. Kim, T. Chalidabhongse, D. Harwood, and L. Davis. Background modeling and subtraction by codebook construction. *IEEE International Conference on Image Processing, ICIP 2004*, 2004.

288. K. Kim, T. Chalidabhongse, D. Harwood, and L. Davis. Background modeling and subtraction by codebook construction. *IEEE International Conference on Image Processing, ICIP 2004*, 2004.

289. K. Kim, T. Chalidabhongse, D. Harwood, and L. Davis. Real-time foreground-background segmentation using codebook model. *Elsevier Real-Time Imaging*, 2015.

290. W. Kim and C. Kim. Background subtraction for dynamic texture scenes using fuzzy color histograms. *IEEE Signal Processing Letters*, 19(3):127–130, March 2012.

291. W. Kim and Y. Kim. Background subtraction using illumination-invariant structural complexity. *IEEE Signal Processing Letters*, 3(4):634–638, May 2016.

292. K.Iwata, Y.Satoh, R.Ozaki, and K.Sakaue. Robust background subtraction based on statistical reach feature method. *IEICE Transactions on Information and Systems*, pages 1251–1259, 2009.

293. B. Klare. Background subtraction using ensembles of classifiers with an extended feature set. *Master Thesis, University of South Florida*, June 2008.

294. B. Klare and S. Sarkar. Background subtraction in varying illuminations u an ensemble based on an enlarged feature set. *OTCBVS 2009*, June 2009.

295. G. Klinker, S. Shafer, and T. Kanade. A physical approach to colour image understanding. *International Journal on Computer Vision*, 4(1):7–38, 1990.

296. T. Ko, S. Soatto, and D. Estrin. Background subtraction on distributions. *European Conference on Computer Vision, ECCV 2008*, pages 276–289, 2008.

297. T. Ko, S. Soatto, and D. Estrin. Background subtraction on distributions. *European Conference on Computer Vision, ECCV 2008*, pages 276–289, 2008.

298. F. Kristensen, P. Nilsson, and V. Wall. Background segmentation beyond RGB. *ACCV 2006*, pages 602–612, 2006.

299. B. Kumar, H. Reddy, and C. Jayachandra. Automated surveillance system and data communication. *IOSR Journal of Computer Engineering*, 12(2):31–38, June 2013.

300. P. Kumar, D. Rout, A. Kumar, M. Verma, and D. Kuma. Detection of video objects in dynamic scene using local binary pattern subtraction method. *Intelligent Computing, Communication and Devices, ICCD 2014*, pages 385–391, 2014.

301. P. Kumar, K Sengupta, and A. Lee. A comparative study of different color spaces for foreground and shadow detection for traffic monitoring system. *International Conference on Intelligent Transportation Systems*, pages 100–105, 2002.

302. A. Kushwah and R. Srivastava. Automatic moving object segmentation methods under varying illumination conditions for video data: comparative study, and an improved method. *Multimedia Tools Applications*, 2015.

303. A. Kushwah and R. Srivastava. Framework for dynamic background modeling and shadow suppression for moving object segmentation in complex wavelet domain. *SPIE Journal of Electronic Imaging*, 2015.

304. A. Kushwah and R. Srivastava. A framework for moving object segmentation under rapidly changing illumination conditions in complex wavelet domain. *International Conference on Futuristic Trends on Computational Analysis and Knowledge Management, ABLAZE 2015*, pages 148–153, February 2015.

305. A. Kushwaha and R. Srivastava. Complex wavelet based moving object segmentation using approximate median filter based method for video surveillance. *IEEE International on Advance Computing Conference, IACC 2014*, pages 973–978, February 2014.

306. H. Lai, C. Lin, M. Chen, L. Kang, and C. Yeh. Moving objects detection based on hysteresis thresholding. *Advances in Smart Innovation, Systems and Technologies*, 21:289–298, 2013.

307. B. Langmann, S. Ghobadi, K. Hartmann, and O. Loffeld. Multi-model background subtraction using Gaussian mixture models. *Symposium on Photogrammetry Computer Vision and Image Analysis, PCV 2010*, pages 61–66, 2010.

308. B. Langmann, K. Hartmann, and O. Loffeld. Depth assisted background subtraction for color capable ToF-Cameras. *International Conference on Image and Video Processing and Computer Vision, IVPCV 2010*, pages 75–82, July 2010.

309. L. Latecki, R. Miezianko, V. Megalooikonomou, and D. Pokrajac. Using spatiotemporal blocks to reduce the uncertainty in detecting and tracking moving objects in video. *International Journal of Intelligent Systems Technologies and Applications, IJISTA 2006*, 2006.

310. L. Latecki, R. Miezianko, and D. Pokrajac. Evaluating reliability of motion features in surveillance videos. *NIST Performance Metrics for Intelligent Systems Workshop, NIST 2004*, August 2004.

311. L. Latecki, R. Miezianko, and D. Pokrajac. Motion detection based on local variation of spatiotemporal texture. *CVPR Workshop on Object Tracking and Classification Beyond the Visible Spectrum, CVPR 2004*, July 2004.

312. L. Latecki, R. Miezianko, and D. Pokrajac. Activity and motion detection based on measuring texture change. *International Conference on Machine Learning and Data Mining in Pattern Recognition, MLDM 2005*, pages 476–486, July 2005.

313. L. Latecki, R. Miezianko, and D. Pokrajac. Activity and motion detection based on measuring texture change. *Industrial Conference on Data Mining, ICDM 2005*, 2005.

314. L. Latecki, R. Miezianko, and D. Pokrajac. Instantaneous reliability assessment of motion features in surveillance videos. *SPIE Conference Storage and Retrieval Methods and Applications for Multimedia*, January 2005.

315. L. Latecki, R. Miezianko, and D. Pokrajac. Reliability of motion features in surveillance videos. *Integrated Computer-Aided Engineering, ICAE 2005*, 12(3):279–290, 2005.

316. Y. Lee, J. Jung, and I. Kweon. Hierarchical on-line boosting based background subtraction. *Workshop on Frontiers of Computer Vision, FCV 2011*, pages 1–5, February 2011.

317. J. Leens, O. Barnich, S. Pierard, M. Van Droogenbroeck, and J. Wagner. Combining color, depth, and motion for video segmentation. *Computer Vision Systems*, pages 104–113, 2009.

318. K. Lertniphonphan, S. Aramvith, and T. Chalidabhongse. A background modeling for non-empty scene videos. *Chulalongkorn University, Thailand*, 2014.

319. A. Leykin and R. Hammoud. Robust multi-pedestrian tracking in thermal-visible surveillance videos. *International Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum, OTCBVS 2006*, 2006.

320. A. Leykin, Y. Ran, and R. Hammoud. Thermal-visible video fusion for moving target tracking and pedestrian classification. *International Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum, OTCBVS 2007*, 2007.

321. A. Li, Z. Tang, B. Yuan, and Z. Miao. Segmentation of moving foreground objects using codebook and local binary patterns. *Congress on Image and Signal Processing, CISP 2008*, pages 239–243, 2008.

322. B. Li, J. Yu, D. Jiang, Q. Chen, and Z. Zhang. ST-CSLBP based space-time mixed background modeling algorithm. *Chinese Journal of Scientific Instrument*, 36(12):2781–2786, 2011.

323. B. Li, B. Yuan, and Z. Miao. Moving object detection in dynamic scenes using nonparametric local kernel histogram estimation. *IEEE International Conference on Multimedia and Expo, ICME 2008*, pages 1461–164, 2008.

324. J. Li. Moving objects segmentation based on histogram for video surveillance. *Modern Applied Science*, 3(11), November 2009.

325. J. Li, J. Wang, and W. Shen. Moving object detection in framework of compressive sampling. *Journal of Systems Engineering and Electronics*, 5(740-745), October 2010.

326. L. Li, W. Huang, I. Gu, and Q. Tian. Statistical modeling of complex background for foreground object detection. *IEEE Transaction on Image Processing*, 13(11):1459–1472, November 2004.

327. S. Li, H. Huang, and R. Li. Accurate detection of moving regions via a nested model. *International Conference on Intelligent Transportation Systems, ITSC 2014*, October 2014.

328. S. Li and H. Qi. Recursive low-rank and sparse recovery of surveillance video using compressed sensing. *International Conference on Distributed Smart Cameras, ICDSC 2014*, 2014.

329. W. Li, Y. Ding, and Y. Wang. Color-texture feature based background subtraction of video. *International Workshop on Information Technology, IWIT 2009*, September 2009.

330. Z. Li, P. Jiang, H. Ma, J. Yang, and D. Tang. A model for dynamic object segmentation with kernel density estimation based on gradient features. *Image and Vision Computing*, 2009.

331. Z. Liang, X. Liu, H. Liu, and W. Chen. A refinement framework for background subtraction based on color and depth data. *IEEE International Conference on Image Processing, ICIP 2016*, 2016.

332. S. Liao, G. Zhao, V. Kellokumpu, M. Pietikinen, and S. Li. Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes. *International Conference on Computer Vision and Pattern Recognition, CVPR 2010*, June 2010.

333. C. Lin, C. Chang, W. Chang, M. Chen, and L. Kang. Real-time robust background modeling based on joint color and texture descriptions. *IEEE International Conference on Genetic and Evolutionary Computing, ICGEC 2010*, pages 622–625, December 2010.

334. G. Lin, S. Yang, C. Wang, and C. Lin. Background subtraction based on codebook model and texture feature. *International Symposium on Computer, Consumer and Control, IS3C 2016*, pages 498–501, 2016.

335. H. Lin, T. Liu, and J. Chuang. A probabilistic SVM approach for background scene initialization. *IEEE International Conference on Image Processing, ICIP 2002*, September 2002.

336. H. Lin, T. Liu, and J. Chuang. Learning a scene background model via classification. *IEEE Transactions on Signal Processing*, 57(5):1641–1654, May 2009.

337. L. Lin, Y. Xu, X. Liang, and J. Lai. Complex background subtraction by pursuing dynamic spatio-temporal models. *IEEE Transactions on Circuits and Systems For Video Technology*, 2013.

338. Y. Lin, Y. Tong, Y. Cao, Y. Zhou, and S. Wang. Visual-attention based background modeling for detecting infrequently moving objects. *IEEE Transactions on Circuits and Systems for Video Technology*, 2016.

339. J. Lindstrom, F. Lindgren, K. Astrom, J. Holst, and U. Holst. Background and foreground modeling using an online EM algorithm. *IEEE International Workshop on Visual Surveillance VS 2006 in conjunction with ECCV 2006*, May 2006.

340. C. Liu. A robust texture-based background subtraction for moving object detection in video sequences. *Thesis, I-Shou University, Kaohsiung, Taiwan*, June 2007.

341. C. Liu, J. Yuen, and A. Torralba. SIFT flow: dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):978–994, 2010.

342. G. Liu and J. Yang. Image retrieval based on the texton co-occurrence matrix. *Pattern Recognition*, 41(12):3521–3527, 2008.

343. M. Liu, H. Liu, and C. Chen. Robust 3D Action Recognition through Sampling Local Appearances and Global Distributions. *IEEE Transactions on Multimedia*, December 2017.

344. W. Liu, S. Lin, M. Yang, and C. Huang. Real-time binary descriptor based background modeling. *IAPR Asian Conference on Pattern Recognition*, pages 722–726, November 2013.

345. Y. Liu and D. Pados. Compressed-sensed-domain l1-pca video surveillance. *IEEE Transactions on Multimedia*, 2016.

346. Z. Liu, F. Zhou, X. Chen, X. Bai, and C. Sun. Iterative infrared ship target segmentation based on multiple features. *Pattern Recognition*, 2014.

347. E. Lopez-Rubio, R. Luque-Baena, and E. Domnguez. Foreground detection in video sequences with probabilistic self-organizing maps. *International Journal of Neural Systems*, pages 225–246, 2011.

348. F. Lopez-Rubio and R. Luque Baena. Stochastic approximation for background modelling. *Computer Vision and Image Understanding*, 116(6):735–749, 2011.

349. F. Lopez-Rubio and E. Lopez-Rubio. Features for stochastic approximation based foreground detection. *Computer Vision and Image Understanding*, December 2014.

350. F. Lopez-Rubio, E. Lopez-Rubio, R. Luque-Baena, E. Dominguez, and E. Palomo. Color space selection for self-organizing map based foreground detection in video sequences. *International Joint Conference on Neural Networks, IJCNN 2014*, pages 3347–3354, July 2014.

351. X. Lu, T. Izumi, T. Takahashi, and L. Wang. Moving vehicle detection based on fuzzy background subtraction. *IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2014*, pages 529–532, July 2014.

352. X. Lu, T. Izumi, L. Teng, T. Hori, and L. Wang. A novel background subtraction method for moving vehicle detection. *IEEJ Transactions on Fundamentals and Materials*, 132(10):857–863, 2012.

353. R. Luo, L. Li, and I. Gu. Efficient adaptive background subtraction based on multi-resolution background modelling and updating. *Pacific-RIM Conference on Multimedia, PCM 2007*, December 2007.

354. F. Ma and N. Sang. Background subtraction based on multi-channel SILTP. *Asian Conference on Computer Vision, ACCV 2012*, November 2012.

355. Y. Ma and H. Zhang. Detecting motion object by spatial-temporal entropy. *IEEE International Conference on Multimedia and Expo, ICME 2001*, pages 265–268, 2001.

356. L. Maddalena and A. Petrosino. A self-organizing neural system for background and foreground modeling. *International Conference on Artificial Neural Networks, ICANN 2008*, pages 652–661, 2008.

357. L. Maddalena and A. Petrosino. A fuzzy spatial coherence-based approach to background/foreground separation for moving object detection. *Neural Computing and Applications*, 19(2):179–186, March 2010.

358. K. Maeno, H. Nagahara, A. Shimada, and R. Taniguchi. A background invariant feature for transparent object recognition. *Joint Workshop on Machine Perception and Robotics*, 2013.

359. K. Makantasis, A. Doulamis, and N. Doulamis. Variational inference for background subtraction in infrared imagery. *International Symposium on Visual Computing, ISVC 2015*, 2015.

360. K. Makantasis, A. Nikitakis, A. Doulamis, N. Doulamis, and Y. Papaefstathiou. Data-driven background subtraction algorithm for in-camera acceleration in thermal imagery. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017.

361. R. Marie, A. Potelle, and E. Mouaddib. Dynamic background subtraction using moments. *IEEE International Conference on Image Processing, ICIP 2011*, 20(4):2369–2372, 2011.

362. I. Martins, P. Carvalho, L. Corte-Real, and J. Luis Alba-Castro. Bio-inspired boosting for moving objects segmentation. *International Conference on Image Analysis and Recognition, ICIAR 2016*, 2016.

363. M. Mason and Z. Duric. Using histograms to detect and track objects in color video. *Applied Imagery Pattern Recognition Workshop, AIPR 2001*, pages 154–159, October 2001.

364. N. McFarlane and C. Schofield. Segmentation and tracking of piglets in images. *British Machine Vision and Applications*, pages 187–193, 1995.

365. A. Mendizabal and L. Salgado. A region based approach to background modeling in a wavelet multi-resolution framework. *International Conference on Acoustics, Speech and Signal Processing, ICASSP 2011*, May 2011.

366. T. Menp and M. Pietikinen. Classification with color and texture: jointly or separately? *Pattern Recognition*, 2004.

367. R. Miezianko and D. Pokrajac. Multi-layer background change detection based on spatiotemporal texture projections. *International Conference on Computer Vision and Graphics, ICCVG 2008*, pages 282–291, 2008.

368. R. Miezianko and D. Pokrajac. Texture dissimilarity measures for background change detection. *International Conference on Image Analysis and Recognition, ICIAR 2008*, pages 680–687, 2008.

369. I. Mikic, P. Cosman, G. Kogut, and M. Trivedi. Moving shadow and object detection in traffic scenes. *International Conference on Pattern Recognition, ICPR 2000*, 1:321–324, 2000.

370. Y. Ming, J. Jiang, and J. Ming. Background modeling and subtraction using a local linear dependence based cauchy statistical model. *Digital Image Computing, Techniques and Applications, DICTA 2003*, December 2003.

371. Y. Ming, J. Jiang, and J. Yang. Statistical change detection based on cauchy distribution for visual surveillance. *International Conference on Visual Information Engineering, VIE 2006*, pages 559–564, September 2006.

372. I. Mitsugami, H. Fukui, and M. Minoh. Extraction of potential sunny region for background subtraction under sudden illumination changes. *International Journal of Computer Vision and Signal Processing*, 4:22–28, 2014.

373. A. Mittal. Motion-based background subtraction using adaptive kernel density estimation. *International Conference on Computer Vision and Pattern Recognition, CVPR 2004*, July 2004.

374. S. Miyamori, K. Saito, Y. Fukumizu, and H. Yamauchi. Adaptive BP-RRC mixture model for background subtraction. *International Congress on Image and Signal Processing, CISP 2011*, 3:1180–1183, October 2011.

375. S. Miyamori, K. Saito, Y. Fukumizu, and H. Yamauchi. Background subtraction using multi radial proportion filter. *Journal of Signal Processing*, 6(4):339–342, July 2012.

376. S. Miyamori, K. Saito, Y. Fukumizu, and H. Yamauchi. Background subtraction using multi radial proportion filter. *International Workshop on Nonlinear Circuits, Communications and Signal Processing, NCSP 2012*, March 2012.

377. M. Molina-Cabello, E. Lopez-Rubio, R. Luque-Baena, E. Domnguez, and E. Palomo. Pixel features for self-organizing map based detection of foreground objects in dynamic environments. *International Joint Conference SOCO 2016-CISIS 2016-ICEUTE 2016*, pages 247–255, October 2016.

378. S. Molina-Giraldo, J. Carvajal-Gonzlez, A. lvarez Meza, and G. Castellanos-Domnguez. Video segmentation based on multi-kernel learning and feature relevance analysis for object classification. *International Conference on Pattern Recognition Applications and Methods, ICPRAM 2013*, pages 396–401, February 2013.

379. S. Molina-Giraldo, A. lvarez Meza, D. Peluffo-Ordonez, and G. Castellanos-Domnguez. Image segmentation based on multi-kernel learning and feature relevance analysis. *Ibero-American Conference on Artificial Intelligence*, pages 501–510, 2012.

380. A. Morde, X. Ma, and S. Guler. Learning a background model for change detection. *IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2012*, pages 15–20, 2012.

381. V. Morellas, I. Pavlidis, and P. Tsiamyrtzis. DETER: detection of events for threat evaluation and recognition. *Machine Vision Applications*, 15:29–45, 2003.

382. J. Mota, N. Deligiannis, A. Sankaranarayanan, V. Cevher, and M. Rodrigues. Adaptive-rate sparse signal reconstruction with application in compressive background subtraction. *Preprint*, March 2015.

383. J. Mota, N. Deligiannis, A. Sankaranarayanan, V. Cevher, and M. Rodrigues. Dynamic sparse state estimation using l1-l1 minimization: Adaptive-rate measurement bounds, algorithms and applications. *International Conference on Acoustics, Speech and Signal Processing, ICASSP 2015*, April 2015.

384. G. Moya-Alcover, A. Elgammal, A. Jaume i Capo, and J. Varona. Modelling depth for nonparametric foreground segmentation using RGBD devices. *Pattern Recognition Letters*, 2016.

385. K. Muchtar, C. Lin, L. Kang, and C. Yeh. Robust background modeling based on multiscale color description. *APSIPA Annual Summit and Conference special session on Image/Video Processing and Analysis, APSIPA ASC 2011*, October 2011.

386. D. Mukherjee, Q. Wu, and T. Nguyen. Gaussian mixture model with advanced distance measure based on support weights and histogram of gradients for background suppression. *Applied Intelligence*, 2013.

387. A. Mumtaz, W. Zhang, and A. Chan. Joint motion segmentation and background estimation in dynamic scenes. *International Conference on Computer Vision and Pattern Recognition, CVPR 2014*, June 2014.

388. O. Munteanu, T. Bouwmans, E. Zahzah, and R. Vasiu. The detection of moving objects in video by background subtraction using Dempster-Shafer theory. *Transactions on Electronics and Communications*, 60(1), March 2015.

389. J. Murgia, C. Meurie, and Y. Ruichek. An Improved Colorimetric Invariants and RGB-Depth based Codebook Model for Background Subtraction using Kinect. *MICAI 2014*, 2014.

390. M. Murshed and O. Chae. Statistical background modeling: an edge segment based moving object detection approach. *International Conference on Advanced Video and Signal Based Surveillance, AVSS 2010*, September 2010.

391. M. Murshed, A. Ramirez-Rivera, J. Kim, and O. Chae. Statistical binary edge frequency accumulation model for moving object detection. *International Journal of Innovative Computing, Information and Control*, 2012.

392. S. Nadimi and B. Bhanu. Physics-based models of color and ir video for sensor fusion. *IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 161–166, July 2003.

393. S. Nadimi and B. Bhanu. Physics-based cooperative sensor fusion for moving object detection. *IEEE Workshop on Learning in Computer Vision and Pattern Recognition, CVPR 2004*, June 2004.

394. M. Narayana, A. Hanson, and E. Learned-Miller. Background modeling using adaptive pixelwise kernel variances in a hybrid feature space. *International Conference on Computer Vision and Pattern Recognition, CVPR 2012*, 2012.

395. M. Narayana, A. Hanson, and E. Learned-Miller. Improvements in joint domain-range modeling for background subtraction. *BMVC 2012*, 2012.

396. M. Narayana, E. Learned-Miller, and A. Hanson. Background subtraction - separating the modeling and the inference. *Machine Vision and Applications*, 2014.

397. D. Needell and J. Tropp. CoSaMP: iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, June 2008.

398. A. Nghiem and F. Bremond. Background subtraction in people detection framework for RGB-D cameras. *IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2014*, pages 241–246, 2014.

399. T. Nguyen and K. Miyata. Multi-scale region perpendicular local binary pattern: an effective feature for interest region description. *Visual Computer*, pages 391–406, April 2015.

400. V. Nguyen, H. Vu, and T. Tran. Background Subtraction with Kinect data: An Efficient Combination RGB and Depth. *NAFOSTED Conference on Information and Computer Science, NICS 2014*, pages 160–169, March 2014.

401. V. Nguyen, H. Vu, and T. Tran. An efficient combination of RGB and depth for background subtraction. *Some Current Advanced Researches on Information and Computer Science in Vietnam*, pages 49–63, 2015.

402. A. Nikitakis, I. Papaefstathiou, K. Makantasis, and A. Doulamis. A novel background subtraction scheme for in-camera acceleration in thermal imagery. *Design, Automation and Test in Europe Conference and Exhibition, DATE 2016*, pages 1497–1500, 2016.

403. S. Noh and M. Jeon. A new real-time foreground extraction method using cubic codebook and graph-cut algorithm. *IASTED Signal Processing, Pattern Recognition, and Applications, SPPRA 2011*, February 2011.

404. S. Noh and M. Jeon. A new framework for background subtraction using multiple cues. *Asian Conference on Computer Vision, ACCV 2012*, November 2012.

405. Y. Nonaka, A. Shimada, H. Nagahara, and R. Taniguchi. Evaluation report of integrated background modeling based on spatio-temporal features. *IEEE Workshop on Change Detection, CVPRW 2012*, June 2012.

406. P. Noriega, B. Bascle, and O. Bernier. Local kernel color histograms for background substraction. *International Conference on Computer Vision Theory and Applications, VISAPP 2006*, 1:213219, February 2006.

407. P. Noriega and O. Bernier. Real time illumination invariant background subtraction using local kernel histograms. *British Machine Vision Conference, BMVC 2006*, September 2006.

408. T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, July 2002.

409. O. Okun. Feature selection and ensemble methods for bioinformatics:algorithmic classification and implementations. *Information Science*, 2011.

410. S. Ottonelli, P. Spagnolo, P. Mazzeo, and M. Leo. Foreground segmentation by combining color and depth images. *Iberian Conference on Pattern Recognition and Image Analysis, IbPRIA 2013*, June 2013.

411. S. Ottonelli, P. Spagnolo, P. Mazzeo, and M. Leo. Improved video segmentation with color and depth using a stereo camera. *IEEE International Conference on Industrial Technology, ICIT 2013*, February 2013.

412. C. Ouyang and P. Chen. A robust texture-based background subtraction algorithm for moving object detection in video sequences. *International Conference on Genetic and Evolutionary Computing, ICGEC 2012*, pages 480–483, 2012.

413. G. Pajares, J. Ruz, and J. Manuel de la Cruz. Performance analysis of homomorphic systems for image change detection. *IBPRIA 2005*, pages 563–570, 2005.

414. D. Panda and S. Meher. Dynamic background subtraction using texton co-occurrence matrix. *IEEE India Conference Emerging Trends and Innovation in Technology, INDICON 2014*, December 2014.

415. D. Panda and S. Meher. Dynamic background subtraction using local binary pattern and histogram of oriented gradients. *International Conference on Image Information Processing, ICIIP 2015*, pages 306–311, 2015.

416. D. Panda and S. Meher. Detection of moving objects using fuzzy color difference histogram based background subtraction. *IEEE Signal Processing Letters*, 23(1):45–49, January 2016.

417. T. Parag, A. Elgammal, and A. Mittal. A framework for feature selection for background subtraction. *International Conference on Pattern Recognition, CVPR 2006*, June 2006.

418. V. Parameswaran, M. Singh, and V. Ramesh. Illumination compensation based change detection using order consistency. *IEEE Computer Vision and Pattern Recognition Conference*, June 2010.

419. D. Park and H. Byun. Object-wise multilayer background ordering for public area surveillance. *International Conference on Advanced Video and Signal Based Surveillance, AVSS 2009*, pages 484–489, September 2009.

420. J. Park and C. Lee. Improved background modeling through color de-correlation. *European Signal Processing Conference, EUSIPCO 2011*, September 2011.

421. J. Park, G. Lee, W. Cho, N. Toan, S. Kim, and S. Park. Moving object detection based on Clausius entropy. *IEEE International Conference on Computer and Information Technology, CIT 2010*, pages 517–521, June 2010.

422. J. Park, G. Lee, N. Toan, W. Cho, and S. Park. Moving object detection using Clausius entropy and adaptative Gaussian mixture model. *Journal of The Institute of Electronics Engineers of Korea*, 47(1), January 2010.

423. J. Paruchuri, S. Edwin, S. Cheung, and C. Chen. Spatially adaptive illumination modeling for background subtraction. *International Conference on Computer Vision, ICCV 2011 Workshop on Visual Surveillance*, November 2011.

424. K. Patwardhan, G. Sapiro, and V. Morellas. Robust foreground detection in video using pixel layers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(4):746–751, April 2008.

425. M. Paul, W. Lin, C. Lau, and B. Lee. Video coding with dynamic backgrounds. *EURASIP Journal on Advances in Signal Processing*, 2013.

426. N. Pejcic, N. Reljin, S. McDaniel, D. Pokrajac, and A. Lazarevic. Detection of moving objects using incremental connectivity outlier factor algorithm. *ACM Conference SE*, 2009.

427. D. Penciuc, F. El Baf, and T. Bouwmans. Comparison of background subtraction methods for an interactive learning space. *NETTIES 2006*, September 2006.

428. J. Pilet, C. Strecha, and P. Fua. Making background subtraction robust to sudden illumination changes. *European Conference on Computer Vision, ECCV 2008*, October 2008.

429. D. Pokrajac and L. Latecki. Spatiotemporal blocks-based moving objects identification and tracking. *IEEE Visual Surveillance and Performance Evaluation of Tracking and Surveillance, VS-PETS 2003*, October 2003.

430. D. Pokrajac and L. Latecki. Entropy-based approach for detecting feature reliability. *ETRAN 2004*, 2004.

431. D. Pokrajac, V. Zeljkovic, and L. Lateckci. Spatial-temporal algorithm for moving objects detection in infra red video sequences. *International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services, TELSIKS 2005*, September 2005.

432. D. Pokrajac, V. Zeljkovic, and L. Latecki. Noise-resilient detection of moving objects based on spatial-temporal blocks. *International Symposium ELMAR focused on Multimedia Systems and Applications, ELMAR 2005*, June 2005.

433. C. Poppe, S. Bruyne, T. Paridaens, P. Lambert, and R. Wall. Moving object detection in the H.264/AVC compressed domain for video surveillance applications,. *Journal of Vision Communication and Image Representation*, 20(6):428–437, August 2009.

434. F. Porikli. Integral histogram: A fast way to extract histograms in cartesian spaces. *International Conference on Computer Vision and Pattern Recognition, CVPR 2005*, 1(1):829–836, 2005.

435. F. Porikli. Real-time video object segmentation for MPEG-encoded video sequences. *SPIE Conference on Real-Time Imaging*, pages 195–203, 2014.

436. F. Porikli, F. Bashir, and H. Sun. Compressed domain video object segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 20:2–14, 2010.

437. F. Porikli and C. Wren. Change detection by frequency decomposition: Waveback. *International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2005*, April 2005.

438. Y. Qiao, K. Yuan, C. Song, and X. Xiang. Detection of moving objects with fuzzy color coherence vector. *Mathematical Problems in Engineering*, 2014.

439. Z. Qu, M. Yu, and J. Liu. Real-time traffic vehicle tracking based on improved MoG background extraction and motion segmentation. *International Symposium on Systems and Control in Aeronautics and Astronautics, ISSCAA 2010*, pages 676–680, June 2010.

440. Z. Qu, S. Yu, and M. Fu. Motion background modeling based on context-encoder. *IEEE International Conference on Artificial Intelligence and Pattern Recognition, ICAIPR 2016*, September 2016.

441. L. Ramesh and P. Shah. R-SpaRCS: An Algorithm for Foreground-Background Separation of Compressively-Sensed Surveillance Video. *IEEE International Conference on Advanced Video and Signal based Surveillance, AVSS 2015*, 2015.

442. A. Ramirez, J. Castillo, and O. Chae. Recognition of face expressions using local principal texture pattern. *International Conference on Image Processing, ICIP 2012*, September 2012.

443. A. Ramirez-Rivera, M. Murhsed, J. Kim, and O. Chae. Background modeling through statistical edge-segment distributions. *IEEE Transactions on Circuits and Systems for Video Technology*, 2012.

444. A. Ramirez-Rivera, M. Murshed, and O. Chae. Object detection through edge behavior modeling. *International Conference on Advanced Video and Signal Based Surveillance, AVSS 2011*, 2011.

445. B. Rani and V. Naidu. A novel CWT based moving object detection using background subtraction. *International Journal of Computational Science, Mathematics and Engineering*, 2(4), April 2015.

446. A. Reckley, W. Hsu, C. Chen, G. Ma, and E. Huang. Sensor selection and integration to improve video segmentation in complex environments. *Mathematical Problems in Engineering*, 2014.

447. V. Reddy, C. Sanderson, and B. Lovell. An efficient and robust sequential algorithm for background estimation in video surveillance. *IEEE International Conference on Image Processing, ICIP 2009*, pages 1109–1112, 2009.

448. N. Reljin, S. McDaniel, D. Pokrajac, N. Pejcic, T. Vance, A. Lazarevic, and L. Latecki. Small moving targets detection using outlier detection algorithms. *SPIE 2010*, 2010.

449. C. Reyneke, P. Robinson, and A. Nel. Comparison of background subtraction techniques under sudden illumination changes. *Pattern Recognition Association of South Africa, PRASA 2014*, November 2014.

450. H. Ribeiro and A. Gonzaga. Hand image segmentation in video sequence by GMM: a comparative analysis. *XIX Brazilian Symposium on Computer Graphics and Image Processing, SIBGRAPI 2006*, pages 357–364, 2006.

451. A. Rodrigues, L. Milen, E. Nascimento, and W. Schwartz. Change detection based on features invariant to monotonic transforms and spatial constrained matching. *International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2014*, 2014.

452. K. Roy, J. Kim, M. Iqbal, F. Makhmudkhujaev, B. Ryu, and O. Chae. An adaptive fusion scheme of color and edge features for background subtraction. *International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2017*, 2017.

453. N. Sengupta S. Biswas, J. Sil. Background modeling and implementation using discrete wavelet transform: a review. *Journal ICGST-GVIP*, 11(1):29–42, March 2011.

454. H. Sajid and S. Cheung. Background subtraction under sudden illumination change. *IEEE workshop on Multimedia Signal Processing*, pages 1–6, 2014.

455. A. Salhi, M. Kardouchi, and N. Belacel. Histograms of fuzzy oriented gradients for face recognition. *Computer Applications Technology, ICCAT 2013*, pages 1–5, 2013.

456. J. Santoyo-Morales and R. Hasimoto-Beltran. Video background subtraction in complex environments. *Journal of Applied Research and Technology*, 12(3):527–537, June 2014.

457. Y. Satoh. Robust background subtraction by bi-polar radial reach correlation. *IEIC 2005*, 104(741):73–78, 2005.

458. Y. Satoh. Object tracking using renewal of a background based on radial reach filter. *Korea-Japan Joint Workshop Frontiers of Computer Vision, FCV 2006*, pages 13–17, 2006.

459. Y. Satoh, S. Kaneko, and S. Igarashi. Robust object detection and segmentation by peripheral increment sign correlation image. *Systems and Computers in Japan*, 35(9):70–80, 2004.

460. Y. Satoh, S. Kaneko, Y. Niwa, and K. Yamamoto. Robust object detection using a radial reach filter (rrf). *2004 Wiley Periodicals*, 35(10):63–73, 2004.

461. Y. Satoh and K. Sakaue. Robust background subtraction based on bi-polar radial reach correlation. *IEEE International Conference on Computers, Communications, Control and Power Engineering, TENCON 2005*, 2005.

462. Y. Satoh, H. Tanahashi, Y. Niwa, S. Kaneko, and K. Yamamoto. Robust object detection and segmentation based on radial reach correlation. *IAPR Workshop on Machine Vision Applications, MVA 2002*, pages 512–517, 2002.

463. Y. Satoh, H. Tanahashi, Y. Niwa, and K. Yamamoto. Robust human detection from complex background by radial reach filter. *IEEE Conference on Computers, Communications, Control and Power Engineering, TENCON 2002*, pages 533–536, 2002.

464. Y. Satoh, H. Tanahashi, C. Wang, S. Kaneko, Y. Niwa, and K. Yamamoto. Robust event detection by radial reach filter (RRF). *International Conference on Pattern Recognition, ICPR 2002*, 2002.

465. Y. Satoh, C. Wang, H. Tanahashi, Y. Niwa, and K. Yamamoto. Robust object detection for intelligent surveillance systems based on radial reach correlation (RRC). *IEEE International Conference on Intelligent Robots and Systems, IROS 2003*, pages 224–229, 2003.

466. A. Satpathy, H. Eng, and X. Jiang. Difference of gaussian edge-texture based background modeling for dynamic traffics. *International Symposium on Visual Computing, ISVC 2008*, 1:406–417, December 2008.

467. R. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, 37(3), March 1999.

468. J. Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 2015.

469. M. Sedky, M. Moniri, and C. Chibelushi. Spectral-360: A Physics-Based Technique for Change Detection. *IEEE Change Detection Workshop, CDW 2014*, June 2004.

470. M. Sedky, M. Moniri, and C. Chibelushi. Object segmentation using full-spectrum matching of albedo derived from colour images. *US patent no. 2374109 12.10.2011*, 2011.

471. N. Setiawan, S. Hong, J. Kim, and C. Lee. Gaussian mixture model in improved IHLS color space for human silhouette extraction. *International Conference on Artificial Reality and Telexistence, ICAT 2006*, pages 732–741, 2006.

472. M. Shafiee, P. Siva, P. Fieguth, and A. Wong. Embedded motion detection via neural response mixture background modeling. *International Conference on Computer Vision and Pattern Recognition, CVPR 2016*, June 2016.

473. M. Shah, J. Deng, and B. Woodford. Video background modeling: recent approaches, issues and our proposed techniques. *Machine Vision and Applications*, 25(5):1105–1119, July 2014.

474. S. Shah, T. Goldstein, and C. Studer. Estimating sparse signals with smooth support via convex programming and block sparsity. *Preprint*, 2016.

475. Y. Shan, F. Yang, and R. Wang. Color space selection for moving shadow elimination. *International Conference on Image and Graphics, ICIG 2007*, pages 496–501, 2007.

476. Z. Shang, Z. Qiang, and H. Liu. An algorithm of moving object detection based on texture and color model. *International Conference on Graphic and Image Processing, ICGIP 2012*, October 2012.

477. G. Sharma, W. Wu, and E. Dalal. The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations. *Color Research and Application*, 30(1):21–30, February 2005.

478. L. Sharma, D. Yadav, and K. Sing. Fishers linear discriminant ratio based threshold for moving human detection in thermal video. *Infrared Physics and Technology*, 78:118–128, September 2016.

479. Y. Sheikh and M. Shah. Bayesian modeling of dynamic scenes for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI 2005*, 27(11):1778–1792, November 2005.

480. Y. Sheikh and M. Shah. Bayesian modeling of dynamic scenes for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1778–1792, 2005.

481. Y. Sheikh and M. Shah. Bayesian object detection in dynamic scenes. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2005*, June 2005.

482. Y. Shen, W. Hu, M. Yang, J. Liu, B. Wei, S. Lucey, and C. Chou. Real-time and robust compressive background subtraction for embedded camera networks. *IEEE Transactions on Mobile Computing*, 2016.

483. L. Shi and J. Liu. Perception-inspired background subtraction in complex scenes based on spatiotemporal features. *SPIE Infrared Technology and Applications, and Robot Sensing and Advanced Control*, October 2016.

484. A. Shimada, H. Nagahara, and R. Taniguchi. Object detection based on spatio-temporal light field sensing. *IPSJ Transactions on Computer Vision and Applications*, 5:129–133, 2013.

485. A. Shimada, H. Nagahara, and R. Taniguchi. Change detection on light field for active video surveillance. *IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2015*, 2015.

486. A. Shimada and R. Taniguchi. Hybrid background model using spatial-temporal LBP. *IEEE International Conference on Advanced Video and Signal based Surveillance, AVSS 2009*, September 2009.

487. X. Shu. Advanced imaging via multiplexed sensing and compressive sensing. *PhD Thesis,University of Illinois at Urbana-Champaign*, 2013.

488. X. Shu and N. Ahuja. Imaging via three-dimensional compressive sampling (3DCS). *International Conference on Computer Vision, ICCV 2011*, pages 439–446, November 2011.

489. M. Sigari and M. Fathy. Real-time background modeling/subtraction using two-layer codebook model. *International Multiconference on Engineering and Computer Science, IMECS 2008*, March 2008.

490. C. Silva, T. Bouwmans, and C. Frelicot. An eXtended center-symmetric local binary pattern for background modeling and subtraction in videos. *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISAPP 2015*, March 2015.

491. C. Silva, T. Bouwmans, and C. Frelicot. Online weighted one-class ensemble for feature selection in background/foreground separation. *International Conference on Pattern Recognition, ICPR 2016*, December 2016.

492. C. Silva, T. Bouwmans, and C. Frelicot. Superpixel-based online wagging one-class ensemble for feature selection in background/foreground separation. *Submitted to Pattern Recognition Letters*, 2017.

493. J. Silveira, C. Jung, and S. Musse. Background subtraction and shadow detection in grayscale video sequences. *Brazilian Symposium on Computer Graphics and Image Processing, SIBGRAPI 2005*, pages 189–196, 2005.

494. D. Silvestre. Video surveillance using a time-of-light camera. *Master Thesis, Informatics and Mathematical Modelling, University of Denmark*, 2007.

495. M. Singh, V. Parameswaran, and V. Ramesh. Order consistent change detection via fast statistical significance testing. *IEEE Computer Vision and Pattern Recognition Conference, CVPR 2008*, June 2008.

496. P. Siva, M. Shafiee, F. Li, and A. Wong. PIRM: fast background subtraction under sudden, local illumination changes via probabilistic illumination range modeling. *IEEE International Conference on Image Processing, ICIP 2015*, 2015.

497. A. Sobral and A. Vacavant. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Computer Vision and Image Understanding*, 2014.

498. J. Son, S. Kim, and K. Sohn. Fast illumination-robust foreground detection using hierarchical distribution map for real-time video surveillance system. *Expert Systems With Applications*, page 4, 2016.

499. Y. Song, S. Noh, J. Yu, C. Park, and B. Lee. Background subtraction based on Gaussian mixture models using color and depth information. *International Conference on Control, Automation and Information Sciences, ICCAIS 2014*, pages 132–135, May 2014.

500. Y. Song, S. Noh, J. Yu, C. Park, and B. Lee. Background subtraction based on Gaussian mixture models using color and depth information. *International Conference on Control, Automation and Information Sciences, ICCAIS 2014*, pages 117–120, December 2014.

501. C. Spampinato, Y. Burger, G. Nadarajan, and R. Fisher. Detecting, tracking and counting fish in low quality unconstrained underwater videos. *VISAPP 2008*, pages 514–519, 2008.

502. C. Spampinato, G. Nadaraja, J. Chen-Burger, and R. Fisher. Detecting, tracking and counting fish in low quality unconstrained underwater videos. *International Conference on Computer Vision Theory and Applications, VISAPP 2008*, 2:514–520, 2008.

503. C. Spampinato, S. Palazzo, and D. Giordano. Kernel density estimation using joint spatial-color-depth data for background modeling. *International Conference on Pattern Recognition, ICPR 2014*, 2014.

504. C. Spampinato, S. Palazzo, and I. Kavasidis. A texton-based kernel density estimation approach for background modeling under extreme conditions. *Computer Vision and Image Understanding, CVIU 2014*, 122:74–83, 2014.

505. P. St-Charles and G. Bilodeau. Improving background subtraction using local binary similarity patterns. *IEEE Winter conference on Applications of Computer Vision, WACV 2014*, March 2014.

506. P. St-Charles, G. Bilodeau, and R. Bergevin. Flexible background subtraction with self-balanced local sensitivity. *IEEE Change Detection Workshop, CDW 2014*, June 2014.

507. P. St-Charles, G. Bilodeau, and R. Bergevin. SuBSENSE: a universal change detection method with local adaptive sensitivity. *IEEE Transactions on Image Processing*, 2014.

508. P. St-Charles, G. Bilodeau, and R. Bergevin. SuBSENSE: A Universal Change Detection Method with Local Adaptive Sensitivity. *Workshop on Change Detection in conjunction with CVPR 2014*, 2014.

509. P. St-Charles, G. Bilodeau, and R. Bergevin. A self-adjusting approach to change detection based on background word consensus. *IEEE Winter conference on Applications of Computer Vision, WACV 2015*, January 2015.

510. P. St-Charles, G. Bilodeau, and R. Bergevin. Universal background subtraction using word consensus models. *IEEE Transactions on Image Processing*, 25(10):4768–4781, October 2016.

511. U. Stanczyk and L. Jain. Feature selection for data and pattern recognition. *Recent research trends in feature selection for data and pattern recognition, Springer*, 2015.

512. C. Stauffer and E. Grimson. Adaptive background mixture models for real-time tracking. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 1999*, pages 246–252, 1999.

513. L. Di Stefano, F. Tombari, S. Mattoccia, and E. De Lisi. Robust and accurate change detection under sudden illumination variations. *International Workshop on Multi-dimensional and Multi-view Image Processing, ACCV 2007*, pages 103–109, 2007.

514. A. Stormer, M. Hofmann, and G. Rigoll. Depth gradient based segmentation of overlapping foreground objects in range images. *IEEE Conference on Information Fusion*, pages 1–4, 2010.

515. J. Stuckler and S. Behnke. Efficient Dense 3D Rigid-Body Motion Segmentation in RGB-D Video. *British Machine Vision Conference, BMVC 2013*, page 3, 2013.

516. B. Subudhi, S. Ghosh, S. Cho, and A. Ghosh. Integration of fuzzy Markov random field and local information for separation of moving objects and shadows. *Information Sciences*, 331:15–31, 2016.

517. B. Subudhi, S. Ghosh, S. Shiu, and A. Ghosh. Statistical feature bag based background subtraction for local change detection. *Information Sciences*, 2016.

518. J. Suhr, H. Jung, G. Li, and J. Ki. Mixture of gaussians-based background subtraction for bayer-pattern image sequences. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(3):365–370, March 2011.

519. T. Sun, S. Hsu, and C. Huang. A hybrid codebook background model for background subtraction. *IEEE Conference on Signal Processing Systems*, pages 96–101, 2011.

520. Y. Sun, B. Li, B. Yuan, Z. Miao, and C. Wan. Better foreground segmentation for static cameras via new energy form and dynamic graph-cut. *International Conference on Pattern Recognition, ICPR 2006*, pages 49–52, 2006.

521. G. Szwoch and P. Dalka. Layered background modeling for automatic detection of unattended objects in camera images. *International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2011*, April 2011.

522. K. Takahara, T. Tori, and T. Zin. Making background subtraction robust to various illumination changes. *International Journal of Computer Science and Network Security, IJCSNS 2011*, 11(3):241–248, March 2011.

523. W. Tan, B. Li, and W. Zhang. Research on background subtraction based on center-symmetric local binary patterns. *Journal of University of Science and Technology of China*, 40(11):1112–1116, 2011.

524. X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Transactions on Image Processing*, 19(6), June 2010.

525. T. Tanaka, A. Shimada, D. Arita, and R. Taniguchi. A fast algorithm for adaptive background model construction using parzen density estimation. *IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2007*, 2007.

526. T. Tanaka, A. Shimada, D. Arita, and R. Taniguchi. Object detection under varying illumination based on adaptive background modeling considering spatial locality. *International Workshop on Computer Vision, MIRU 2008*, July 2008.

527. T. Tanaka, A. Shimada, D. Arita, and R. Taniguchi. Object segmentation under varying illumination based on combinational background modeling. *Joint Workshop on Machine Perception and Robotics, MPR 2008*, 2008.

528. T. Tanaka, A. Shimada, D. Arita, and R. Taniguchi. Object segmentation based on adaptive background model considering spatio-temporal features. *Pattern Recognition and Media Understanding, PRMU 2009*, March 2009.

529. T. Tanaka, A. Shimada, D. Arita, and R. Taniguchi. Object segmentation under varying illumination: stochastic background model considering spatial locality. *Progress in Informatics*, pages 21–31, 2010.

530. T. Tanaka, A. Shimada, R. Taniguchi, T. Yamashita, and D. Arita. Towards robust object detection: integrated background modeling based on spatio-temporal features. *Asian Conference on Computer Vision, ACCV 2009*, September 2009.

531. F. Tang, S. Lim, N. Chang, and H. Tao. A novel feature descriptor invariant to complex brightness changes. *International Conference on Pattern Recognition, CVPR 2009*, 2009.

532. P. Tang, L. Gao, and Z. Liu. Salient moving object detection using stochastic approach filtering. *International Conference on Image and Graphics, ICIG 2007*, pages 530–535, 2007.

533. S. Tang, B. Wang, Y. Zhao, X. Hu, and Y. Gong. Efficient background model based on multi-level feedback for video surveillance. *International Conference on Digital Image Processing, ICDIP 2016*, May 2016.

534. H. Tezuka and T. Nishitani. A precise and stable foreground segmentation using fine-to-coarse approach in transform domain. *International Conference on Image Processing, ICIP 2008*, pages 2732–2735, October 2008.

535. H. Tezuka and T. Nishitani. Multiresolutional gaussian mixture model for precise and stable foreground segmentation in transform domain. *IEICE Transaction on Fundamentals*, E92-A(3):772–779, March 2009.

536. H. Tezuka and T. Nishitani. Precise object contour detection using variable block-based video foreground segmentation in transform domain. *International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, January 2009.

537. A. Thangarajah, Q. Wu, and J. Huo. A unified threshold updating strategy for multivariate Gaussian mixture based moving object detection. *International Conference on High Performance Computing and Simulation, HPCS 2016*, pages 570–574, 2016.

538. A. Thangarajah, Q. Wu, A. Singh, B. Mandon, and A. Chowdhury. Video foreground detection in non-static background using multi-dimensional color space. *International Conference on Eco-friendly Computing and Communication Systems*, 70:55–61, 2015.

539. T. Thongkamwitoon, S. Aramvith, and T. Chalidabhongse. An adaptive real-time background subtraction and moving shadows detection. *IEEE International Conference on Multimedia and Expo, ICME 2004*, June 2004.

540. T. Thongkamwitoon, S. Aramvith, and T. Chalidabhongse. Non-linear learning factor control for statistical adaptive background subtraction algorithm. *International Symposium on Circuits and Systems, ISCAS 2005*, May 2005.

541. L. Tong, F. Dai, D. Zhang, D. Wang, and Y. Zhang. Encoder combined video moving object detection. *Neurocomputing*, 139:150–162, September 2014.

542. D. Toth and T. Aach. Detection and recognition of moving objects using statistical motion detection and fourier descriptors. *International Conference on Image Analysis and Processing, ICIAP 2003*, pages 430–435, 2003.

543. D. Toth, T. Aach, and V. Metzler. Bayesian spatio-temporal motion detection under varying illumination. *European Signal Processing Conference, EUSIPCO 2000*, pages 2081–2084, 2000.

544. D. Toth, T. Aach, and V. Metzler. Illumination-invariant change detection. *IEEE Southwest Symposium on Image Analysis and Interpretation, SSIAI 2000*, pages 3–7, April 2000.

545. K. Toyama, J. Krumm, B. Brumiit, and B. Meyers. Wallflower: Principles and practice of background maintenance. *International Conference on Computer Vision*, pages 255–261, September 1999.

546. R. Trabelsi, I. Jabri, F. Smach, and A. Bouallegue. Efficient and Fast Multi-Modal Foreground-Background Segmentation using RGB-D Data. *Pattern Recognition Letters*, 2017.

547. D. Tsai and W. Chiu. Motion detection using Fourier image reconstruction. *Pattern Recognition Letters*, 29:2145–2155, 2008.

548. G. Tu, H. Karstoft, L. Pedersen, and E. Jorgensen. Illumination and reflectance estimation with its application in foreground detection. *Sensors 2015*, September 2015.

549. Q. Tu, Y. Xu, and M. Zho. Box-based codebook model for real-time objects detection. *World Congress on Intelligent Control and Automation, WCICA 2008*, pages 7621–7625, June 2008.

550. O. Tuzel, F. Porikli, and P. Meer. Region covariance: A fast descriptor for detection and classification. *European Conference on Computer Vision, ECCV 2006*, pages 589–600, 2006.

551. L. Unzueta, M. Nieto, A. Cortes, J. Barandiaran, O. Otaegui, and P. Sanchez. Adaptive multi-cue background subtraction for robust vehicle counting and classification. *IEEE Transactions on Intelligent Transportation Systems*, 2011.

552. A. Vacavant, T. Chateau, A. Wilhelm, and L. Lequievre. A benchmark dataset for foreground/background extraction. *International Workshop on Background Models Challenge, ACCV 2012*, November 2012.

553. S. Varadarajan, P. Miller, and H. Zhou. Spatial mixture of Gaussians for dynamic background modelling. *IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2013*, pages 63–68, 2013.

554. S. Varadarajan, P. Miller, and H. Zhou. Region-based mixture of gaussians modelling for foreground detection in dynamic scenes. *Pattern Recognition*, 48(11):3488–3503, 2015.

555. S. Varma and M. Sreeraj. Hybrid background subtraction in video using bi-level codebook model. *IEEE International Conference on the Applications of Digital Information and Web Technologies, ICADIWT 2014*, February 2014.

556. J. Vijverberg, M. Loomans, C. Koeleman, and P. de With. Global illumination compensation for background subtraction using gaussian-based background difference modeling. *International Conference on Advanced Video and Signal-based Surveillance, AVSS 2009*, pages 448–453, September 2009.

557. P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *Computer Vision and Pattern Recognition*, 2001.

558. B. Vishnyakov, V. Gorbatsevich, S. Sidyakin, Y. Vizilter, I. Malin, and A. Egorov. Fast moving objects detection using iLBP background model. *International Archives on Photogrammetry and Remote Sensing*, pages 347–350, 2014.

559. B. Vishnyakov, S. Sidyakin, and Y. Vizilter. Diffusion background model for moving objects detection. *International Archives on Photogrammetry and Remote Sensing*, May 2015.

560. N. Wajima, S. Takahashi, M. Ito, Y. Satoh, and S. Kaneko. Object detection based on radial reach filter under the change of background. *SPIE International Symposium on Optomechatronic Technologies*, pages 1–10, 2005.

561. B. Wang and P. Dudek. AMBER: Adapting Multi-Resolution Background Extractor. *IEEE International Conference on Image Processing, ICIP 2013*, 2013.

562. B. Wang, Y. Liu, W. Xu, W. Wang, and M. Zhang. Background subtraction using spatiotemporal condition information. *Optik*, page 14061411, 2014.

563. H. Wang and P. Miller. Regularized online mixture of Gaussians for background subtraction. *Advanced Video and Signal-Based Surveillance, AVSS 2011*, pages 249–254, 2003.

564. H. Wang, Q. Wang, Y. Li, Y. Liu, and Y. Dai. An illumination-robust algorithm based on visual background extractor for moving object detection. *Asian Control Conference, ASCC 2015*, pages 1–6, 2015.

565. J. Wang, F. He, X. Zhang, and Y. Gao. A moving objects detection algorithm using iterative division and Gaussian mixture model. *International Conference on Advanced Computer and Control, ICACC 2010*, 5:229–233, 2010.

566. J. Wang and Y. Yagi. Efficient background subtraction under abrupt illumination variations. *Asian Conference on Computer Vision, ACCV 2012*, November 2012.

567. K. Wang, Y. Liu, C. Gou, and F. Wang. A multi-view learning approach to foreground detection for traffic surveillance applications. *IEEE Transactions on Vehicular Technology*, 65(6):4144–4158, June 2016.

568. L. Wang and C. Pan. Fast and effective background subtraction based on ELBP. *International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2010*, March 2010.

569. L. Wang and C. Pan. Effective multi-resolution background subtraction. *International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2011*, pages 909–912, 2011.

570. L. Wang, H. Wu, and C. Pan. Adaptive ELBP for background subtraction. *Asian Conference on Computer Vision, ACCV 2010*, pages 560–571, November 2010.

571. R. Wang, F. Bunyak, G. Seetharaman, and K. Palaniappan. Static and moving object detection using flux tensor with split Gaussian models. *IEEE Workshop on Change Detection in conjunction with CVPR 2014, CDW 2014*, 2014.

572. S. Wang, Z. Wang, and R. Hu. Surveillance video synopsis in the compressed domain for fast video browsing. *Journal of Visual Communication and Image Representation*, pages 1431–1442, November 2013.

573. T. Wang, J. Liang, X. Wang, and S. Wang. Background modeling using local binary patterns of motion vector. *IEEE Visual Communications and Image Processing, VCIP 2012*, pages 1–5, November 2012.

574. W. Wang, D. Chen, W. Gao, and J. Yang. Modeling background from compressed video. *International Conference on Computer Vision, ICCV 2005*, 2005.

575. W. Wang, W. Gao, and R. Wang. A local hierarchical approach for background modeling and moving targets detection. *MIPPR 2009*, 2009.

576. W. Wang and R. Wu. Fusion of luma and chroma GMMs for HMM-based object detection. *First Pacific Rim Symposium on Advances in Image and Video Technology, PSIVT 2006*, pages 573–581, December 2006.

577. W. Wang, J. Yang, and W. Gao. Modeling background and segmenting moving objects from compressed video. *IEEE Transactions on Circuits and Systems for Video Technology*, 18:670–681, 2008.

578. X. Wang, F. Liu, and Z. Ye. Background modeling in compressed sensing scheme. *ESEP 2011*, 13(4776-4783), December 2011.

579. X. Wang and W. Wan. Motion segmentation via multi-task robust principal component analysis. *Journal of Applied Sciences, Electronics and Information Engineering*, 32(5):473–480, September 2014.

580. Y. Wang, K. Loe, and J. Wu. A dynamic conditional random field model for foreground and shadow segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):279–289, February 2006.

581. Y. Wang, Q. Lu, D. Wang, and W. Liu. Compressive background modeling for foreground extraction. *Journal of Electrical and Computer Engineering, Hindawi Publishing Corporation*, March 2015.

582. Y. Wang, Z. Luo, and P. Jodoin. Interactive deep learning method for segmenting moving objects. *Pattern Recognition Letters, Special Issue on Scene Background Modeling and Initialization*, 2016.

583. G. Ward. The radiance lighting simulation and rendering system. *Computer Graphics, SIGGRAPH 1994*, pages 459–472, July 1994.

584. G. Warnell, S. Bhattacharya, R. Chellappa, and T. Basar. Adaptive-rate compressive sensing via side information. *IEEE Transactions on Image Processing*, 24(11):3846–3857, November 2015.

585. G. Warnell, D. Reddy, and R. Chellappa. Adaptive rate compressive sensing for background subtraction. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, March 2012.

586. A. Waters, A. Sankaranarayanan, and R. Baraniuk. SpaRCS: recovering low-rank and sparse matrices from compressive measurements. *Neural Information Processing Systems, NIPS 2011*, December 2011.

587. A. Waters, A. Sankaranarayanan, and R. Baraniuk. SpaRCS: recovering low-rank and sparse matrices from compressive measurements. *Technical Report*, 2011.

588. Z. Wei, S. Jiang, and Q. Huang. A pixel-wise local information-based background subtraction approach. *International Conference on Multimedia and Expo, ICME 2008*, pages 1501–1504, April 2008.

589. C. Wren and A. Azarbayejani. Pfinder : Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780–785, July 1997.

590. C. Wren and A. Azarbayejani. Pfinder : Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780–785, July 1997.

591. C. Wren and F. Porikli. Waviz: Spectral similarity for object detection. *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, PETS 2005*, January 2005.

592. H. Wu, N. Liu, X. Luo, J. Su, and L. Chen. Real-time background subtraction-based video surveillance of people by integrating local texture patterns. *Signal, Image and Video Processing*, pages 665–676, 2014.

593. J. Wu and X. Zu. Research on background subtraction based on center-symmetric local binary patterns. *International Symposium on Instrumentation and Measurement, Sensor Network and Automation, IMSNA 2013*, pages 491–484, December 2013.

594. Q. Wu and B. Jeng. Background subtraction based on logarithmic intensities. *Pattern Recognition Letters*, 23(13):1529–1536, November 2002.

595. Y. Wu, D. Zeng, and H. Li. Layered Video Objects Detection Based on LBP and Codebook. *International Workshop on Education Technology and Computer Science, ETCS 2009*, 1:207–213, March 2009.

596. Y. Wu, D. Zeng, and H. Li. Layered Video Objects Detection Based on LBP and Codebook. *International Journal on Modern Education and Computer Science, IJMECS 2009*, 1:1–8, 2009.

597. Y. Wu, D. Zeng, and H. Li. Layered Video Objects Detection Based on LBP and Codebook. *International Workshop on Education Technology and Computer Science, ETCS 2009*, 1:207–213, March 2009.

598. B. Xie, V. Ramesh, and T. Boult. Sudden illumination change detection using order consistency. *Image and Vision Computing*, 22(2):117–125, February 2004.

599. Q. Xiong and C. Jaynes. Multi-resolution background modeling of dynamic scenes using weighted match filters. *ACM International Workshop on Video Surveillance and Sensor Networks, VSSN 2004*, 2004.

600. J. Xu, X. Ding, S. Wang, and Y. Wu. Background subtraction based on a combination of local texture and color. *Acta Automatica Sinica*, 35(9), September 2009.

601. M. Xu and T. Ellis. Illumination-invariant motion detection using color mixture models. *British Machine Vision Conference, BMVC 2001*, pages 163–172, September 2001.

602. M. Xu and J. Lu. K-cluster-valued compressive sensing for imaging. *EURASIP Journal on Advances in Signal Processing*, 2011.

603. P. Xu, M. Ye, X. Li, Q. Liu, Y. Yang, and J. Ding. Dynamic background learning through deep auto-encoder networks. *ACM International Conference on Multimedia*, November 2014.

604. P. Xu, M. Ye, X. Li, Q. Liu, Y. Yang, and J. Ding. Motion detection via a couple of auto-encoder networks. *International Conference on Multimedia and Expo, ICME 2014*, 2014.

605. Y. Xu. Moving object segmentation by pursuing local spatio-temporal manifolds. *PhD Thesis*, 2013.

606. G. Xue, L. Song, J. Sun, and M. Wu. Hybrid center-symmetric local pattern for dynamic background subtraction. *International Conference in Multimedia and Exposition, ICME 2011*, July 2011.

607. G. Xue, J. Sun, and L. Song. Background subtraction based on phase and distance transform under sudden illumination change. *IEEE International Conference on Image Processing, ICIP 2010*, 2010.

608. G. Xue, J. Sun, and L. Song. Dynamic background subtraction based on spatial extended center-symmetric local binary pattern. *International Conference in Multimedia and Exposition, ICME 2010*, pages 1050–1055, July 2010.

609. G. Xue, J. Sun, and L. Song. Background subtraction based on phase feature and distance transform. *Pattern Recognition Letters*, 33(12):1601–1613, September 2012.

610. D. Yadav and K. Sing. Moving object detection for visual surveillance using quasi-euclidian distance. *IC3T 2015*, 3:225–233, September 2015.

611. D. Yadav and K. Sing. A combined approach of Kullback-Leibler divergence and background subtraction for moving object detection in thermal video. *Infrared Physics and Technology*, 76:21–31, May 2016.

612. D. Yan, Q. Yu, and M. Wang. Real-time objects detection using layered codebook model. *Advanced Materials Research*, pages 694–697, May 2013.

613. J. Yan, S. Wang, T. Xie, Y. Yang, and J. Wang. Variational bayesian learning for background subtraction based on local fusion feature. *IET Computer Vision*, October 2016.

614. Y. Yan, J. Ren, H. Zhao, G. Sun, Z. Wang, J. Zheng, S. Marshall, and J. Soraghan. Cognitive fusion of thermal and visible imagery for effective detection and tracking of pedestrians in videos. *Cognitive Computation*, December 2017.

615. F. Yang, H. Jiang, Z. Shen, W. Deng, and D. Metaxas. Adaptive low rank and sparse decomposition of video using compressive sensing. *IEEE International Conference on Image Processing, ICIP 2013*, 2013.

616. J. Yang, J. Wang, and H. Lu. A hierarchical approach for background modeling and moving objects detection. *International Journal of Control, Automation and Systems*, pages 940–947, October 2010.

617. J. Yang, S. Wang, Z. Lei, Y. Zhao, and S. Li. Spatial-temporal LBP-based moving object segmentation in compressed domain. *IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2012*, 2012.

618. L. Yang, H. Cheng, J. Su, and X. Chen. Pixel-to-model background modeling in crowded scenes. *IEEE International Conference on Multimedia and Expo, ICME 2014*, pages 1–6, 2014.

619. L. Yang, H. Cheng, J. Su, and X. Li. Pixel-to-model distance for robust background reconstruction. *IEEE Transactions on Circuits and Systems for Video Technology*, April 2015.

620. M. Yang, C. Huang, W. Liu, S. Lin, and K. Chuang. Binary descriptor based nonparametric background modeling for foreground extraction using detection theory. *IEEE Transactions on Circuits and Systems for Video Technology*, 2014.

621. R. Yang, Z. Wang, P. Heng, and K. Leung. Fuzzified Choquet integral with a fuzzy-valued integrand and its application on temperature prediction. *IEEE Transactions on Systems, Man, and Cybernetics*, 38(2):367–380, 2008.

622. S. Yang and C. Hsu. Background modeling from GMM likelihood combined with spatial and color coherency. *ICIP 2006*, pages 2801–2804, 2006.

623. T. Yang, S. Li, Q. Pan, and J. Li. Real-time and accurate segmentation of moving objects in dynamic scene. *ACM International Workshop on Video Surveillance and Sensor Networks, VSSN 2004*, October 2004.

624. W. Yang, L. Dou, and J. Zhang. A multi-histogram clustering approach toward markov random field for foreground segmentation. *International Journal of Image and Graphics*, October 2013.

625. Y. Yang and M. Levine. The background primal sketch: An approach for tracking moving objects. *International Conference on Machine Visiona and Applications*, 5(4):17–34, December 1992.

626. G. Yao, T. Lei, J. Zhong, P. Jiang, and W. Jia. Comparative evaluation of background subtraction algorithms in remote scene videos captured by MWIR sensors. *MDPI Sensors*, 2017.

627. C. Yeh, C. Lin, K. Muchtar, and L. Kang. Color invariant background modeling for moving objects detection. *IPPR Computer Vision, Graphics, CVGIP 2012*, August 2012.

628. H. Yin, H. Yang, H. Su, and C. Zhan. Dynamic background subtraction based on appearance and motion pattern. *International Conference on Multimedia and Expo Workshops, ICMEW 2013*, pages 1–6, July 2013.

629. K. Yokoi. Illumination-robust change detection using texture based features. *Meeting on Image Recognition and Understanding, MIRU 2006*, July 2006.

630. K. Yokoi. Illumination-robust change detection using texture based features. *Machine Vision Applications, MVA 2007*, pages 320–327, May 2007.

631. K. Yokoi. Probabilistic BPRRC: robust change detection against illumination changes and background movements. *Conference on Machine Vision Applications, MVA 2009*, pages 148–151, May 2009.

632. S. Yoo and C. Kim. Background subtraction using hybrid feature coding in the bag-of-features framework. *Pattern Recognition Letters*, 34(16):2086–2093, 2013.

633. S. Yoshinaga, A. Shimada, H. Nagahara, and R. Taniguchi. Object detection using local difference patterns. *Asian Conference on Computer Vision, ACCV 2010*, pages 216–227, November 2010.

634. S. Yoshinaga, A. Shimada, H. Nagahara, and R. Taniguchi. Statistical local difference pattern for background modeling. *IPSJ Transactions on Computer Vision and Applications*, 3:198–210, 2011.

635. S. Yoshinaga, A. Shimada, H. Nagahara, and R. Taniguchi. Background model based on statistical local difference pattern. *International Workshop on Background Models Challenge, ACCV 2012*, pages 327–332, November 2012.

636. G. Yuan, Y. Gao, D. Xu, and M. Jiang. A new background subtraction method using texture and color information. *ICIC 2011*, pages 541–558, 2011.

637. Y. Yue, Z. An, and H. Wu. Adaptive targets-detecting algorithm based on LBP and background modeling under complex scenes. *Procedia Engineering*, 15:2489–2494, December 2011.

638. J. Yuk and K. Wong. An efficient pattern-less background modeling based on scale invariant local states. *IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2011*, 2011.

639. R. Yumiba, M. Miyoshi, and H. Fujiyoshi. Moving object detection with background model based on spatio-temporal texture. *IEEE Winter Vision Meetings, WACV 2011*, pages 352–359, January 2011.

640. R. Yumiba, M. Miyoshi, H. Fujiyoshi, and Y. Murai. Moving object detection with background model based on spatio-temporal texture. *Journal*, J94-D:1101–1112, 2011.

641. A. Zaharescu and M. Jamieson. Multi-scale multi-feature codebook-based background subtraction. *IEEE International Conference on Computer Vision Workshops, ICCV 2011*, pages 1753–1760, November 2011.

642. Q. Zang and R. Klette. Robust background subtraction and maintenance. *International Conference on Pattern Recognition, ICPR 2004*, pages 90–93, 2004.

643. V. Zeljkovic, D. Pokrajac, and L. Latecki. Noise-robust spatial-temporal algorithm for moving object detection. *ETRAN 2005*, 2005.

644. D. Zeng, M. Zhu, T. Zhou, F. Xu, and H. Yang. Robust background subtraction via the local similarity statistical descriptor. *MDPI Applied Sciences*, 2017.

645. W. Zeng, W. Gao, and D. Zhao. Automatic moving object extraction in MPEG video. *International Symposium on Circuits and Systems*, pages 524–527, 2003.

646. Z. Zeng and J. Jia. Arbitrary cylinder color model for the codebook based background subtraction. *Optics Express*, 22(18):577–588, 2014.

647. Z. Zeng, J. Jia, Z. Zhu, and D. Yu. Adaptive maintenance scheme for codebook-based dynamic background subtraction. *Computer Vision and Image Understanding*, 152:5866, 2016.

648. B. Zhang, Y. Gao, S. Zhao, and B. Zhong. Kernel similarity modeling of texture pattern flow for motion detection in complex background. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(1):29–38, January 2011.

649. B. Zhang, N. Lin, and H. Zheng. Motion detection based on directional rectangular pattern and adaptive threshold propagation in the complex background. *Chinese Conference on Pattern Recognition, CCPR 2009*, pages 1–5, 2009.

650. B. Zhang, B. Zhong, and Y. Cao. Complex background modeling based on texture pattern flow with adaptive threshold propagation. *Journal Of Visual Communication And Image Representation*, pages 516–521, 2011.

651. D. Zhang and S. Chen. Kernel-based fuzzy and possibilistic c-means clustering. *International Conference on Artificial Neural Networks*, pages 122–125, 2003.

652. D. Zhang and Z. Wang. Fuzzy integrals of fuzzy-valued functions. *Fuzzy Sets Systems*, 54(1):63–67, 1993.

653. E. Zhang, Y. Li, and J. Duan. Moving object detection based on confidence factor and CSLBP features. *Imaging Science Journal*, May 2016.

654. H. Zhang and D. Xu. Fusing color and gradient features for background model. *International Conference on Signal Processing, ICSP 2006*, 2(7), 2006.

655. H. Zhang and D. Xu. Fusing color and texture features for background model. *International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2006*, 4223(7):887–893, September 2006.

656. J. Zhang. Highway background identification and background modeling based on projection statistics. *Global Applications of Pervasive and Ubiquitous*, 2012.

657. J. Zhang, L. Zhang, and H. Tai. Efficient video object segmentation using adaptive background registration and edge-based change detection techniques. *ICME 2004*, 2004.

658. S. Zhang, T. Jiang, Y. Peng, and X. Peng. A new pixel-level background subtraction algorithm in machine vision. *International Conference on Intelligent Robotics and Applications, ICIRA 2017*, pages 520–531, 2017.

659. S. Zhang, H. Yao, and S. Liu. Dynamic background and subtraction using spatio-temporal local binary patterns. *IEEE International Conference on Image Processing, ICIP 2008*, pages 1556–1559, October 2008.

660. S. Zhang, H. Yao, and S. Liu. Dynamic background subtraction based on local dependency histogram. *International Workshop on Visual Surveillance, VS2008*, 2008.

661. S. Zhang, H. Yao, and S. Liu. Dynamic background subtraction based on local dependency histogram. *International Journal of Pattern Recognition and Artificial Intelligence, IJPRAI 2009*, 23(7), 2009.

662. S. Zhang, H. Yao, and S. Liu. Spatialtemporal nonparametric background subtraction in dynamic scenes. *International Conference on Multimedia and Expo, ICME 2009*, pages 518–521, 2009.

663. S. Zhang, H. Yao, S. Liu, X. Chen, and W. Gao. Covariance-based method for dynamic background subtraction. *IEEE International Conference on Pattern Recognition, ICPR 2008*, pages 1–4, December 2008.

664. W. Zhang, F. Chen, W. Xu, and E. Zhang. Real-time video intelligent surveillance system. *ICME 2006*, pages 1021–1024, 2006.

665. X. Zhang, T. Huang, Y. Tian, and W. Gao. Background-modeling-based adaptive prediction for surveillance video coding. *IEEE Transactions on Image Processing*, 23(2):769–784, 2014.

666. X. Zhang, Y. Tian, T. Huang, S. Dong, and W. Gao. Optimizing the hierarchical prediction and coding in HEVC for surveillance and conference videos with background modeling. *IEEE Transactions on Image Processing*, October 2014.

667. Y. Zhang, X. Li, Z. Zhang, F. Wu, and L. Zhao. Deep learning driven blockwise moving object detection with binary scene modeling. *Neurocomputing*, 2015.

668. Z. Zhang, C. Wang, B. Xiao, S. Liu, and W. Zhou. Multi-scale fusion of texture and color for background modeling. *AVSS 2012*, pages 154–159, 2012.

669. Z. Zhang, C. Wang, B. Xiao, S. Liu, and W. Zhou. Multi-scale fusion of texture and color for background modeling. *International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2012*, September 2012.

670. Z. Zhang, B. Xiao, C. Wang, W. Zhou, and S. Liu. Background modeling by exploring multi-scale fusion of texture and intensity in complex scenes. *Asian Conference on Pattern Recognition, ACPR 2011*, pages 402–406, November 2011.

671. C. Zhao, T. Zhang, Q. Huang, X. Zhang, D. Yang, Y. Qu, and S. Huang. Background subtraction based on superpixels under multi-scale in complex scenes. *Chinese Conference on Pattern Recognition, CCPR 2016*, October 2016.

672. G. Zhao, T. Ahonen, J. Matas, and M. Pietikainen. Rotation-invariant image and video description with local binary pattern features. *IEEE Transactions on Image Processing*, 21(4):1465–1467, 2012.

673. L. Zhao, Y. Tian, and T. Huang. Background-foreground division based search for motion estimation in surveillance video coding. *International Conference on Multimedia and Expo, ICME 2014*, 2014.

674. L. Zhao, X. Zhang, Y. Tian, R. Wang, and T. Huang. A background proportion adaptive Lagrange multiplier selection method for surveillance video on high HEVC. *International Conference on Multimedia and Expo, ICME 2013*, July 2013.

675. P. Zhao, Y. Zhao, and A. Cai. Hierarchical codebook background model using Haar-like features. *IEEE International Conference on Network Infrastructure and Digital Content, IC-NIDC 2012*, pages 438–442, 2012.

676. Q. Zhao and H. Tao. Object tracking using color correlogram. *IEEE Workshop on Visual Surveillance and Performance Evaluation, VS-PETS in conjunction with ICCV*, pages 263–270, October 2015.

677. X. Zhao, Y. Satoh, H. Takauji, S. Kaneko, K. Iwata, and R. Ozakie. Object detection based on a robust and accurate statistical multi-point-pair model. *Pattern Recognition*, 2011.

678. Y. Zhao, X. Song, and Y. Jia. On the dimensionality of video bricks under varying illumination. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2012*, pages 222–229, June 2012.

679. Z. Zhao, T. Bouwmans, X. Zhang, and Y. Fang. A fuzzy background modeling approach for motion detection in dynamic backgrounds. *International Conference on Multimedia and Signal Processing*, pages 772–781, December 2012.

680. A. Zheng, M. Xu, H. Shi, B. Luo, and C. Li. CLASS: Collaborative Low-rank and Sparse Separation for Moving Object Detection. *Cognitive Computation*, pages 1–14, February 2017.

681. B. Zhong, Y. Chen, Y. Chen, R. Ji, Y. Chen, D. Chen, and H. Wang. Background subtraction driven seeds selection for moving objects segmentation and matting. *Neurocomputing*, 103:132–142, March 2013.

682. B. Zhong, X. Hong, H. Yao, S. Shan, X. Chen, and W. Gao. Texture and motion pattern fusion for background subtraction. *Joint Conference on Information Sciences*, pages 1–7, 2008.

683. B. Zhong, S. Liu, H. Yao, and B. Zhang. Multi-resolution background subtraction for dynamic scenes. *International Conference on Image Processing, ICIP 2009*, pages 3193–3196, 2009.

684. B. Zhong, H. Yao, and S. Li. Neighboring image patches embedding for background modeling. *International Conference on Image Processing, ICIP 2009*, 2009.

685. B. Zhong, H. Yao, and S. Liu. Robust background modeling via standard variance feature. *International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2010*, pages 1182–1185, 2010.

686. B. Zhong, H. Yao, S. Shan, X. Chen, and W. Gao. Hierarchical background subtraction using local pixel clustering. *IEEE International Conference on Pattern Recognition, ICPR 2008*, December 2008.

687. B. Zhong, H. Yao, and X. Yuan. Local histogram of figure/ground segmentations for dynamic background subtraction. *EURASIP Journal on Advances in Signal Processing*, 2010:14, 2010.

688. G. Zhong, L. Wang, and J. Dong. An overview on data representation learning: From traditional feature learning to recent deep learning. *Preprint*, 2016.

689. Q. Zhong, L. Dai, Y. Song, and R. Wang. A hierarchical motion detection algorithm with the fusion of the two types of motion information. *Pattern Recognition Artificial Intelligence*, 18(5):552–557, 2005.

690. D. Zhou and H. Zhang. Modified GMM background modeling and optical flow for detection of moving objects. *IEEE International Conference on Systems, Man and Cybernetics, SMC 2005*, pages 2224–2229, October 2005.

691. H. Zhou, Z. Zeng, and J. Zhou. Motion detection with background clutter suppression based on KDE model. *International Conference on Industrial Instrumentation and Control, ICIC 2008*, pages 466–473, 2008.

692. W. Zhou, Y. Liu, W. Zhang, L. Zhuang, and N. Yu. Dynamic background subtraction using spatial-color binary patterns. *International Conference on Image and Graphics, ICIG 2011*, August 2011.

693. J. Zhu, S. Schwartz, and B. Liu. A transform domain approach to real-time foreground segmentation in video sequences. *International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2005*, 2005.

694. M. Zitouni, H. Bhaskar, and M. Al-Mualla. Robust background modeling and foreground detection using dynamic textures. *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISAPP 2016*, 2016.

695. Z. Zivkovic. Improved adaptive Gaussian mixture model for background subtraction. *International Conference Pattern Recognition, ICPR 2004*, pages 28–31, 2004.

696. D. Zonoobi and A. Kassim. Lowrank and sparse matrix reconstruction with partial support knowledge for surveillance video processing. *International Conference on Image Processing, ICIP 2013*, September 2013.