IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 1, 2019, accepted April 24, 2019, date of publication May 6, 2019, date of current version May 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2914961

A Comprehensive Survey of Video Datasets for Background Subtraction

RUDRIKA KALSOTRA AND SAKSHI ARORA
Department of Computer Science and Engineering, Shri Mata Vaishno Devi University, Katra 182320, India
Corresponding author: Sakshi Arora (sakshi@smvdu.ac.in)

ABSTRACT Background subtraction is an effective method of choice when it comes to detection of moving objects in videos and has been recognized as a breakthrough for the wide range of applications of intelligent video analytics (IVA). In recent years, a number of video datasets intended for background subtraction have been created to address the problem of large realistic datasets with accurate ground truth. The use of these datasets enables qualitative as well as quantitative comparisons and allows benchmarking of different algorithms. Finding the appropriate dataset is generally a cumbersome task for an exhaustive evaluation of algorithms. Therefore, we systematically survey standard video datasets and list their applicability for different applications. This paper presents a comprehensive account of public video datasets for background subtraction and attempts to cover the lack of a detailed description of each dataset. The video datasets are presented in chronological order of their appearance. Current trends of deep learning in background subtraction along with top-ranked background subtraction methods are also discussed in this paper. The survey introduced in this paper will assist researchers of the computer vision community in the selection of appropriate video dataset to evaluate their algorithms on the basis of challenging scenarios that exist in both indoor and outdoor environments.

INDEX TERMS Background model, background subtraction, challenges, datasets, deep neural networks, foreground, intelligent video analytics (IVA), video frames.

I. INTRODUCTION

Integration of advanced camera technology and intelligent video analytics (IVA) has sparked a growing interest among researchers in the area of automated video surveillance. However, such technological advancement results in an increase in the number, complexity, and size of surveillance videos and this, in turn, entail the need for new algorithms to handle huge video data efficiently and effectively. Background subtraction aims to detect foreground regions that are in motion from background of a video sequence and is a prerequisite of many intelligent video analytics (IVA) applications [1] such as automated video surveillance [2]–[5], optical motion capture [6], [7], computational imaging [8], [9], video inpainting [10]–[12], target tracking [13]–[15], video coding [16], [17], and human-machine interaction [18]. Since the 1990s, researchers have been exploring this field on the subject of different applications. Although, there are significant publications on background subtraction, currently most efficacious algorithms result in false detections under complex situations [19]. A quick search for background subtraction on IEEE Xplore shows 2430 publications in the last ten years (2008-2018). There are several extensive surveys on background subtraction in the literature. For instance, Bouwmans [20] surveyed different background modeling approaches for foreground detection and briefly discussed 9 traditional and 8 recent video datasets for background subtraction. The author also outlined the challenges and applications of background subtraction. In another work, Bouwmans and Garcia-Garcia [173] reviewed real-time applications, current models and challenges of background subtraction. Many researchers evaluated the state-of-the-art methods on different video datasets for comparative analysis of background subtraction algorithms [21]–[24], [25], [27].

In recent years, new video datasets devoted to background subtraction have been increasingly added to the list of background subtraction video datasets to address the problem of a large realistic dataset for the evaluation of different algorithms. These datasets are accessible to the research community and help in the application of different methods on same dataset for comparative analysis. This survey provides

The associate editor coordinating the review of this manuscript and approving it for publication was Yan-Jun Liu.

the complete description of publicly available video datasets in this field by stating the number of video sequences, total video frames, type of video cameras used, kind of annotations, nature of the environment and video categories under specific dataset. The problems of change detection, motion detection, and foreground extraction are closely associated with background subtraction. The process of background subtraction hinges on effective background modeling strategies and generally occurs in three steps [26], [27]: background initialization, foreground detection, and background maintenance.

Video frames are initialized at background initialization step to create background model [28], [29]. The next step, foreground detection, is to compare each incoming video frame with the background model to distinguish between two types of pixels: foreground pixels and background pixels. At the background maintenance stage, the background model is updated in order to acclimatize new information from video scene into the model. Table 1 summarizes the challenges of background subtraction considering three factors: background, foreground (moving objects), and camera. Video sequences in the dataset are structured into different categories that are typically based on the challenges of background subtraction listed below:

TABLE 1. Challenges of Background Subtraction.

Background							Foreground						Camera		
Illumination Changes	Dynamic Background	Shadows	Challenging Weather	Bootstrapping	Moved Background Objects	Inserted Background Objects	Shadows	Camouflage	Intermittent Object Motion	Occlusion	Foreground Aperture	Sleeping Foreground	Video Noise	Moving Camera	Camera Jitter

A. ILLUMINATION CHANGES

The illumination changes affect the pixels in the video scene and interrupt background model [30]. Gradual illumination changes in an outdoor environment due to variation in light intensity during day time generate an erroneous classification of pixels. There are sudden illumination changes that produce fallacious foreground detection such as switching on/off lights in an indoor scene or fluctuations in an outdoor scene due to the fast transition from a bright sunny day to cloudy.

B. DYNAMIC BACKGROUND

There are some periodical or irregular movements in an outdoor as well as indoor scene and is a challenging task to handle [31]. Dynamic background such as waving trees, fluttering leaves, swing fountains, swinging of pendulum, moving escalator, and swaying curtains lead to detection of uninterested objects [44].

C. SHADOWS

The shadows cast by moving objects or fixed background structures misinterpreted as foreground regions and also result into object merging and object distortion [30], [32], [33].

D. VIDEO NOISE

Poor quality sensors and compression artifacts degrade video signals and add faulty pixels to the video frames [34].

E. CAMOUFLAGE

The correspondence between foreground pixels and background pixels create camouflaged regions. The background pixels are occluded by foreground pixels and make it difficult to distinguish background from the foreground [35].

F. INTERMITTENT OBJECT MOTION

The foregrounds such as abandoned objects or cars in a parking area that become motionless for a short period of time are incorporated into the background and generate ghosting artifacts with detected foreground [36].

G. MOVING CAMERA

Videos captured with aerial vehicles, pan-tilt-zoom (PTZ) cameras, mobile devices, and hand-held cameras induced background motion to the video scene and complicate the process of detection as compared to static cameras [37], [38].

H. OCCLUSION

The partial or full occlusion complicates the computation of background model [39]. There are many instances of occlusion in real-life such as a moving car occluded by sign boards, a moving person passes behind tree or partially occluded by pole and some regions of moving object may not be visible due to any fixed infrastructure.

I. CHALLENGING WEATHER

Videos recorded in challenging weather conditions such as fog, rain, snow, and air turbulence generate false detections [40].

J. BOOTSTRAPPING

Foreground movement during background initialization makes it difficult to compute a representative background image and generates defective background model [41].

Background subtraction has been an open area of research for decades because of partially solved and unsolved challenges which need to be investigated for many computer vision applications. A detailed study of applications of background subtraction can be found in [173]. Here, we briefly discuss the real-time applications of background subtraction:

A. INTELLIGENT VISUAL SURVEILLANCE OF HUMAN ACTIVITIES

The detection and tracking of foreground objects of interest such as people, vehicles, and abandoned objects are important in order to assure social security [95], [193].

B. INTELLIGENT VISUAL SURVEILLANCE OF ANIMALS

Automatic detection of animals either in the natural habitat or in an artificial setup helps researchers to study and analyze their behavior [194]. It is also required in zoos or other protected areas to keep vision on any unusual activities.

C. HUMAN-MACHINE INTERACTION:

Several video games and multimedia applications require human-machine interaction through a video collected by fixed cameras [180].

D. REAL-TIME HAND GESTURE RECOGNITION

Hand gesture recognition requires detection module to detect moving hand area in a video sequence followed by tracking and recognition modules for applications like robotics, sign language interpretation, and human-computer interface [195].

E. OPTICAL MOTION CAPTURE

The goal is to acquire the exact capture of a human with cameras. For instance, Background subtraction is used to extract movements in the optical motion capture systems.

F. BACKGROUND SUBSTITUTION

Background subtraction is a fundamental step in this application where the object of interest is first extracted from the video and then combined with a new background [196].

G. CONTENT-BASED VIDEO CODING

Background subtraction is used to segment video into video objects in content-based video coding.

The contribution of this survey is threefold. First, it provides insights into the main characteristics of video datasets for background subtraction. Second, different video datasets are compared on the basis of considerable parameters in this paper. Third, detailed literature has been studied for all the standard video datasets relating to background subtraction to label them for the specific area of application. Thus it allows a direct comparison of all the datasets. To the best of our knowledge, most comprehensive compilation of all the background subtraction datasets has been done in this paper to guide researchers in the selection of the appropriate video dataset for evaluating their algorithms, taking different challenges into consideration.

The remainder of the paper is organized as follows: In Section II, significant details of video datasets for background subtraction dedicated to studying the intelligent visual surveillance of human activities are provided. Section III details the RGB-D videos datasets for background subtraction. The video datasets specifically designed for background initialization are mentioned in Section IV. We compared the video datasets from different points of view and mentioned the highlights and gaps in Section V. Section VI details the application specific video datasets. The performance measures in the field discussed in Section VII. Section VIII outlines the background subtraction libraries. The current trends with future research perspective in background subtraction are highlighted in Section IX. The paper finally concludes in Section X.

II. VIDEO DATASETS FOR BACKGROUND SUBTRACTION

The multitudinous datasets have been released for the evaluation of background subtraction methods. Intelligent visual surveillance is the prime application of background subtraction [173]. So, we mentioned video datasets for background subtraction dedicated to studying the intelligent visual surveillance of human activities in this section. They are presented in chronological order of their appearance. The chronological appearance of the different video datasets runs parallel to the challenges that the scientific community has been considering in the domain of detecting moving objects in video frames. Fig. 1 represents the citation frequency of background subtraction datasets. It is evident from the citation counts that CDnet 2012 dataset followed by SABS dataset outclassed other datasets. The latest background subtraction datasets such as Remote Scene IR and CAMO-UOW are less cited but they provide significant videos to test critical challenges.

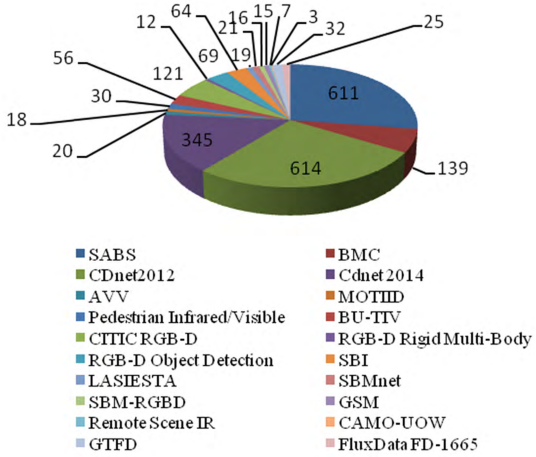

FIGURE 1. Citation frequency of background subtraction datasets (2011-2018).

The details of each dataset are presented in a structured manner, concentrating on a set of significant features, specifically, video sequences, challenges, objective, ground truth, environment, example video frames, number of video frames, and reference papers. The video datasets are categorized into two classes: Small-scale video datasets and Large-scale video dataset.

TABLE 2. Details of video sequences of wallflower dataset.

Video Sequence	Total Frames	Video Scenes	Challenges
Moved objects	1745	Indoor	Moved Background
Time of day	5890	Indoor	Gradual Illumination Changes
Light switch	2715	Indoor	Sudden Illumination Changes
Waving trees	287	Outdoor	Dynamic Background
Camouflage	353	Indoor	Camouflage
Bootstrapping	3055	Indoor	Bootstrapping
Foreground aperture	2113	Indoor	Foreground Aperture

A. SMALL-SCALE VIDEO DATASETS

1) WALLFLOWER DATASET

The well-known, Wallflower dataset [45], was recorded with a 3-CCD camera for studying algorithms of background maintenance in 1999. It is one of the first datasets that focuses on the detection of moving objects. This dataset consists of 7 video sequences to investigate specific challenging scenarios such as illumination variations, dynamic background, camouflage, etc. The complete details of each video sequence are mentioned in Table 2. It provides 16,158 total video frames along with one manually segmented ground truth with pixel-level labeling from each sequence for evaluation. The example video frames including initial video frame, random video frame and their corresponding ground truth of four video sequences from Wallflower dataset are shown in Fig. 2. Some examples of recent works using this dataset for evaluation of background subtraction are [46] and [47].

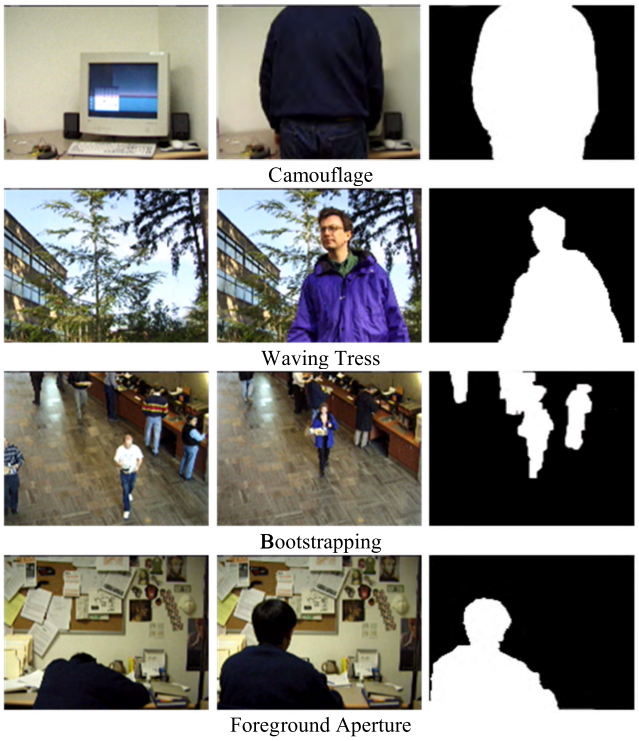

FIGURE 2. Example video frames from Wallflower dataset [45]: Initial video frame (first), random video frame (second), and ground truth (third).

2) PETS DATASET

The Performance Evaluation of Tracking and Surveillance (PETS) workshop [48] was organized since 2000 with the objective of evaluating detection and tracking algorithms in context to surveillance. The PETS datasets provide numerous videos and their associated ground-truths as bounding boxes for qualitative comparisons of different algorithms. The latest series of PETS datasets, PETS 2016 [49], and PETS 2017 [50] focused on the evaluation of event detection and tracking algorithms for onboard surveillance systems with the goal of extending security to crucial assets from terrorists' attacks. Some example works in which different versions of PETS datasets are employed for moving object detection in [51]–[53]. One recent published work [156] used PETS dataset for multi-human tracking.

3) ATON DATASET

ATON dataset [32] consists of 5 video sequences having total 5274 video frames and was released as a part of the ATON project. The primary objective of this project was to investigate the consequences of shadows that interfere with moving object detection and to develop methods for shadow detection and analysis. For that, videos were recorded in both the indoor and outdoor environment. Multiple sensors such as omniview (ODVS) cameras, rectilinear cameras, pan-tilt cameras were combined to fix many of the challenges in moving object detection. Therefore the datasets include video scenes with hard and soft shadows and provide manually segmented ground truths for each video sequence. The ATON project has published many works [54]–[56] with a focus on shadow detection using this dataset.

4) IBM DATASET

IBM dataset was recorded by the IBM Smart Surveillance Research team in order to provide effective solutions to challenging research areas: moving object detection, tracking, face tracking, color classification, and object classification. This dataset consists of 15 video sequences, recorded in both indoor and outdoor environment and some of them are taken from PETS 2001 to deal with complex video scenes. Bounding box associated with each moving object for few video frames is provided as ground truth. This dataset has been used for moving object detection in [57] and [58].

5) CAVIAR DATASET

This dataset was created as a part of the Context Aware Vision using Image-based Active Recognition (CAVIAR) project (2002 – 2005) to address challenges in automated surveillance and automatic behavior analysis of customers. Video dataset consists of 54 video sequences representing

TABLE 3. Details of video sequences of I2R Dataset.

Video Sequence	Total Frames	Video Scenes	Challenges
Curtains	23893	Indoor	Dynamic Background, Camouflage
Campus	1439	Outdoor	Dynamic Background, Shadows, Gradual Illumination Changes
Lobby	2545	Indoor	Shadows, Sudden Illumination Changes
Shopping Mall	1286	Indoor	Shadows, Bootstrapping
Airport	3584	Indoor	Shadows, Bootstrapping
Restaurant	3055	Indoor	Shadows, Bootstrapping
Water Surfaces	633	Outdoor	Dynamic Background
Fountain	1523	Outdoor	Dynamic Background
Subway Station	2634	Outdoor	Dynamic Background, Sudden Illumination Changes, Video Noise
Sidewalks	-	Outdoor	Dynamic Background, Challenging Weather

9 different human activities: walking, fighting, resting, browsing, slumping, group meeting, leaving baggage, shop entering and shop exiting. The video clips were recorded at two different locations: indoor office lobby and shopping center. A bounding box corresponding to each moving object, an activity label and a scenario label for each person are provided as ground truth. A large number of research papers on human activity recognition [59], target detection [60], motion segmentation [61], and tracking [62] were published under this project. For instance, three published works that use this dataset for human detection and tracking are [63]–[65].

6) I2R DATASET

I2R dataset [66] was recorded by Lin and Huang to deal with the challenges of background modeling for detection of moving objects in complex environments. They evaluated their statistical based background modeling algorithm and Bayesian framework for foreground detection on I2R dataset in 2004. It consists of 10 video sequences captured in an indoor as well as an outdoor environment to include video scenes with bootstrapping problems, shadows, camouflage, sudden or gradual illumination variations, video noise, challenging weather, and dynamic backgrounds. The complete details of each video sequence are presented in Table 3. Pixel-wise manually segmented foreground masks are provided for 20 video frames of each video sequence as ground truth. Some of the application examples of this dataset are [67]–[69]. This dataset is also quoted as STAR dataset in some of the published works [70]–[72].

7) OTCBVS BENCHMARK DATASETS

The OTCBVS benchmark is a collection of datasets used for different applications such as person detection, weapon detection, facial analysis, maritime imagery, etc. The main objective of this benchmark dataset is to assess the quality of computer vision algorithms, contributing to the object detection and classification research field. It provides infrared images and videos to instigate research in thermal imagery. Out of total 12 datasets, 7 datasets were recorded to investigate moving object detection in thermal videos. Fig. 3 shows example video frames from OTCBVS benchmark datasets.

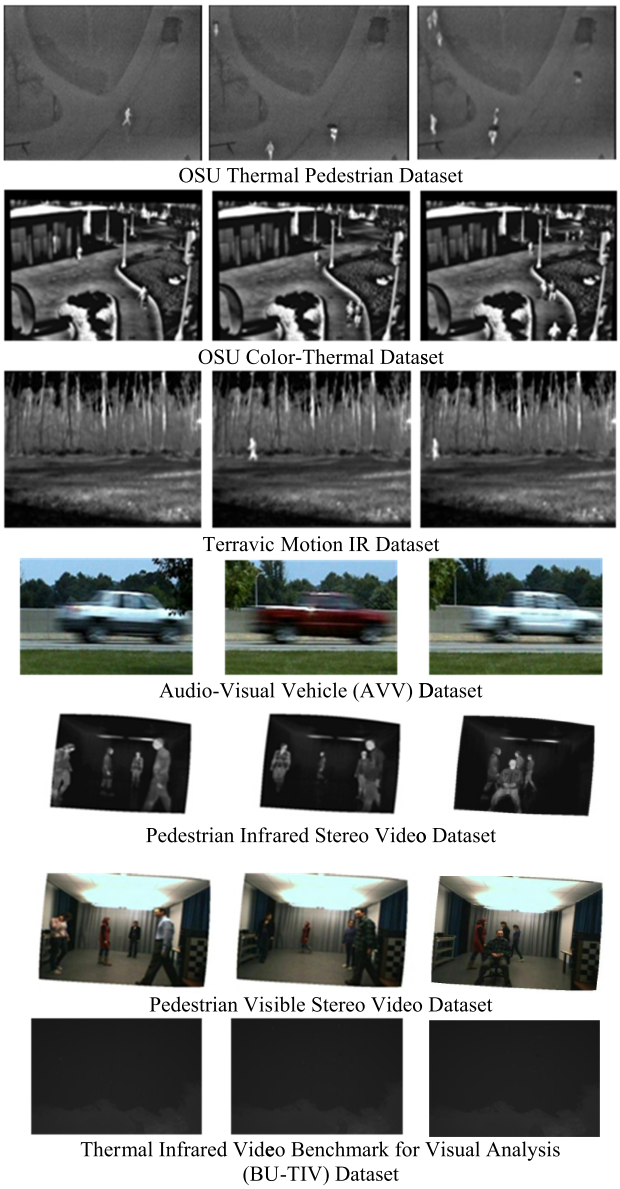

FIGURE 3. Example video frames from OTCBVS Benchmark dataset.

OSU Thermal Pedestrian Database [73], recorded in 2004, was one of the first datasets under OTCBVS benchmark and was recorded for pedestrian detection in thermal videos. It includes 10 video sequences with 284 total video frames and list of bounding boxes as ground truths. Some of the videos were captured during rainy days to cover challenging weather condition. This dataset was used to test person detection in thermal imagery by using background subtraction with AdaBoost classifier. Terravic Motion IR Database [75] is a varied collection of 18 thermal video sequences,

covering both indoor and outdoor video challenges. The video sequences were recorded in a different context for studying detection and tracking algorithms with thermal imagery. Another dataset, OSU Color-Thermal Database [74] consists of 6 video sequences with 17089 total video frames, recorded with both thermal sensor and color sensor at busy pathway intersections in a university campus. The dataset was created with the intention of studying object detection algorithms with the fusion of color and thermal imagery. It was used to evaluate the background subtraction technique for object detection in thermal and visible imagery.

Audio-Visual vehicle (AVV) Dataset [121] was recorded in 2012 to study moving vehicle detection and classification algorithms under various challenging scenarios. They collected 961 sets of vehicles samples and each set of the sample consists of an original image shot, a remodeled visual image, and an audio clip. CSIR-CSIO Moving Object Thermal Infrared Imagery Dataset (MOTIID) [122] was created in 2013 and contains 18 video files to investigate moving object detection in thermal infrared imagery. This dataset was also used in [123] to evaluate statistical based background subtraction technique in infrared videos. Pedestrian Infrared/Visible Stereo Video Dataset [124] was designed to investigate methodologies for human silhouettes registration in infrared/visible videos. This dataset contains 4 infrared and visible video pairs with video frames ranging between 100 and 4400 and provides foreground masks and ground truth point pairs. There is a maximum of five pedestrians circumambulating and occluding each other in all videos. The video sequences in Thermal Infrared Video Benchmark for Visual Analysis (BU-TIV) dataset [125] were designed in 2014 to evaluate detection and tracking algorithm in infrared videos. It contains 11 videos with more than 60,000 video frames of different sizes having single or multiple moving objects and annotated data.

8) CARNEGIE MELLON DATASET

The Carnegie Mellon dataset [76] was created in 2005 to evaluate object detection algorithm based on background subtraction. It has one video with 500 video frames and provides pixel-wise foreground masks for all the video frames as ground truth. One example of recently published work that applied this dataset to background subtraction is [33].

9) ETISEO DATASET

ETISEO dataset was created as a part of video scene understanding evaluation project, ETISEO (2005-2006) to advance the development of video surveillance algorithms [77]. It contains more than 80 realistic videos covering major challenges such as occlusion, shadows, dynamic background, challenging weather conditions, and lighting variations. There are indoor as well as outdoor video scenes recorded with a static camera. As for ground truth, bounding boxes associated with moving objects are provided. Object detection and tracking approaches are evaluated on this dataset and published in [78] and [79].

10) VSSN DATASET

The video sequences in VSSN dataset was recorded in 2006 to advance the research in video surveillance by focusing on different application areas such as traffic monitoring, object detection and recognition, multi-camera tracking, augmented video analysis, and home surveillance. This dataset contains 9 semi-synthetic videos with real backgrounds and synthetic moving objects. The details of video sequences can be consulted in [80]. As ground truths, pixel-wise labeled foreground masks are provided for each video frame. The challenges like shadows, gradual and sudden illumination changes, bootstrapping, and dynamic background are covered under VSSN dataset. Some examples of works using this dataset applied to background subtraction are [81]–[83].

11) BEHAVE DATASET

BEHAVE video dataset [84] was released as a part of the BEHAVE project (2004-2007) with two main objectives: (1) Detection and classification of human interactions such as people greeting, meeting or fighting to check on criminal-oriented behavior, (2) Crowded scene analysis in order to distinguish between normal and abnormal behavior. This dataset contains 4 videos recorded with a same video camera. There are around 90,000 video frames having different types of human activities such as walking in a group, meeting, fighting with each other, etc and bounding boxes associated with each moving object. Two application examples of this dataset are [85] and [86].

12) LIMU DATASET

The Laboratory for Image and Media Understanding (LIMU) provides video dataset for moving object detection. It contains 8 videos, out of which 5 are self-captured and other 3 are borrowed from PETS2001 [90] dataset. This dataset has 2 indoor videos and 6 outdoor videos. Pixel-based labeling of the foreground is provided for 1 video frame out of 15 as ground truth for self-captured 5 video clips.

13) UCSD DATASET

The Statistical Visual Computing Laboratory (SVCL) conducts research for the development of intelligent systems and provides datasets for evaluating different algorithms such object detection, recognition, and classification to accomplish the goal of highly sophisticated intelligent systems. The UCSD Background Subtraction dataset contains 18 video clips recorded in an outdoor environment. The video scenes comprise of complicated moving backgrounds and camera motion in order to evaluate background subtraction technique for highly dynamic scenes. It provides video frames in JPEG format and ground truth masks in MATLAB array form, where 1 represents foreground mask and 0 represents background. This dataset has been used in [91] and [92] for evaluating background subtraction in highly dynamic scenes.

14) cVSG DATASET

The Video Processing and Understanding Lab (VPU) developed a Chroma Video Segmentation Ground Truth (cVSG) dataset [87] in 2008 for the research community to evaluate and compare video segmentation algorithms. This dataset contains 14 semi-synthetic video sequences with rigid and non-rigid foregrounds varying in size (small, medium and large), motion (slow and fast), and textural semblances. The specific details of each video sequence are mentioned in Table 4. Foreground objects were recorded in chroma studio and mounted over different backgrounds. Backgrounds were recorded in an indoor as well as an outdoor environment with both static and moving camera. As for ground truth information, pixel-wise segmented foreground masks are provided for the objective evaluation of motion-based video segmentation algorithms. In [88] and [89], cVSG dataset has been used by researchers for foreground detection in video sequences.

15) I-LIDS DATASET

The Imagery Library for Intelligent Detection Systems (i-LIDS), UK government benchmark for video analytics systems, provides datasets for real-time automated video analysis with the objective to foster the research in 6 areas: detection of abandoned baggage, doorways surveillance, detection in infrared and thermal imagery, parked vehicle detection, sterile zone monitoring, and multi-camera tracking. The i-LIDS dataset [93] consists of several video sequences covering challenging scenarios like illumination variations, all weather conditions, dynamic background, shadows, and occlusion. The 24 hours video footage is segmented into short video clips (30 to 60 minutes) and different camera views are provided in each dataset.

Ground-truths are not provided for full video sequences due to its large size. Along with manually labeled ground-truths, information regarding temporal events and spatial data of moving objects are also provided in XML files. Some examples of works using this dataset are intelligent traffic monitoring [94], [95] and person re-identification [96], [97].

16) SZTAKI SURVEILLANCE DATASET

This surveillance benchmark consists of 5 videos: 2 videos from ATON dataset and other 3 from their personal collection with pixel-wise manually segmented foreground ground truths. A password from the author is required to download this dataset. This dataset was used in the validation of foreground and shadow detection in [98] and [99]. In [100] an outdoor video sequence representing strong shadows of moving objects has been selected from SZTAKI surveillance dataset for the evaluation of moving shadow detection algorithm. This dataset is also used in [101] to evaluate both indoor and outdoor video scenes with illumination variations and variability in foreground movement.

17) SABS DATASET

The Stuttgart Artificial Background Subtraction (SABS) dataset [105] was created in 2011 with two main targets: to test moving object detection methods and to evaluate tracking algorithms. The dataset contains synthetic videos for different challenging scenarios of background subtraction in the context of a prototypical surveillance system. The videos were categorized into nine challenges: basic surveillance scenes, dynamic background, gradual illumination changes, sudden illumination changes, shadow, bootstrapping, camouflage, video compression, and video noise. Video frames of each video are divided into training frames and test frames. Each video has initial 801 video frames as training frames, excluding bootstrap category. For test data, 600 video frames are used except for bootstrap category and gradual illumination changes category (1400 video frames are used as test frames in both of the categories). Shadow masks are also provided along with labeled ground-truth foreground masks of training frames. An example video frame from SABS dataset with its ground-truth mask and shadow mask is shown in Fig. 4. This dataset has been recently used by researchers in [42], [70], and [107] to test their background subtraction algorithm for moving object detection.

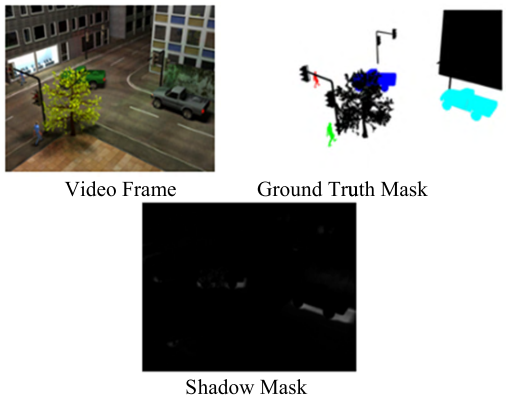

FIGURE 4. Example video frame from SABS dataset [105].

18) FLUXDATA FD-1665 DATASET

The FluxData FD-1665 dataset [201] is a collection of 5 multispectral videos recorded with a FD-1665-MS camera. It has 1 indoor and 4 outdoor video sequences containing between 250 and 2300 video frames. This dataset was created in order to investigate the use of multispectral videos of more than three bands for background subtraction. The challenges such as shadows, camouflage, gradual illumination changes, and intermittent object motion are covered under this dataset. As ground truths, pixel-wise labeled foreground masks are provided for more than 7400 video frames.

19) REMOTE SCENE IR DATASET

This dataset [155] contains realistic videos representing several background subtraction challenges: camouflage, video noise, camera jitter, dynamic background, foreground size,

TABLE 4. Details of video sequences of cVSG Dataset.

Video Sequence	Total Frames	Video Scenes	Foreground Objects	Camera	Challenges
Video 1	3651	Indoor	Non-rigid	Static Camera	Occlusion
Video 2	3651	Indoor	Non-rigid	Moving Camera	Occlusion, Complex Motion
Video 3	3651	Outdoor	Non-rigid	Static Camera	Occlusion, Shadows
Video 4	3651	Outdoor	Non-rigid	Moving Camera	Occlusion, Shadows, Complex Motion
Video 5	367	Outdoor	Non-rigid & Rigid	Moving Camera	Occlusion
Video 6	258	Outdoor	Non-rigid & Rigid	Moving Camera	Occlusion, Intermittent Object Motion
Video 7	1251	Outdoor	Non-rigid& Rigid	Static Camera	Shadows
Video 8	752	Outdoor	Non-rigid	Static Camera	Shadows, Sleeping Foreground
Video 9	671	Outdoor	Non-rigid & Rigid	Static Camera	Intermittent Object Motion
Video 10	619	Outdoor	Non-rigid &Rigid	Static Camera	Shadows, Complex Motion
Video 11	793	Outdoor	Non-rigid	Static Camera	Sleeping Foreground
Video 12	1378	Outdoor	Non-rigid & Rigid	Static Camera	Intermittent Object Motion
Video 13	307	Outdoor	Non-rigid & Rigid	Static Camera	Occlusion, Shadows
Video 14	732	Outdoor	Rigid	Static Camera `	Occlusion

TABLE 5. Challenges in video sequences of remote scene IR Dataset.

Challenges	V1	V2	V3	V4	V5	V6	V7	V8
Dynamic Background	✓	✓	✓	×	×	×	×	×
Ghosts	✓	×	✓	✓	×	×	×	×
Camera Jitter	×	×	×	✓	×	✓	×	×
Video Noise	×	×	×	✓	✓	✓	×	×
Camouflage	×	✓	✓	×	×	×	×	×
Foreground Size	×	×	×	×	✓	✓	×	×
Low Speed	×	×	×	×	×	×	✓	✓
High Speed	×	×	×	×	×	×	✓	✓

speed of the foreground, and ghosts. All video sequences were recorded in 2017 with a medium-wave infrared sensor, designed by the authors. Some examples video frames with their ground truth from Remote Scene IR dataset are shown in Fig. 5. There are a total of 1263 frames in 12 videos and are available in. BMP format. Pixel-wise foreground ground truths are provided for each frame by manually segmented them. The overall objective of this dataset is to evaluate background subtraction algorithms on real remote scene infrared video sequences.

Table 5 shows challenges in different videos of Remote Scene IR dataset. The video scenes of each video have two or more challenges to test. There are three same series of video sequences with different frame sample rates representing the low speed of foreground under V7 video sequence and three same series of video sequences with different frame sample rates representing the high speed of foreground under V8 video sequence. Application examples of this dataset for background subtraction are [159]–[161].

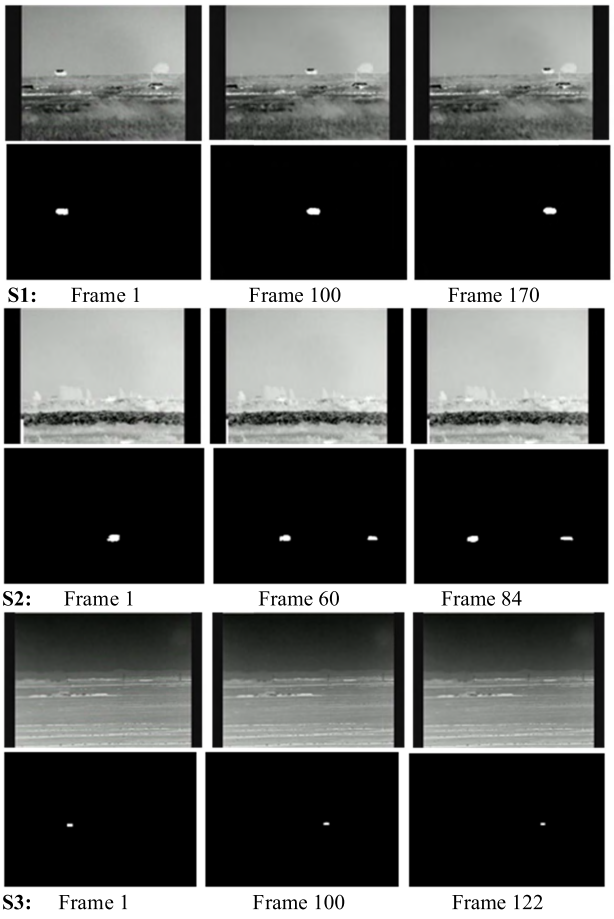

FIGURE 5. Example video frames and their corresponding ground truth from Remote Scene IR dataset [155]: S1 (Dynamic background + ghosts), S2 (Camouflage + Dynamic background), and S3 (Video noise + Small and Dim Foreground).

20) CAMO-UOW DATASET

CAMO-UOW dataset was recorded in 2017 to address the problem of camouflaged moving foreground detection in real scenes [174], [175]. It comprises of 10 videos having 3,517 total video frames. The complete details of each video sequence are mentioned in Table 6. Each video consists of one or two persons wearing similar color clothes as that of

TABLE 6. Details of video sequences of CAMO-UOW Dataset.

Video Sequence	Total Frames	Format
Video 1	371	Grayscale
Video 2	176	Grayscale
Video 3	371	Grayscale
Video 4	371	Grayscale
Video 5	371	Grayscale
Video 6	373	Grayscale
Video 7	272	RGB
Video 8	466	RGB
Video 9	288	RGB
Video 10	458	RGB

the background. The manually segmented ground truth with pixel-based labeling is provided for all video frames. The video scenes were recorded in both the indoor and outdoor environment.

B. LARGE-SCALE VIDEO DATASETS

Besides the above-mentioned datasets, there are other widely used video datasets that primarily focused on the evaluation of background subtraction strategies. The datasets discussed below cover a wide range of challenging scenarios of background subtraction.

1) BMC DATASET

BMC dataset [108] was created in 2012 for the Background Models Challenge workshop to evaluate and rank background subtraction algorithms. It comprises of 20 synthetic videos and 9 real videos, recorded in an outdoor environment with a focus on different weather situations such as rain, sun, fog or wind. The video sequences are categorized into two sets: learning and evaluation. The first set contains 10 synthetic videos and the second set contains other 10 synthetic videos and 9 real videos. The real videos with a long sequence of frames are captured with the static camera to investigate the reliability of algorithms in challenging scenarios such as sudden illumination changes, casted shadows, challenging weather, big foreground, and dynamic background. The hand-segmented ground truth with pixel-based labeling is provided for 10 synthetic videos under learning set and real videos under evaluation set. Fig. 6 shows example video frames and their associated ground truths from BMC dataset. Some examples of works using this dataset to evaluate their background subtraction algorithms are [19], [109], [110].

2) CDNET DATASET

There are two versions of CDnet dataset: CDnet 2012 and CDnet 2014, presented at IEEE Change Detection Workshops for benchmarking change detection algorithms. The CDnet 2012 dataset [111] was recorded in 2012 with distinct cameras including PTZ camera, low-resolution IP cameras, mil-resolution camcorders, and thermal cameras. It consists of 31 videos having total 90,000 video frames and is grouped into six categories to cover a wide range of challenges that

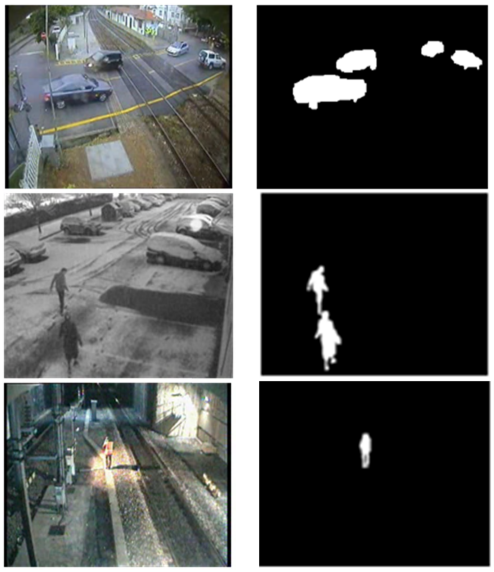

FIGURE 6. Video frames and their associated ground truth masks from BMC dataset [108].

exist in most video analytics applications. Out of 31 videos, one video was borrowed from the PETS 2006 dataset and other 30 videos were self-captured by the authors. A distinguishing mark of this dataset is that video frames are annotated for shadow regions and background along with manually segmented ground truth foreground masks. Two examples of published works using CDnet 2012 dataset for detecting moving objects using background subtraction are [112] and [113].

In 2014, researchers released the CDnet 2014 dataset [114] with 22 additional videos having more than 70,000 video frames. This dataset incorporates complex video scenes and adds 5 new video categories each representing a specific situation. The video categories under both benchmark datasets are mentioned in Table 7. Fig. 7 shows example video frames of CDnet datasets. There are a lot of works, for instance [115]–[120], using CDnet 2014 dataset in the validation of background subtraction algorithms.

3) LASIESTA DATASET

The Labeled and Annotated Sequences for Integral Evaluation of Segmentation Algorithms (LASIESTA) dataset [140], created in 2016, is a collection of 48 videos recorded with static and moving cameras in indoor and outdoor scenarios. LASIESTA aims to evaluate the quality of algorithms focused on the detection and tracking of moving objects. This dataset consists of 18,425 total video frames and videos are categorized into indoor sequences, outdoor sequences, indoor sequences with simulated motion and outdoor sequences with simulated motion. The details of four video categories are mentioned in Table 8. The ground truths are provided for each video frame at both pixel and object level. A video sequence contains a maximum of three moving objects and is labeled as: red pixels for a first moving object, green pixels for

TABLE 7. Video categories in CDnet Dataset.

CDnet 2012			
Video Categories	Videos	Indoor	Outdoor
Baseline	4	2	2
Dynamic Background	6	×	6
Camera Jitter	4	1	3
Shadows	6	2	4
Intermittent Object Motion	6	1	5
Thermal	5	2	3
CDnet 2014			
Video Categories	Videos	Indoor	Outdoor
Challenging Weather	4	×	4
Low Frame-Rate	4	×	4
Night	6	×	6
PTZ	4	×	4
Air Turbulence	4	×	4

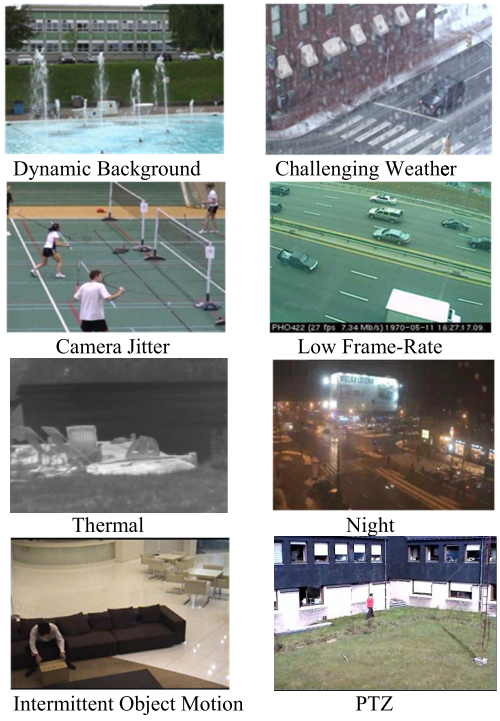

FIGURE 7. Example video frames from CDnet dataset: CDnet 2012 (first column) [111] and CDnet 2014 (second column) [114].

a second moving object, and yellow pixels for a third moving object. Along with original videos and ground truths, it also provides information of temporarily static moving objects which is useful for abandoned object detection. The video sequences are of short duration ranging between 7 seconds to 56 seconds. The challenges in the dataset include shadows, dynamic background, illumination changes, occlusion, camouflage, moving camera, bootstrapping, stationary foreground objects, and challenging weather. Few video frames of both indoor and outdoor sequences with their associated ground truths are shown in Fig. 8. Some examples of works

TABLE 8. Video categories in lasiesta dataset.

Video Categories	Videos	Video Frames
Indoor sequences	14	4,700
Outdoor sequences	10	5,025
Simulated Motion with Indoor sequences	12	3,600
Simulated Motion with Outdoor sequences	12	5,100

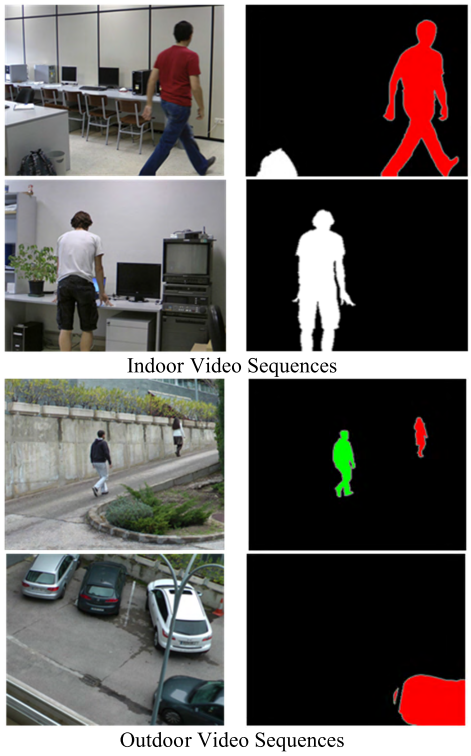

FIGURE 8. Example video frames from LASIESTA dataset [140]: Original video frames (first column) and ground truth masks (second column).

using this dataset for moving object detection using background subtraction approach are [70], [141]–[144].

4) GTFD DATASET

The Grayscale-Thermal Foreground Detection (GTFD) dataset [176], created in 2017, is a comprehensive grayscale-thermal video benchmark for moving object detection. It is a collection of 25 videos having 1067 total video frames that include rigid and non-rigid objects. The videos were recorded in both the indoor and outdoor environment to cover different challenging scenarios. The dataset includes 7 major challenges: intermittent motion, bad weather, low illumination, intense shadow, dynamic scene, background clutter, and thermal crossover. The complete details of each video sequence of GTFD dataset are mentioned in Table 9. GTFD dataset provides a pair of grayscale and thermal video frames, annotated ground truths, implemented baselines and evaluation metrics to study and benchmark foreground detection algorithms for grayscale-thermal videos. The manually annotated ground

TABLE 9. Details of video sequences of GTFD Dataset.

Video Sequence	Total Frames	Video Sequence	Total Frames
Pedestrian 1	101	Pedestrian 5	30
Pedestrian 2	112	Pond walker	30
Pedestrian 3	131	Rain	30
Truck	50	Pedestrian 6	30
Night 1	40	Pedestrian 7	30
White car	30	Moving clouds	24
Pedestrian 4	63	Pedestrian 8	25
Car 1	29	Bad thermal	30
Car 2	30	Night 3	25
Swam	45	Car 4	25
Night 2	40	Car 5	30
Car 3	25	Car 6	30
Bus	31		

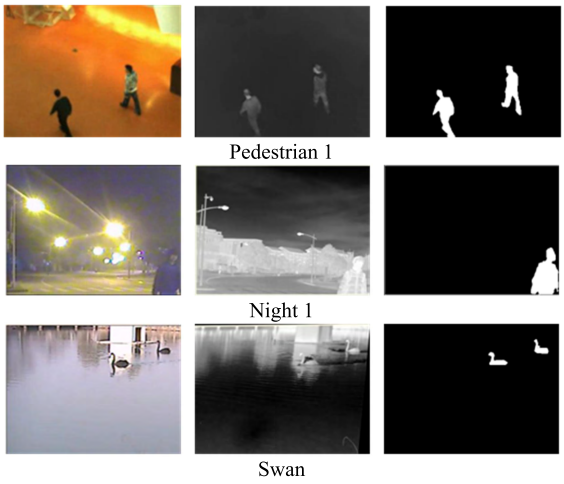

FIGURE 9. Example video frames from GTFD dataset [176]: Grayscale video frames (first column), Infrared video frames (second column), and ground truth masks (third column).

truths with pixel-based labeling are provided for all video frames. Li *et al.* [176] designed this dataset to investigate the fusion of thermal and grayscale data for effective foreground detection. Example video frames of GTFD dataset with their associated ground truths are shown in Fig. 9. Some examples of works using this dataset for moving object detection using background subtraction approach are [177]–[179].

III. RGB-D VIDEO DATASETS FOR BACKGROUND SUBTRACTION

The depth data produced by RGBD sensors are of great significance for resolving major challenges of many computer vision applications and have opened new opportunities for detecting moving objects. Nowadays, researchers are trying to exploit depth information generated by RGBD sensors to cope with challenges of background subtraction. The RGB-D video datasets described below can be useful to benchmark those background subtraction algorithms that work with videos recorded with depth sensors.

TABLE 10. Details of video sequences of Citic Rgb-d Dataset.

Set I: Stereo Camera		
Video Sequence	Total Frames	Challenges
Suitcase	288	Illumination Changes, Low Light, Camouflage
Crossing	624	Camouflage, Shadows
LCD Screen	525	Camouflage, Illumination Changes, Flickering Lights
Lab Door	1416	Shadows, Sudden Illumination Changes, Occlusion, Flickering Lights
Set II: Microsoft Kinetic Sensor		
Video Sequence	Total Frames	Challenges
Chair Box	529	Flickering lights
Wall	218	Shadows, Illumination Changes
Shelves	554	Inserted Background, Complicated Depth Estimation
Hallway	618	Shadows, Sudden Illumination Changes, Camouflage

A. CITIC RGB-D DATASET

This dataset consists of two different set of video sequences recorded in 2013 for the evaluation of background subtraction methods and is publically available on the MULTIVISION website [126]. The first set [127] contains total 2,853 video frames in 4 indoor video sequences recorded by using the stereo cameras with the objective of ranking different background subtraction technique based on depth-computation algorithms. Along with video frames, pixel-based manually segmented foreground masks of some video frames and disparity details by three different approaches are provided as ground-truth data. The second one [128] contains total 1,919 video frames in 4 video sequences recorded by using the Microsoft kinetic sensor in an indoor environment and was devoted to evaluate color and depth based background subtraction technique using sensors. In addition to RGB and Depth images, manually segmented foreground masks are provided for some video frames as ground-truth information. The detailed description of each video sequence of both sets is provided in Table 10. Example video frames from set I and set II of CITIC RGB-D dataset are presented in Fig. 10(a) and Fig. 10(b). This dataset is also named as MULTIVISION dataset by authors as in paper [129].

B. RGB-D RIGID MULTI-BODY DATASET

The RGB-D Rigid Multi-Body dataset [130] consists of 3 video sequences with 3300 total video frames to test motion estimation and segmentation algorithms on moving objects of varying sizes such as large objects (two chairs), medium objects (watering can and box), and small objects (tea can and cereal box) in RGB-D video scenes. This dataset was recorded in 2013 by non-static RGB-D camera (Asus Xtion Pro Live Camera) in an indoor environment. Each video sequence contains 1100 video frames and is considered as

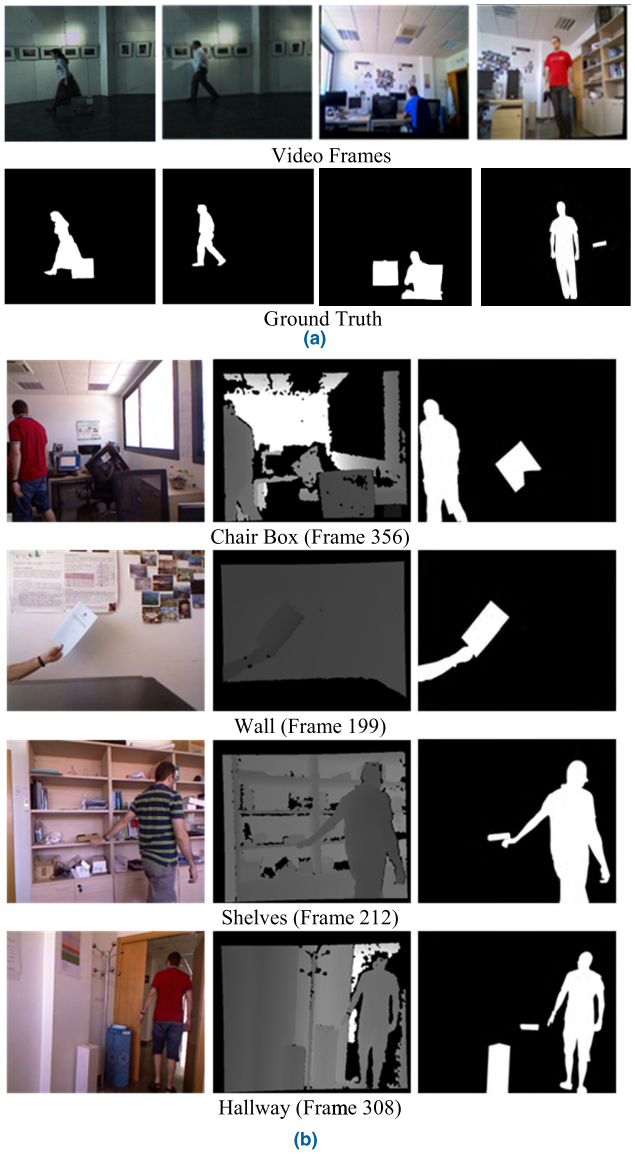

FIGURE 10. (a). Example video frames from set I of CITIC RGB-D dataset [127]: Frame 190 of suitcase (first column), frame 565 of crossing (second column), frame 435 of LCD screen (third column), and frame 1003 of Lab door (fourth column). (b). Example video frames from set II of CITIC RGB-D dataset [128]: Original video frames (first column), depth images (second column), and ground truth (third column).

a case of moved background objects. The objects of interest for segmentation are rigid objects such as chairs, boxes and can. Fig. 11 shows example video frames of RGB-D Rigid Multi-Body dataset.

Non-rigid objects such as arms and legs of person were annotated as don't care labels in ground truth data. Foreground masks of rigid objects were obtained with a motion capture system for ground truth data. The video frames were also manually annotated at every 5 seconds throughout the sequence. One published work that uses this dataset to improve the performance of the background subtraction algorithm in complicated video scenes is [131].

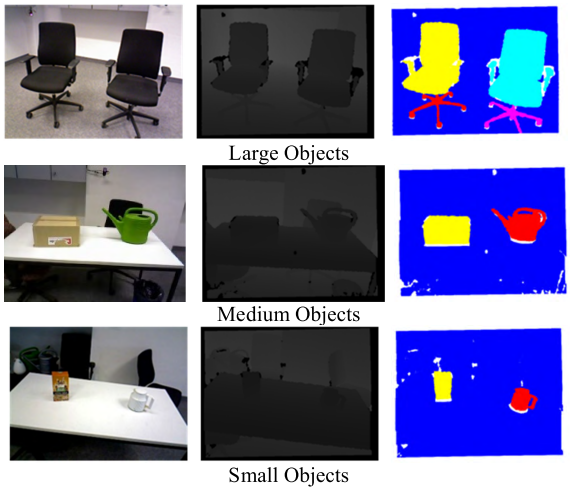

FIGURE 11. Example video frames from RGB-D Rigid Multi-Body dataset [130]: Original video frames (first column), depth images (second column), and ground truth (third column).

C. RGB-D OBJECT DETECTION DATASET

The RGB-D Object Detection dataset [132] was created in 2014 to study different challenging situations in foreground and background segmentation. The authors of this dataset proposed an algorithm by considering both color and depth information in video scene to improve the accuracy of background subtraction. It contains 5 video sequences with 1830 total video frames. All the videos were recorded in an indoor laboratory by using the Microsoft Kinect RGB-D camera at a frame rate of 30 fps. Out of 5 video sequences, the first one was recorded to test the overall performance of the algorithm in complex scenes taking into account different challenges and the others were recorded to focus on a specific challenge. Along with color and depth video frames, pixel-wise hand labeled foreground masks are provided as ground truth data. The details of each video sequence are provided in Table 11. Fig. 12 shows color and depth video frames of RGB-D Object Detection dataset.

TABLE 11. Details of video sequences of Rgb-d Object Dataset.

Video Sequence	Total Frames	Ground Truth	Challenges
Genseq	300	39	Shadows, Color Camouflage, Noisy Depth Data, Move Background Objects
Shseq	250	25	Shadows
ColCamSeq	360	45	Color Camouflage
DCamSeq	670	102	Depth Camouflage
MoveBGSeq	250	37	Moved Background Objects

D. SBM-RGBD DATASET

SBM-RGBD dataset [150] consists of 33 RGBD videos with 15033 total video frames recorded in an indoor environment by a Microsoft Kinect sensor. Some videos were selected from five public datasets and others were

FIGURE 12. Example video frames from RGB-D Object dataset [132]: Color video frames (first column) and depth images (second column).

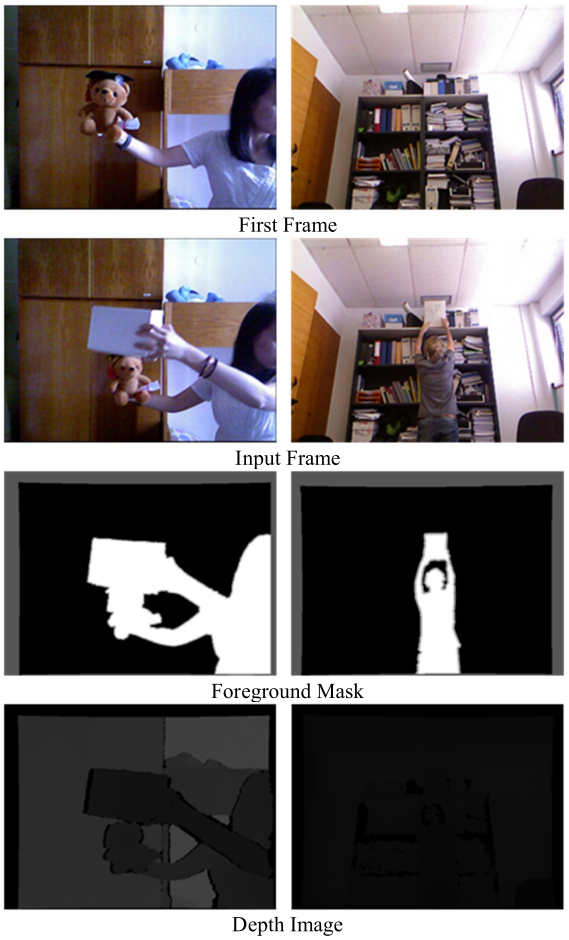

FIGURE 13. Example video frames from SBM-RGBD dataset [150]: Bootstrapping (first column) and Color Camouflage (second column).

self-captured by the organizers in 2017. The videos were categorized into 7 categories, each representative of specific background modeling challenges: *Illumination Changes* (4 videos with 2,579 video frames), Color Camouflage (4 videos with 1,707 video frames), Depth Camouflage (4 videos with 1,953 video frames), Intermittent Motion (6 videos with 1,854 video frames), Out of Sensor Range (5 videos with 4,610 video frames), Shadows (5 videos with 1,301 video frames), and Bootstrapping (5 videos with 1,029 video frames). Fig. 13 shows example color video frames with their first video frame, foreground masks, and depth sequences. Pixel-wise foreground masks are provided for all the videos as ground truth but only a few of them are available publicly for testing purpose. This dataset aims to benchmarking of background modeling algorithms on RGBD videos. One published work [151] uses this dataset to test their background subtraction algorithm for illumination variations and color camouflage in an indoor environment.

E. GSM DATASET

GSM dataset [152] was created in 2017 to provide a comprehensive RGBD dataset for evaluation and comparison of background subtraction algorithms. This dataset covers all the typical challenges of background modeling that obstruct the detection of moving objects in RGBD videos. It consists of 7 indoor video sequences with 3361 total video frames. A small number of video frames from each video sequence are manually labeled pixel-wise for ground truth data. Foreground masks as ground truth data are selectively provided for those video frames where a particular challenge is prominent in the video scene. Details of each video sequence of GSM dataset are presented in Table 12. The intention of this dataset is to investigate background subtraction algorithm for video scenes recorded with different RGBD sensors. In paper [153], researchers surveyed various publicly available RGBD datasets including GSM dataset for background subtraction. Example color and depth video frames from each video sequence of GSM dataset representing different challenging scenario is shown in Fig. 14. An application example paper which uses this dataset for background subtraction in RGBD videos can be consulted in [154].

IV. VIDEO DATASETS FOR BACKGROUND INITIALIZATION

Background subtraction is a three-stage process in which background initialization module forms the initial stage to

TABLE 12. Video categories in GSM Dataset.

Video Categories	Video Frames	Ground Truth Frames	Challenges
Time of Day	1231	23	Gradual Illumination Changes
Color camouflage	428	11	Camouflage
Depth camouflage	465	12	Camouflage
Shadows	330	11	Shadows
Light switch	407	9	Sudden Illumination Changes
Bootstrapping	300	11	Bootstrapping
Walking object	200	10	Moved Background Object

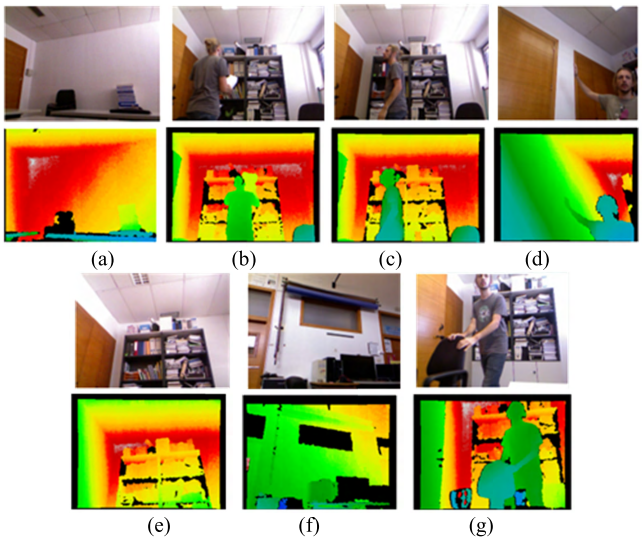

FIGURE 14. One example video frame and an associated depth image from each video categories of GSM dataset [152].

provide a clean background from a video sequence. There are background initialization based applications such as privacy protection, video inpainting, and computational photography that only need a clean background without foreground detection [173]. In this section, we mentioned video datasets specifically designed to benchmark background initialization algorithms.

A. SBI DATASET

The Scene Background Initialization (SBI) dataset [133] contains videos from 8 publically accessible datasets. It was created in 2015 for the evaluation and comparison of different background initialization algorithms. This dataset provides 14 videos along with single ground truth background of each video. One of the video frames, free of foreground is manually segmented for ground truth. Both indoor and outdoor video scenes such as moving person with mild shadows, occlusion of car by waving trees in a parking area, sleeping foreground, etc, were included in the dataset in order to rank algorithms on typical challenges. The researchers compared 15 background initialization algorithms on SBI dataset to highlight the most propitious approaches as well as open research challenges in this area in paper [28]. Fig. 15 shows example frames of SBI dataset and associated ground truth background. This dataset has been used, for instance, in papers [134]–[139] to evaluate background modeling and background initialization methods.

FIGURE 15. Example video frames from SBI dataset [133]: Original video frames (first column) and background ground truths (second column).

B. SBM NET DATASET

The Scene Background Modeling.NET (SBMnet) dataset [29] provides a diverse set of realistic videos with total 73,357 video frames captured from different video cameras in 2016. There are 79 videos, grouped into 8 video categories with the objective of covering a wide range of motion detection challenges. The video categories are named as: *Basic, Intermittent motion, Clutter, Jitter, Illumination changes, Background motion, Very long, and Very short*. The 'Very long' video category contains video with more than 3,500 video frames and 'Very short' video category has video less than 20 video frames. The details of each video category are provided in Table 13.

Some videos were borrowed from publicly available video datasets and others were self-captured. As ground truth, a colored background video frame is provided after removing foreground with the semi-automatic method. See Fig. 16 for series of video frames from SBMnet dataset. Application examples of this dataset for background initialization and background modeling are [145]–[149].

TABLE 13. Video categories in SBMnet Dataset.

Video Categories	Videos	Indoor	Outdoor	Video Frames
Basic	16	3	13	9,539
Intermittent motion	16	8	8	15,805
Clutter	11	3	8	6,499
Jitter	9	2	7	6,497
Illumination changes	6	4	2	5,434
Background motion	6	×	6	3,626
Very long	5	×	5	25,878
Very short	10	×	10	79

FIGURE 16. Series of video frames from SBMnet dataset [29]: (a) Illumination changes, (b) Clutter, and (c) Intermittent Motion.

V. DISCUSSION

In this section, the key attributes of different video datasets for background subtraction and background initialization described in this survey are summarized clearly in four tables (Tables 14-17). The best results obtained to date for background subtraction on significant benchmark datasets are also discussed.

Table 14 presents the following information: the year of creation of dataset, the reference paper that provides detailed information of dataset, the total number of videos, and the web page from where the dataset can be downloaded. Table 15 summarizes the following characteristics: the type of dataset, the kind of sensor used, the type of video scenes, and the type of camera used to record the videos. The video datasets contain synthetic, semi-synthetic or realistic video frames and this information aids in the selection of video dataset for evaluating different algorithms. The other valuable information is the type of sensor used because it will strongly influence the option of dataset to be used. In recent years, many video datasets dedicated to background subtraction have used thermal sensors to record video scenes at the different time of day as these sensors have proved to be potential in the area of automated video surveillance. The problems associated with background subtraction method for detecting moving objects from an outdoor video scene are more challenging than indoor ones. The information about the type of video scenes acts as a distinguishing mark for video datasets. Videos captured by different cameras such as Pan-tilt-zoom cameras, hand-held cameras, and moving cameras add new challenges to background subtraction method and necessitate different algorithms in comparison to static cameras. The information about a type of camera will guide researchers for testing and ranking their proposed algorithm on appropriate datasets.

Table 16 categorizes video datasets on the basis of ground truth data, evaluation domain, and area of application. One of the essential information is ground truth as it provides relevant details of the video scenes and depicts the versatility of datasets. The ground truth assists in the task of evaluation and analysis of the algorithms. The second column shows two general entries for ground truth data: pixel-wise labeling and bounding boxes. For interpretation of background subtraction algorithms, pixel-wise labeling of video frames is appropriate as ground truth information whereas bounding boxes ground truth is specific to tracking algorithms. The video datasets such as SBI [133] and SBMnet [29] provide background video frame as ground truth information because they are specifically designed for interpretation of background initialization and modeling algorithms. The third column is devoted to evaluation domain of datasets as each dataset is explicitly designed to investigate certain challenges. The last column indicates an area of application and provides references in which each video dataset has been used for benchmarking. All the characteristics used to categorize video datasets in this paper are pertinent to the selection of appropriate dataset. Some of the datasets are publically available for download while the others require a license agreement.

The challenges of background subtraction discussed in Section 1 have a significant role in the selection of datasets as specific challenges are taken into account by the researchers. Therefore, video datasets are classified on the basis of different challenges of background subtraction in Table 17. There are several video categories under dataset representing either a particular challenge or group of challenges so the same dataset can be seen in various groups. For example, Wallflower dataset has seven video sequences to evaluate different challenges and can appear in seven categories. The video categories representing different challenges of background subtraction is the cornerstone of video dataset and direct researchers to choose the best dataset for evaluating and comparing their algorithms with the state-of-the-art algorithms. As it is clearly seen from table 17, most of the video datasets have been recorded to cover challenges such as dynamic background, shadows, gradual illumination changes, sudden illumination changes, occlusion, color camouflage, challenging weather and thermal videos. Automated

TABLE 14. Details of video datasets dedicated to background subtraction for detection of moving objects.

Name/Description/Year	Videos	Web
Small-Scale Video Datasets for Background Subtraction		
Wallflower [45] (1999)	7	https://www.microsoft.com/en-us/research/project/test-images-for-wallflower-paper/
PETS [48] (2000-2017)	-	http://www.cvg.reading.ac.uk/slides/pets.html
ATON [32] (2003)	5	http://cvrr.ucsd.edu/aton/shadow/
IBM	15	http://www.research.ibm.com/peoplevision
CAVIAR [157], [158] (2002-2005)	54	http://homepages.inf.ed.ac.uk/rbf/CAVIAR/
I2R [66] (2004)	10	http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html
OSU Thermal Pedestrian Database [73] (2004)	10	http://vcipl-okstate.org/pbvs/bench/
Terravic Motion IR Database (2005)	18	http://vcipl-okstate.org/pbvs/bench/
OSU Color-Thermal Database [74] (2007)	6	http://vcipl-okstate.org/pbvs/bench/
Carnegie Mellon [76] (2005)	1	https://www.cmu.edu/ira/CDS/index.html
ETISEO [77] (2005)	80	https://www-sop.inria.fr/orion/ETISEO
VSSN [80] (2006)	9	http://mmc36.informatik.uni-augsburg.de/VSSN06%20OSAC/
BEHAVE [84] (2004-2007)	4	http://groups.inf.ed.ac.uk/vision/BEHAVEDATA/INTERACTIONS/
cVSG [87] (2008)	14	http://www-vpu.ii.uam.es/CVSG/
LIMU	8	http://limu.ait.kyushu-u.ac.jp/dataset/en/
UCSD [91] (2007)	18	http://svcl.ucsd.edu/projects/background_subtraction/
i-LIDS [93]	-	https://www.gov.uk/guidance/imagery-library-for-intelligent-detection-systems#i-lids-datasets
SZTAKI Surveillance	5	http://web.eee.sztaki.hu/~bcsaba/FgShBenchmark.htm
SABS [105] (2011)	9	https://www.vis.uni-stuttgart.de/
AVV [121] (2012)	-	http://vcipl-okstate.org/pbvs/bench/
MOTIID [122] (2013)	18	http://vcipl-okstate.org/pbvs/bench/
Pedestrian Infrared/Visible Stereo Video Dataset [124] (2013)	4	http://vcipl-okstate.org/pbvs/bench/
BU-TIV [125] (2014)	11	http://vcipl-okstate.org/pbvs/bench/
FluxData FD-1665 [201] (2014)	5	http://ilt.u-bourgogne.fr/benezeth/projects/ICRA2014/
Remote Scene IR [155] (2017)	12	https://github.com/JerryYaoGl/BSEvaluationRemoteSceneIR
CAMO-UOW [175] (2017)	10	https://www.uow.edu.au/~wanqing/#Datasets
Large-Scale Video Datasets for Background Subtraction		
BMC [108] (2012)	29	http://bmc.iut-auvergne.com/?page_id=24
CDnet 2012 [111] (2012)	31	http://www.changedetection.net/
CDnet 2014 [114] (2014)	22	http://www.changedetection.net/
LASIESTA [140] (2016)	48	http://www.gti.ssr.upm.es/data/LASIESTA
GTFD [176] (2017)	25	http://chenglongli.cn/people/lcl/dataset-code.html
RGB-D Video Datasets for Background Subtraction		
CITIC RGB-D [127], [128] (2013)	8	http://atcproyectos.ugr.es/mvision/
RGB-D Rigid Multi-Body [130] (2013)	3	http://www.ais.uni-bonn.de/download/rigidmultibody/
RGB-D Object Detection [132] (2014)	5	https://seis.bristol.ac.uk/~mc13306/
SBM-RGBD [150] (2017)	33	http://rgbd2017.na.icar.cnr.it/SBM-RGBDdataset.html
GSM [152] (2017)	7	http://gsm.uib.es/#dataset
Video Datasets for Background Initialization		
SBI [133] (2015)	14	http://sbmi2015.na.icar.cnr.it/SBIdataset.html
SBMnet [29] (2016)	79	http://scenebackgroundmodeling.net/

video surveillance systems need to deal with distinct challenges in the real-time environment. More the challenges covered in video datasets, greater will be its significance. The number of challenges covered in video datasets during 1999-2010 is presented in Fig. 17. Indeed, the influence of background subtraction datasets depends upon the number of critical challenges covered in video datasets. The number of challenges covered in video datasets during 2011-2017 is presented in Fig. 18.

The most recent video datasets have RGB-D videos as the depth data generated by depth sensors are of great significance for resolving major challenges of many computer vision applications. The depth data are more robust to dynamic background and illumination variations and fit for indoor video scenes in comparison to color data. But there are some challenges like depth camouflage and complicated depth estimation associated with depth data in case of moving object detection. The video scenes with such challenges appear in CITIC RGB-D, RGB-D object detection, SBM-RGBD, and GSM datasets. Conventional background subtraction algorithms need to redesign in order to deal with the fusion of color and depth data. The detection of moving objects in real-time applications encountered with different size of foregrounds: small-sized objects, medium-sized objects, and large sized objects. This type of video scenes appears in BMC, cVSG, RGB-D Rigid Multi-Body and Remote scene IR datasets. The real-time detection of moving objects encounters with several challenges. So, the selection of appropriate video datasets will substantially influence the interpretation of the algorithm.

TABLE 15. Characteristics of video datasets dedicated to background subtraction for detection of moving objects.

Name	Type	Sensors	Video Scene	Camera
Wallflower	Realistic	Color	Indoor & Outdoor	Static camera
PETS	Realistic	Color	Indoor & Outdoor	Multiple cameras
ATON	Realistic	Color	Indoor & Outdoor	Multiple cameras
IBM	-	Color	Indoor & Outdoor	Multiple cameras
CAVIAR	Realistic	Color	Indoor	Static camera
I2R	Realistic	Color	Indoor & Outdoor	Multiple cameras
OSU Thermal Pedestrian Database	Realistic	Thermal	Outdoor	Static camera
Terravic Motion IR Database	Realistic	Thermal	Indoor & Outdoor	-
OSU Color-Thermal Database	Realistic	Color & Thermal	Outdoor	Static camera
Carnegie Mellon	Realistic	Color	Outdoor	Moving camera
ETISEO	Realistic	Color	Indoor & Outdoor	Static camera
VSSN	Semi-synthetic	Color	Indoor & Outdoor	Static camera, Moving camera
BEHAVE	Realistic	Color	Outdoor	Static camera
cVSG	Semi-synthetic	Color	Indoor & Outdoor	Static camera , Moving camera
LIMU	Realistic	Color	Indoor & Outdoor	Multiple cameras
UCSD	Realistic	Color	Outdoor	Static camera
i-LIDS	Realistic	Color & Thermal	Outdoor	Multiple cameras
SZTAKI Surveillance	Realistic	Color	Indoor & Outdoor	Multiple cameras
SABS	Synthetic	Color	Outdoor	-
AVV	Realistic	Color	Outdoor	Pan-tilt-zoom cameras
MOTIID	Realistic	Thermal	Outdoor	Static camera
Pedestrian Infrared/ Visible Stereo Video Dataset	Realistic	Color & Thermal	Indoor	Static camera
BU-TIV	Realistic	Thermal	Indoor & Outdoor	Multiple cameras
FluxData FD-1665	Realistic	Color	Indoor & Outdoor	Static camera
Remote Scene IR	Realistic	Thermal	Outdoor	Static cameras
CAMO-UOW	Realistic	Color	Indoor & Outdoor	Static cameras
BMC	Realistic & Synthetic	Color	Outdoor	Static camera
CDnet 2012	Realistic	Color & Thermal	Indoor & Outdoor	Multiple cameras
CDnet 2014	Realistic	Color & Thermal	Outdoor	Multiple cameras
LASIESTA	Realistic	Color	Indoor & Outdoor	Static camera, Moving camera
GTFD	Realistic	Color & Thermal	Indoor & Outdoor	Static cameras
CITIC RGB-D	Realistic	Color	Indoor	Static camera
RGB-D Rigid Multi-Body	Realistic	Color	Indoor	Non-static camera
RGB-D Object Detection	Realistic	Color	Indoor	Static camera
SBM-RGBD	Realistic	Color	Indoor	Multiple cameras
GSM	Realistic	Color	Indoor	Multiple cameras
SBI	Realistic	Color	Indoor & Outdoor	Multiple cameras
SBMnet	Realistic	Color	Indoor & Outdoor	Multiple cameras

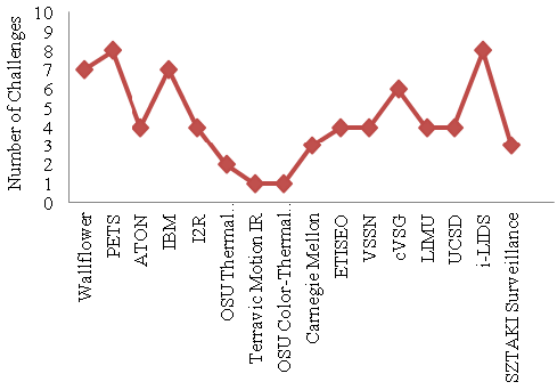

FIGURE 17. Challenges count in video datasets (1999-2010).

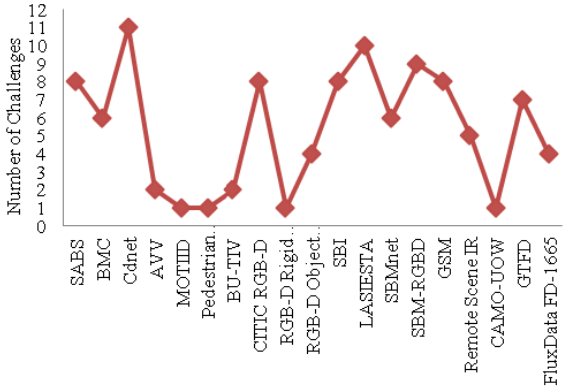

FIGURE 18. Challenges count in video datasets (2011-2017).

It is observed that CDnet dataset is the most prolific dataset in the list of background subtraction datasets dealing with a large number of challenges and publications. Although new datasets devoted to background subtraction have been created in recent years it is noted that the latest publications employed CDnet 2014 dataset for evaluation of models. As we can see, there are very few video datasets that cover existing challenges of background subtraction.

TABLE 16. Summary of Video Datasets Representing Ground Truth, Evaluation Domain and Application Details.

Name	Ground Truth	Evaluation Domain	Example Application Areas
Wallflower	Pixel-wise labeling	Background maintenance	Video surveillance [46], [47]
PETS	Bounding boxes	Object detection and tracking	Moving object detection [51],[52],[53], tracking [157]
ATON	Pixel-wise labeling	Shadow detection and analysis	Moving object and shadow detection for intelligent transport system [54],[55,[56]
IBM	Bounding boxes	Detection, tracking and classification	People counting [57], detection and tracking [58]
CAVIAR	Bounding boxes	Automated video surveillance and automatic behavior analysis	Human detection [64], Human tracking [63], [65]
I2R	Pixel-wise labeling	Background modeling	Moving object detection [67],[68],[69], visual surveillance [70],[71],[72]
OSU Thermal Pedestrian Database	Bounding boxes	Background subtraction	Pedestrians detection in Thermal Imagery [73]
Terravic Motion IR Database	-	Detection and tracking	Detection and tracking in thermal imagery
OSU Color-Thermal Database	Pixel-wise labeling	Background subtraction	Thermal and color fusion based object detection [74]
Carnegie Mellon	Pixel-wise labeling	Background subtraction	Foreground detection and enhancement [33]
ETISEO	Bounding boxes	Video surveillance	Moving object detection [78], tracking [79]
VSSN	Pixel-wise labeling	Video surveillance	Foreground segmentation [81],[82], [83]
BEHAVE	Bounding boxes	Event detection, crowd scenes analysis and tracking	Violence detection [85],[86]
cVSG	Pixel-wise labeling	Video segmentation	Human segmentation [88], static foreground classification [89]
LIMU	Pixel-wise labeling	Background subtraction	Moving object detection
UCSD	Foreground masks in 3d array format	Background subtraction	Moving object detection [91], [92]
i-LIDS	-	Event detection and multi-camera tracking	Traffic monitoring [94], [95], person re-identification [96],[97]
SZTAKI Surveillance	Pixel-wise labeling	Background subtraction	Foreground and shadow detection [98],[99], [100], [101]
SABS	Pixel-wise labeling	Background subtraction	Moving object detection [42],[70], [107]
AVV	-	Detection and classification	Moving Vehicle detection and classification [121]
MOTIID	-	Moving object detection	Background subtraction in infrared imagery [123]
Pedestrian Infrared/ Visible Stereo Video Dataset	Pixel-wise labeling	Pedestrian registration in visible and infrared videos	Human silhouettes registration [124]
BU-TIV	-	Tracking and counting in infrared videos	Detection, Tracking and Counting
FluxData FD-1665	Pixel-wise labeling	Background subtraction	Moving object detection [202]
Remote Scene IR	Pixel-wise labeling	Background subtraction	Foreground detection [160], [161], [162]
CAMO-UOW	Pixel-wise labeling	Camouflaged foreground detection	Foreground detection [175], [176]
BMC	Pixel-wise labeling	Background subtraction	Moving object detection[19],[109], [110]
CDnet 2012	Pixel-wise labeling	Background subtraction	Foreground detection [112], [113]
CDnet 2014	Pixel-wise labeling	Background subtraction	Foreground detection [115], [116], [117], [118], [119], [120]
LASIESTA	Pixel-wise labeling	Background subtraction	Moving object detection [141], [142], [143], [144], [145]
GTFD	Pixel-wise labeling	Background subtraction	Foreground detection [178],[179], [180]
CITIC RGB-D	Pixel-wise labeling	Background subtraction	Foreground segmentation [127], [128], [129]
RGB-D Rigid Multi-Body	Pixel-wise labeling	Motion estimation and segmentation	Background subtraction [131]
RGB-D Object Detection	Pixel-wise labeling	Background subtraction	Foreground and background segmentation [132]
SBM-RGBD	Pixel-wise labeling	Background modeling	Foreground segmentation [152]
GSM	Pixel-wise labeling	Background subtraction	Foreground and background segmentation [155]
SBI	Background frame	Background initialization	Background initialization [134],[135], [136], [137], [138], [139]
SBMnet	Background frame	Background initialization	Background initialization [146], [147], [148], Background and foreground separation [149], Background estimation [150]

The leader background subtraction methods on significant benchmark datasets are presented in Table 18. Different evaluation metrics are employed to evaluate the state-of-the-art algorithms on benchmark datasets in order to rank the algorithms. The evaluation metrics used to measure the performance of the background subtraction algorithms are discussed in Section VII. The average F-measure of leader methods which outperform other background subtraction

TABLE 17. Background subtraction video datasets classification according to different challenges.

Challenges	Datasets
Dynamic Background	Wallflower, PETS, IBM, I2R, VSSN, ETISEO, cVSG, i-LIDS, LIMU, UCSD, BMC, SABS, CDnet 2012, SBI, LASIESTA, Remote Scene IR, GTFD
Shadows	PETS, I2R, ATON, cVSG, Carnegie Mellon, i-LIDS, VSSN, ETISEO, SZTAKI Surveillance, LIMU, BMC, SABS, CDnet 2012, CITIC RGB-D, RGB-D object detection, SBI, LASIESTA, SBM-RGBD, GSM, GTFD, FluxData FD-1665
Gradual Illumination changes	Wallflower, PETS, IBM, I2R, VSSN, ETISEO, SZTAKI Surveillance, ATON, i-LIDS, LIMU, BMC, SABS, CITIC RGB-D, SBI, LASIESTA, SBMnet, SBM-RGBD, GSM, FluxData FD-1665
Sudden Illumination changes	Wallflower, PETS, IBM, I2R, VSSN, ATON, i-LIDS, SZTAKI Surveillance, LIMU, BMC, SABS, CITIC RGB-D, SBI, LASIESTA, SBMnet, SBM-RGBD, GSM
Occlusion	PETS, IBM, ETISEO, cVSG, Carnegie Mellon, i-LIDS, Audio-Visual Vehicle, CITIC RGB-D, SBI, LASIESTA
Color Camouflage	Wallflower, SABS, CITIC RGB-D, RGB-D object detection, LASIESTA, SBM-RGBD, GSM, Remote Scene IR, CAMO-UOW, FluxData FD-1665
Depth Camouflage	RGB-D object detection, SBM-RGBD, GSM, CITIC RGB-D
Bootstrapping	Wallflower, VSSN, SABS, LASIESTA, SBM-RGBD, GSM
Challenging Weather	IBM, ETISEO, i-LIDS, UCSD, OSU Thermal Pedestrian Database, BMC, CDnet 2014, LASIESTA, GTFD
Intermittent Object Motion	cVSG, CDnet 2012, SBI, SBMnet, SBM-RGBD, GTFD, FluxData FD-1665
Camera Jitter	UCSD, CDnet 2012, SBI, SBMnet,, Remote Scene IR
Foreground Aperture	Wallflower
Moved Background	Wallflower, IBM, RGB-D object detection, GSM
Inserted Background	CITIC RGB-D
Sleeping Foreground	cVSG, SBI, LASIESTA
Background Motion	IBM, UCSD, SBMnet
Air Turbulence	CDnet 2014
Clutter	PETS, IBM, i-LIDS, Thermal Infrared Video Benchmark for Visual Analysis , SBMnet, GTFD
Foreground Size	cVSG, BMC, RGB-D Rigid Multi-Body, Remote Scene IR
Video Noise	SABS, Remote Scene IR
Low Frame-rate	CDnet 2014
Thermal Videos	PETS, i-LIDS, CDnet 2012, OSU Thermal Pedestrian Database, OSU Color-Thermal Database, Terravic Motion IR Database, CSIR-CSIO Moving Object Thermal Infrared Imagery Dataset, Pedestrian Infrared/Visible Stereo Video Dataset, Thermal Infrared Video Benchmark for Visual Analysis, GTFD
Night Videos	SABS, CDnet 2014
Pan-tilt-zoom Cameras	ATON, Audio-Visual Vehicle, CDnet 2014
Moving Cameras	Carnegie Mellon, PETS, LASIESTA
Out of Sensor Range	SBM-RGBD
Low Light	CITIC RGB-D, GTFD

algorithms on BMC, CDnet 2012, CDnet 2014, LASIESTA, GTFD, Remote Scene IR, and SBM-RGBD benchmarks are mentioned in Table 18. For SBI and SBMnet datasets, Multi-scale Structural Similarity Index (MS-SSIM), Peak Signal-Noise Ratio (PSNR), and Color Image Quality Measure (CQM) are mentioned. The measures come from the corresponding papers and results to date come from the benchmark dataset website. A multitude of algorithms have been proposed for background subtraction in the literature that can perform efficiently in real-time applications, but still, there are some unsolved issues that require significant attention.

VI. APPLICATION SPECIFIC VIDEO DATASETS

The video datasets described in the previous sections are dedicated to the task of visual surveillance of human activities in the general environment. There are other video datasets for background subtraction in the literature which were recorded in different environs to focus on specific applications. These types of video datasets are included here in a summarized manner.

A. AQU@THEQUE DATASET

The Aqu@theque dataset [180] was created in 2007 to detect and recognize fish species. It contains 5 different image sequences which were filmed in a tank of an aquarium and covered major challenges of background subtraction such as bootstrapping, camouflage, illumination variations, occlusion, and dynamic background. The Aqu@theque dataset is available on email request to the author. This dataset was designed to address the problems of background subtraction for automatic detection, recognition, and behavioral analysis of fish species in an aquarium.

B. FISH4KNOWLEDGE DATASET

The Fish4Knowledge dataset [102] was created as a part of the Fish4Knowledge project, (2010-2013), to investigate target detection algorithm against a complex background of undersea water and to provide the research community with a large scale labeled fish dataset for environmental and behavioral studies. This underwater benchmark dataset consists of 14 video sequences that are categorized into 7 different classes representing key challenges in background modeling. The challenges in the dataset include occlusion, complex textures of background, low contrasted video frames, camouflage, dynamic background, and illumination changes. Videos in this dataset were recorded at a different time of day to test the performance of the algorithm on all environmental conditions. For ground truth data, about 30 video frames from

TABLE 18. Leader methods on significant benchmark datasets.

Datasets	Leader Method	Technique	Performance
BMC	BC [209]	Bayesian classification processed on feature statistics	F-measure = 0.93
CDnet 2012	PAWCS [210]	Word-based model for background subtraction (Integration of local binary similarity patterns features with color values at pixel level including feedback mechanism)	F-measure = 0.8579
CDnet 2014	FgSegNet-V2 [171]	Convolutional neural network (CNN) followed by a feature fusion into feature pooling module for background subtraction	F-measure = 0.9847
LASIESTA	SC-SOBS [211]	Neural network (Self-organizing background subtraction method)	F-measure = 0.7842
GTFD	F-WELD [177]	Low-rank representation on background modeling	F-measure = 0.73
Remote Scene IR	Sigma-delta [212]	Estimation model (Background estimation by $\sum$-Δ filter)	F-measure = 0.5037
SBM-RGBD	SCAD [213]	Fusion of appearance and depth information for background subtraction	F-measure = 0.8757
SBI	SC-SOBS [211]	Neural network (Self-organizing background subtraction method)	MS-SSIM = 0.9765 PSNR = 35.2723 CQM = 50.1138
SBMnet	MSCL [214]	Robust Subspace learning model (Spatial and temporal clustering into robust principal component analysis (RPCA) for background modeling)	MS-SSIM = 0.9410 PSNR = 30.8952 CQM = 31.7049

each video are manually labeled. More than 3500 objects were labeled and provided as binary masks. This dataset is used in [103] and [104] for detecting underwater moving objects using background subtraction.

C. UNDERWATER CHANGE DETECTION DATASET

The Underwater Change Detection dataset [181] is a collection of 5 videos with pixel-based manually segmented ground truth video frames. This dataset was created to evaluate moving object detection algorithms against the difficulties of an underwater environment. Fishes are considered as a foreground and all other as background. Videos were recorded at a different time of the day to cover the major challenges of background subtraction. The video scenes with illumination variations, dynamic background, strong shadows, camouflage, challenging weather (marine snow), and bad lighting conditions were included in the dataset.

D. CCT DATASET

The Caltech Camera Traps (CCT) dataset [184] was created to study the problem of generalization in the detection and classification of animals in an unfamiliar environment. It consists of 243,187 images from 140 camera locations. The camera traps are placed at different locations of interest for monitoring purpose and behavioral studies of the animal population. Several challenges provided by the camera trap data such as bad weather conditions, occlusion, motion blur, discolorations due to camera malfunctions, small region of interest, poor illumination, forced perspective, non-animal variability, and temporal changes in background were included in the CCT dataset. The subset of this dataset was used in [185] to investigate how well the state-of-the-art detection and classification algorithms generalize to the unseen environment. The class-level annotations for all images and the bounding box annotations for a subset of images are provided as ground truth data.

E. CRIM13 DATASET

The Caltech Resident-Intruder Mouse (CRIM13) dataset [186] contains 237*2 videos of 10 minutes (approx). Each video consists of pairs of mice engaging in social behavior, recorded with two static cameras synchronized in a way to cover both top and side views. Social behavior in mice surveillance video scenes was categorized into 13 mutually exclusive actions. Each video is annotated frame-by-frame by the team of behavior experts. The goal of this dataset is to provide continuous videos to analyze social behavior.

F. MARDCT DATASET

The Maritime Detection, Classification, and Tracking (MarDCT) dataset [182] is divided into three classes based on the type of ground truth data: *Detection, Classification, and Tracking*.The objective of this dataset is to evaluate different computer vision techniques for intelligent surveillance of maritime environment. It is a collection of videos and images. The information about location, camera, reflections and time of the day with ground truth data are also available [183]. The videos were recorded with different cameras such as static, moving, and pan-tilt-zoom (PTZ) at a different time of the day to incorporate major challenges related to water background. The foreground masks, bounding boxes, and identification numbers were included as ground truth annotations with videos and images in the dataset.

G. MUHAVI-MAS DATASET

The Multicamera Human Action Video and Manually Annotated Silhouette (MuHAVi-MAS) Data [187] is multi-action dataset developed specifically for evaluation of human action recognition methods. It consists of 17 human action classes performed by 14 actors. The video scenes were recorded in a realistic site with challenging illumination conditions provided by multiple sources of night street lights using 8 CCTV cameras. From 1904 video segments, 952 video segments are provided to the research community in order

to investigate segmentation and human action recognition methods. Videos from MuHAVi data and manually annotated silhouettes from MAS data are protected with username and password. A subset of sample videos and annotated data are publically available [188].

H. HUMANEVA DATASET

There are two versions of HumanEva dataset [189]: HumanEva-I and HumanEva-II, designed to provide a testbed for the research community to investigate the state-of-the-art methods and unsolved problems in human pose estimation and tracking. It contains 80,000 video frames (approx) in 56 video sequences. The video scenes were recorded in a laboratory setting by using 4 grayscale and 3 color video cameras. The cameras were synchronized to cover multiple subjects performing a set of actions such as walking, boxing, and jogging. The HumanEva-I dataset contains 6 set of predefined actions performed by 4 subjects. The HumanEva-II dataset contains an extended set of actions performed by only 2 subjects. The challenges for the evaluation of background subtraction algorithms on this dataset include strong illumination and shadows.

I. KINDERGARTEN VIDEO DATASET

The Kindergarten video dataset [190] was created to cover realistic video sequences in order to evaluate kindergarten video surveillance system. It consists of 100 videos (approx) ranging from 1 minute to 30 minutes. The video scenes were recorded in both indoor and outdoor environs of kindergarten. This dataset can be used to investigate major challenges of background subtraction such as shadows, illumination variations, sleeping foreground objects, noise, camouflage, and moved background objects. Abdelhedi *et al.* [191] proposed an automatic fall detection system in order to monitor children falls in the kindergarten and used kindergarten video dataset for evaluation purpose.

J. EDINBURGH CEILIDH OVERHEAD VIDEO DATA DATASET

The Edinburgh Ceilidh Overhead Video Data [192] is a collection of 16 dance videos with two different dance styles, recorded in 2016 at the University of Edinburgh. This dataset contains video data, individual video frames, and annotated ground truth file. There are 4,577 total video frames in the dataset. The ground truth information includes the labeled position and the current state of each dancer in the vide frame. A highly structured human behavior was filmed in order to evaluate segmentation and action recognition algorithms on complex video scenes.

VII. PERFORMANCE MEASURE IN BACKGROUND SUBTRACTION

The quality assessment of background subtraction algorithm is essential in order to check its validity. Finding an accurate evaluation metric to evaluate the performance of a method which detects moving objects from the video sequences is not trivial. In this section, the standard evaluation metrics used for the quality evaluation of background subtraction algorithms are discussed. The evaluation metrics based on the number of true positives, false positives, true negatives, and false negatives for performance evaluation of background subtraction algorithms are presented in Table 19. True positives (TP) is the number of foreground pixels classified as foreground, False positives (FP) is the number of background pixels classified as foreground, True negatives (TN) is the number of background pixels classified as background, and False negatives (FN) is the number of foreground pixels classified as background.

TABLE 19. Evaluation metrics for background subtraction.

Metrics	Description
Recall (Re)	TP/ (TP + FN)
Precision (Pr)	TP/ (TP + FP)
F-Measure	2 (Pr. Re) / (Pr + Re)
Specificity (Sp)	TN / (TN + FP)
False Positive Rate (FPR)	FP / (FP + TN)
False Negative Rate (FNR)	FN / (TN + FP)
Percentage of Wrong Classification (PWC)	100 (FN + FP) / (TP + FN + FP + TN)

The classical evaluation metrics such as Recall (Re), Precision (Pr), F-measure, Specificity (Sp), False Positive Rate (FPR), False Negative Rate (FNR), and Percentage of Wrong Classification (PWC) described in Table 19 are widely adopted in the literature for performance evaluation of background subtraction algorithms. The higher the recall, specificity, precision, and F-measure values, the better the background subtraction algorithm. The significant datasets such as CDnet dataset and SBM-RGBD dataset employed all the seven classical evaluation metrics to benchmark the state-of-the-art background subtraction methods. There are other advanced quality metrics in the literature such as PSNR (Peak Signal-Noise Ratio), SSIM (Structural SIMilarity), and D-score. F-measure and PSNR are used to compare the raw behavior of each background subtraction algorithms for moving object detection in BMC dataset [108]. Let I be the set of n images and G be the ground truth video sequences. Then PSNR is defined as

$$\frac{1}{n}\sum_{i=1}^{n} 10\log_{10}\frac{m}{\sum_{j=1}^{m}||I_i(j)-G_i(j)||^2}, \quad (1)$$

where $I_i(j)$ denotes the jth pixel of image i of size m in the sequence I of length n.

PSNR is simple to compute and convenient in the context of optimization. But it is not very well matched to rate visual quality [197]. Another quality metric, SSIM, used for the perceptual assessment of background subtraction methods is defined as

$$SSIM(I,G)=\frac{1}{n}\sum_{i=1}^{n}\frac{(2\mu_{I_i}\mu_{G_i}+c_1)(2cov_{I_iG_i}+c_2)}{(\mu_{I_i}^2+\mu_{G_i}^2+c_1)(\sigma_{I_i}^2+\sigma_{G_i}^2+c_2)}, \quad (2)$$

where μ_{I_i}, μ_{G_i} denote the means, $cov_{I_iG_i}$ the covariance of I_i and G_i, and $\sigma_{I_i}^2, \sigma_{G_i}^2$ the standard deviations. In BMC

benchmark dataset, c_1 and c_2 is defined as: $c_1 = (k_1 \times L)^2$ and $c_1 = (k_1 \times L)^2$, where $L = 255$ for gray-scale images, $k_1 = 0.01$ and $k_2 = 0.03$.

D-score [198] provides dissimilarity criterion between the segmentation result and the ground truth video frame and is defined as

$$\mathbf{D - score}\left(\mathbf{I_i}\left(\mathbf{j}\right)\right) = \mathbf{exp}\left(\left(-\log_2\left(\mathbf{2.DT}\left(\mathbf{I_i}\left(\mathbf{j}\right)\right) - \mathbf{5/2}\right)^2\right)\right), \tag{3}$$

where $DT(I_i(j))$ is given by the minimum distance between the pixel $I_i(j)$ and the nearest reference point. The lower the D-score value, the higher the significance of background subtraction algorithm. The classical evaluation metrics consider errors in the same manner regardless of their localization. But, D-score takes into account the localization and type of errors in relation to real object positions. In addition, there are other frequently used quality metrics for the evaluation of background initialization algorithms such as Average Gray-level Error (AGE), Percentage of Clustered Error Pixels (pCEPs), Percentage of Error Pixels (pEPs), Multi-scale Structural Similarity Index (MS-SSIM), and Color Image Quality Measure (CQM). The detailed description of evaluation metrics for background initialization can be found in [29]. These evaluation metrics measure the visual exactness of an estimated background image against a ground truth background image.

VIII. BACKGROUND SUBTRACTION LIBRARIES

There are a number of libraries available for background subtraction using different algorithms in the field. The first one provides three background subtraction algorithms and is under OpenCV [199]. Parks and Fels [200] developed a library that offers seven background subtraction algorithms. These background subtraction algorithms were evaluated using Visual Microsoft C and OpenCV. Another background subtraction library, Scene [202], developed by Bender and Guerra using OpenCV offers five background subtraction algorithms. It is an open source multiplatform framework designed to perform background subtraction using two traditional algorithms based on gaussian models and three algorithms based on neural networks and fuzzy classification rules.

More recently, the BGSLibrary [203] was developed by Sobral and Bouwmans [204] based on OpenCV. It provides a C++ framework to perform moving object detection using background subtraction algorithms. This free and open source library currently contains more than 43 algorithms. A graphical user interface (BGSLibraryGUI) developed to configure and run BGSLibaray is also available for download on BGSLibrary website [205]. The LRSLibrary [206] is a collection of more than 100 low-rank and sparse decomposition algorithms for motion segmentation in videos. It is implemented in MATLAB and also provides a graphical user interface for background modeling and subtraction [207].

IX. CURRENT RESEARCH TRENDS AND FUTURE RESEARCH DIRECTIONS

Deep neural network based systems are increasingly used in intelligent video analytics and especially for object detection, scene labeling and image classification [163], [164]. Recently, deep neural networks have shown noteworthy results for background subtraction and improved the area of foreground detection. Bouwmans *et al.* [162] provided a detailed survey of recent advances on deep neural networks and presented the comparative evaluation of deep neural network based methods for background subtraction on CDnet 2014 dataset. The convolutional neural network based background subtraction has outperformed the conventional background subtraction algorithms and become the leading methods on a prolific dataset.

The use of convolutional neural networks (CNNs) for background subtraction was first attempted by Braham and Van Droogenbroeck in 2016 and named the model as ConvNet [165]. First, a single grayscale background image is extracted by computing the temporal median value for each pixel on 150 video frames to construct the background model and then the proposed CNN is trained with a scene-specific dataset. To perform the pixel classification task, two patches from the input frame and the background model is feed into the trained network as an input and then a subtraction operation is performed by ConvNet. Experiments on CDnet 2014 dataset output similar results as other state-of-the-art methods and show robust performance with the F-measure of 0.9454 and 0.7565 in case of hard shadows and night videos respectively. Practically, ConvNet works for scene-specific background subtraction. The videos under PTZ and intermittent object motion video category were not considered for the evaluation of the proposed model. The work presented in [165] directs the researchers to further use deep learning and improve the efficiency of deep neural networks for background subtraction. Lim *et al.* [166] designed an encoder-decoder structured convolutional neural network for background subtraction. A set of grayscale images: a target frame, its previous frame, and a background model are concatenated and fed into the network as input. Temporal median filtering is employed on the first 101 video frames to construct an initial background model. The VGG-16 network [167] was modified to design the encoder of the proposed network. The CDnet 2014 dataset is used for training and testing purpose. The encoder-decoder structured CNN produces the segmented foreground as an output which is further refined using super- pixel information to eliminate holes and incorrect boundaries. Experiments on CDnet 2014 dataset show that this method outperforms established background subtraction algorithms in most of the video categories like intermittent object motion, camera jitter, low frame rate, bad weather, and thermal imagery. The videos under PTZ video category were excluded as the algorithm is particularly designed for static cameras.

A semi-automatic method named Cascaded CNN based on multi-scale CNN with a cascaded architecture was designed by Wang *et al.* [118] in order to segment moving objects from the surveillance videos. First, the moving objects are manually segmented from the set of selected training video frames. Secondly, these video frames are used to train the CNN and then generalization is employed by CNN to segment the remaining video frames. The Cascaded CNN [118] was proposed to generate accurate segmentation maps with a little user intervention so that they can be used as ground truth in order to validate background subtraction algorithms. Experiments were conducted on two datasets: CDnet 2014 dataset and SBI dataset [133]. This method is committed to interactive ground-truth generation and is computationally expensive technique. It requires more training frames in the case of complex videos such as Night videos and PTZ videos. The performance drops when segmenting small foreground objects from the surveillance videos.

Babaee *et al.* [47] utilized the outputs of the existing background subtraction algorithms for background model generation and designed a deep convolutional neural network based background subtraction model for moving object detection. The background image was generated by combining the segmentation mask from [36] and the output of [168]. The pairs of RGB image patches from video and background frames along with respective ground truth segmentation masks from CDnet 2014 dataset were used to train the CNN. The background images were obtained by the SuBSENSE algorithm [36]. The video scenes with frequent background changes were discarded from the training phase. The foreground segmentation is done by the proposed CNN followed by the spatial median filtering as a post-processing method. Experiments were conducted on three datasets: Wallflower dataset, CDnet 2014 dataset, and PETS 2009 dataset. The performance evaluation on Wallflower dataset shows that the proposed algorithm outperforms all established background subtraction algorithms. The results on CDnet 2014 dataset reflect that the proposed algorithm does not handle camouflage regions within foreground objects and produces cavities in the foreground mask. This method outputs poor segmentation for PTZ videos and low frame-rate videos. The performance drops significantly with the frequent changes in the background.

In another work, Lim and Keles [117] proposed a multi-inputs CNN based approach for moving object segmentation called FgSegNet-M. A triplet CNN and a Transposed Convolutional Neural Network (TCNN) fixed at the end of it in an encoder-decoder structure were employed for segmenting moving foregrounds. The first four blocks of the pre-trained VGG-16 network [167] were employed as the multi-scale feature encoder at the beginning of the proposed CNN. To map the features to a pixel-level foreground probability map, a decoder network was integrated at the end of the CNN. The final binary segmentation labels were obtained by applying thresholding to the foreground probability map. FgSegNet-M outperforms two CNN based background subtraction methods: Cascaded CNN [118] and DeepBS [47] and other conventional algorithms on CDnet 2014 dataset. Furthermore, Lim and Keles designed a single input CNN followed by a feature pooling module (FPM) and the TCNN for moving object segmentation called FgSegNet-S [169], a variant of FgSegNet-M [117]. This method was evaluated on three datasets: CDnet 2014 [114], SBI [133] and UCSD [91] background subtraction datasets. Experiments on CDnet 2014 show that the FgSegNet-S is less robust against camera motion than FgSegNet-M. For all other video categories, it performs slightly better than FgSegNet-M The performance of both FgSegNet models drop significantly with low frame-rate videos and PTZ videos. In further work, Lim and Keles [170] proposed a more robust network by incorporating feature fusion to feature pooling module (FPM) and named it as FgSegNet-V2. It does not work for low frame-rate video category. This version of FgSegNet method is robust against camera motion and improves the performance in PTZ videos. This method outperforms all other background subtraction algorithms and ranked as number one on the CDnet 2014 dataset.

Zheng *et al.* [171] proposed a background subtraction algorithm based on Bayesian Generative Adversarial Network (BSGAN) to deal with the challenges of sudden and slow illumination variations, non-stationary background, and ghost. The deep convolutional neural networks are utilized to construct the generator and the discriminator of the proposed Bayesian generative adversarial network. In this method, the background model is generated by applying median filtering technique and then a network is trained to classify and label each pixel as foreground and background. Experiments on CDnet 2014 dataset show that BSGAN outputs good segmentation results in most of the video categories. Furthermore, Zheng *et al.* [172] designed a parallel version of BSGAN called BPVGAN.

Wang *et al.* [213] proposed a CNN based approach for background subtraction on depth videos called BGSNet-D (BackGround Subtraction neural Networks for Depth video). The objective is to use only depth data for background subtraction in order to deal with scenarios where color information is not available. A preprocessing strategy based on min-max normalization was employed to process depth images and to reduce noise. Two-channel patches extracted from preprocessed input video frames and a corresponding background frame were fed into the CNN for feature extraction and classification. Experimental results on SBM-RGBD dataset reflect that the proposed method outperforms traditional algorithms which also use only depth data for background subtraction. The performance of background subtraction algorithms which utilize both color and depth information is efficient than BGSNet-D in certain scenarios.

The top six background subtraction methods on CDnet 2014 dataset with average F-measure are presented in Table 20. The average F-measure comes from the changedetection.net website. The leader methods for background subtraction on CDnet 2014 dataset are based on deep

TABLE 20. Leader background subtraction methods with average F-measure on CDnet 2014 dataset.

Methods	Year	Network Architecture	F-Measure
FgSegNet-V2 [171]	2018	Convolutional Neural Network	0.9847
FgSegNet-S [170]	2018	Convolutional Neural Network	0.9804
FgSegNet-M [117]	2018	Convolutional Neural Network	0.9770
BSPVGAN [173]	2018	Generative Adversarial Network	0.9501
BSGAN [172]	2018	Generative Adversarial Network	0.9339
Cascaded CNN [118]	2017	Convolutional Neural Network	0.9209

neural networks following supervised approaches. FgSegNet-V2 [170] tops the list with an average F-measure of 0.9847. In addition to deep neural network based background subtraction, the researchers are also trying to modify conventional background subtraction algorithms by incorporating different techniques. For instance, Chen *et al.* [42] exploited the hierarchical superpixel segmentation for background subtraction to handle ineluctable background motion. In another recent work, Jiang and Lu [113] adopted the concept of weighted samples with minimum-weight update policy for background model initialization and updation. Many background subtraction approaches for moving object detection have been developed and have achieved promising results, but there are still some significant unsolved challenges that need to be considered for robust background subtraction. Some future research directions in background subtraction are as follows:

1) Deep neural networks need to be more exploited for background initialization and background subtraction in order to handle complex backgrounds for real-time applications.
2) A robust and efficient algorithm is required to deal with PTZ videos, Night videos, and low frame-rate video categories. Even the performance of top rated background subtraction algorithms drops significantly while testing on above-mentioned video categories.
3) A large video dataset with wide variety of video categories including thermal videos, low frame-rate videos, strong illumination changes, color and depth camouflage, night videos, intermittent object motion, hard shadows, dynamic backgrounds, and moving camera along with large number of annotated images is needed for the exhaustive evaluation of background subtraction methods.
4) The leading deep convolutional neural networks based background subtraction algorithms work with RGB images. The deep convolutional neural networks need to be more exploited to work with RGB-D images or depth images in order to utilize the advantages of depth information for background subtraction.
5) Most of the background subtraction methods are specifically designed for static cameras and show poor results in case of moving cameras. Thus, conventional algorithms can be modified to expect good segmentation results for moving cameras.
6) The top background subtraction methods on CDnet 2014 dataset are based on supervised learning. Deep neural networks have proven track records in supervised learning tasks. Unsupervised and semi-supervised methods need to be more explored for effective background subtraction. Graph Memory Networks [214], architecture for connecting neural networks to structural knowledge graphs provide a constructive solution to a wide range of problems. The graph-based semi-supervised method has the ability to work with multiple architectures. The researchers can explore graph memory networks for background subtraction.

X. CONCLUSION

The development of video datasets devoted to background subtraction helps the research community to check their algorithms against all major and latest challenges in the area of moving object detection. However, tremendous growth in the number of video datasets increases the difficulty in selecting datasets pertinent to background subtraction. It is also observed that more than one dataset has been used for evaluation in most of the published works.

This survey attempts to provide thorough details of the most important datasets for background subtraction. A comparison of video datasets for background subtraction dedicated to studying the intelligent visual surveillance of human activities, highlighting essential characteristics such as a type of dataset, total number of videos, type of video scenes, kind of sensor used, ground truth details, and area of applications is provided to assist researchers in the selection of the most appropriate dataset. The reference to a paper that provides the details of dataset, the link to each dataset and references of published works using these datasets are also reported in this survey to save both time and efforts of researchers. Although many datasets were introduced for background subtraction, in our opinion still there is a need for general dataset that covers all the challenges of background subtraction to evaluate real-time video scenes. It is also evident from the literature that none of the background subtraction algorithms is effective to address all the key challenges simultaneously. This paves way for the new research and general video dataset in this area.

REFERENCES

[1] H. Liu, S. Chen, and N. Kubota, "Intelligent video systems and analytics: A survey," *IEEE Trans. Ind. Informat.*, vol. 9, no. 3, pp. 1222–1233, Aug. 2013.
[2] L. Unzueta, M. Nieto, A. Cortes, J. Barandiaran, O. Otaegui, and P. Sanchez, "Adaptive multicue background subtraction for robust vehicle counting and classification," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 527–540, Jun. 2012.

[3] S. Lee, N. Kim, I. Paek, M. H. Hayes, and J. Paik, "Moving object detection using unstable camera for consumer surveillance systems," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Las Vegas, NV, USA, Jan. 2013, pp. 145–146.

[4] S.-C. S. Cheung and C. Kamath, "Robust background subtraction with foreground validation for urban traffic video," *EURASIP J. Adv. Signal Process.*, vol. 2005, no. 14, Dec. 2005, Art. no. 726261.

[5] M. Cristani, M. Farenzena, D. Bloisi, and V. Murino, "Background subtraction for automated multisensor surveillance: A comprehensive review," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, Dec. 2010, Art. no. 343057.

[6] G. Guerra-Filho, "Optical motion capture: Theory and implementation," in *Proc. RITA*, 2005, vol. 12, no. 2, pp. 61–90.

[7] I. Mikić, M. Trivedi, E. Hunter, and P. Cosman, "Human body model acquisition and tracking using voxel data," *Int. J. Comput. Vis.*, vol. 53, no. 3, pp. 199–223, Jul. 2003.

[8] S. Darabi, E. Shechtman, C. Barnes, D. B. Goldman, and P. Sen, "Image melding: Combining inconsistent images using patch-based synthesis," *ACM Trans. Graph.*, vol. 31, no. 4, Jul. 2012, Art. no. 82.

[9] A. Agarwala *et al.*, "Interactive digital photomontage," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 294–302, Aug. 2004.

[10] A. Colombari, M. Cristani, V. Murino, and A. Fusiello, "Exemplar-based background model initialization," in *Proc. 3rd ACM Int. Workshop Video Surveill. Sensor Netw.*, 2005, pp. 29–36.

[11] K.-L. Hung and S.-C. Lai, "Exemplar-based video inpainting approach using temporal relationship of consecutive frames," in *Proc. IEEE 8th Int. Conf. Awareness Sci. Technol. (iCAST)*, Taichung, Taiwan, Nov. 2017, pp. 373–378.

[12] V. Shetty, S. Vishwakarma, Harisha, and A. Agrawal, "Design and implementation of video synopsis using online video inpainting," in *Proc. 2nd IEEE Int. Conf. Recent Trends Electron., Inf. Commun. Technol. (RTEICT)*, Bangalore, India, May 2017, pp. 1208–1212.

[13] R. Hoseinnezhad, B.-N. Vo, and B.-T. Vo, "Visual tracking in background subtracted image sequences via multi-Bernoulli filtering," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 392–397, Jan. 2013.

[14] R. G. Abbott and L. R. Williams, "Multiple target tracking with lazy background subtraction and connected components analysis," *Mach. Vis. Appl.*, vol. 20, no. 2, pp. 93–101, Feb. 2009.

[15] Y. Yang, J. Yang, L. Liu, and N. Wu, "High-speed target tracking system based on a hierarchical parallel vision processor and gray-level LBP algorithm," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 6, pp. 950–964, Jun. 2017.

[16] H. Kim and H.-J. Lee, "A low-power surveillance video coding system with early background subtraction and adaptive frame memory compression," *IEEE Trans. Consum. Electron.*, vol. 63, no. 4, pp. 359–367, Nov. 2017.

[17] S. Chakraborty, M. Paul, M. Murshed, and M. Ali, "Adaptive weighted non-parametric background model for efficient video coding," *Neurocomputing*, vol. 226, pp. 35–45, Feb. 2017.

[18] F. El Baf, T. Bouwmans, and B. Vachon, "Comparison of background subtraction methods for a multimedia learning space," in *Proc. SIGMAP*, Jul. 2007, pp. 153–158.

[19] S. Chen, T. Xu, D. Li, J. Zhang, and S. Jiang, "Moving object detection using scanning camera on a high-precision intelligent holder," *Sensors*, vol. 16, no. 10, p. 1758, Oct. 2016.

[20] T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: An overview," *Comput. Sci. Rev.*, vol. 11, pp. 31–66, May 2014.

[21] S. K. Choudhury, P. K. Sa, S. Bakshi, and B. Majhi, "An evaluation of background subtraction for object detection vis-a-vis mitigating challenging scenarios," *IEEE Access*, vol. 4, pp. 6133–6150, 2017.

[22] Y. Xu, J. Dong, B. Zhang, and D. Xu, "Background modeling methods in video analysis: A review and comparative evaluation," *CAAI Trans. Intell. Technol.*, vol. 1, no. 1, pp. 43–60, Jan. 2016.

[23] Y. Benezeth, P.-M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, "Comparative study of background subtraction algorithms," *Proc. SPIE*, vol. 19, no. 3, Jul. 2010, Art. no. 033003.

[24] T. Bouwmans, A. Sobral, S. Javed, S. K. Jung, and E.-H. Zahzah, "Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset," *Comput. Sci. Rev.*, vol. 23, pp. 1–71, Feb. 2017.

[25] D. Ortego, J. C. SanMiguel, and J. M. Martínez, "Stand-alone quality estimation of background subtraction algorithms," *Comput. Vis. Image Understand.*, vol. 162, pp. 87–102, Sep. 2017.

[26] T. Bouwmans, "Recent advanced statistical background modeling for foreground detection-a systematic survey," *Recent Patents Comput. Sci.*, vol. 4, no. 3, pp. 147–176, 2011.

[27] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Comput. Vis. Image Understand.*, vol. 122, pp. 4–21, May 2014.

[28] T. Bouwmans, L. Maddalena, and A. Petrosino, "Scene background initialization: A taxonomy," *Pattern Recognit. Lett.*, vol. 96, pp. 3–11, Sep. 2017.

[29] P.-M. Jodoin, L. Maddalena, A. Petrosino, and Y. Wang, "Extensive benchmark and survey of modeling methods for scene background initialization," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5244–5256, Nov. 2017.

[30] J. Xiang, H. Fan, H. Liao, J. Xu, W. Sun, and S. Yu, "Moving object detection and shadow removing under changing illumination condition," *Math. Problems Eng.*, vol. 2014, Feb. 2014, Art. no. 827461.

[31] S. Chen, J. Zhang, Y. Li, and J. Zhang, "A hierarchical model incorporating segmented regions and pixel descriptors for video background subtraction," *IEEE Trans. Ind. Informat.*, vol. 8, no. 1, pp. 118–127, Feb. 2012.

[32] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, "Detecting moving shadows: Algorithms and evaluation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 7, pp. 918–923, Jul. 2003.

[33] T. Akilan, Q. M. J. Wu, and Y. Yang, "Fusion-based foreground enhancement for background subtraction using multivariate multi-model Gaussian distribution," *Inf. Sci.*, vols. 430–431, pp. 414–431, Mar. 2018.

[34] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011.

[35] X. Zhang, C. Zhu, S. Wang, Y. Liu, and M. Ye, "A Bayesian approach to camouflaged moving object detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 9, pp. 2001–2013, Sep. 2017.

[36] P. L. St-Charles, G. A. Bilodeau, and R. Bergevin, "SuBSENSE: A universal change detection method with local adaptive sensitivity," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 359–373, Jan. 2015.

[37] Y. Wu, X. He, and T. Q. Nguyen, "Moving object detection with a freely moving camera via background motion subtraction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 2, pp. 236–248, Feb. 2017.

[38] M. Yazdi and T. Bouwmans, "New trends on moving object detection in video images captured by a moving camera: A survey," *Comput. Sci. Rev.*, vol. 28, pp. 157–177, May 2018.

[39] P. Christiansen, L. Nielsen, K. Steen, R. Jørgensen, and H. Karstoft, "DeepAnomaly: Combining background subtraction and deep learning for detecting obstacles and anomalies in an agricultural field," *Sensors*, vol. 6, no. 11, p. 1904, Nov. 2016.

[40] S. Varadarajan, P. Miller, and H. Zhou, "Region-based mixture of Gaussians modelling for foreground detection in dynamic scenes," *Pattern Recognit.*, vol. 48, no. 11, pp. 3488–3503, Nov. 2015.

[41] D. Avola, M. Bernardi, L. Cinque, G. L. Foresti, and C. Massaroni, "Adaptive bootstrapping management by keypoint clustering for background initialization," *Pattern Recognit. Lett.*, vol. 100, pp. 110–116, Dec. 2017.

[42] M. Chen, X. Wei, Q. Yang, Q. Li, G. Wang, and M.-H. Yang, "Spatiotemporal GMM for background subtraction with superpixel hierarchy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1518–1525, Jun. 2018.

[43] Y. Lin, Y. Tong, Y. Cao, Y. Zhou, and S. Wang, "Visual-attention-based background modeling for detecting infrequently moving objects," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 6, pp. 1208–1221, Jun. 2017.

[44] W. Kim and C. Jung, "Illumination-invariant background subtraction: Comparative review, models, and prospects," *IEEE Access*, vol. 5, pp. 8369–8384, 2017.

[45] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and practice of background maintenance," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Kerkyra, Greece, vol. 1, Sep. 1999, pp. 255–261.

[46] C.-H. Yeh, C.-Y. Lin, K. Muchtar, H.-E. Lai, and M.-T. Sun, "Three-pronged compensation and hysteresis thresholding for moving object detection in real-time video surveillance," *IEEE Trans. Ind. Electron.*, vol. 64, no. 6, pp. 4945–4955, Jun. 2017.

[47] M. Babaee, D. T. Dinh, and G. Rigoll, "A deep convolutional neural network for video sequence background subtraction," *Pattern Recognit.*, vol. 76, pp. 635–649, Apr. 2018.
[48] D. P. Young and J. M. Ferryman, "PETS metrics: On-line performance evaluation service," in *Proc. IEEE Int. Workshop Vis. Surveill. Perform. Eval. Tracking Surveill.*, Beijing, China, Oct. 2005, pp. 317–324.
[49] L. Patino, T. Cane, A. Vallee, and J. Ferryman, "PETS 2016: Dataset and challenge," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun./Jul. 2016, pp. 1240–1247.
[50] K. Vignesh, G. Yadav, and A. Sethi, "Abnormal event detection on BMTT-PETS 2017 surveillance challenge," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 2161–2168.
[51] Y. Zhang, X. Li, Z. Zhang, F. Wu, and L. Zhao, "Deep learning driven blockwise moving object detection with binary scene modeling," *Neurocomputing*, vol. 168, pp. 454–463, Nov. 2015.
[52] B. N. Subudhi, S. Ghosh, S. C. K. Shiu, and A. Ghosh, "Statistical feature bag based background subtraction for local change detection," *Inf. Sci.*, vol. 366, pp. 31–47, Oct. 2016.
[53] S. Sladojević, A. Anderla, D. Ćulibrk, D. Stefanović, and B. Lalić, "Integer arithmetic approximation of the HoG algorithm used for pedestrian detection," *Comput. Sci. Inf. Syst.*, vol. 14, no. 2, pp. 329–346, Jun. 2017.
[54] M. Trivedi, S. Bhonsle, and A. Gupta, "Database architecture for autonomous transportation agents for on-scene networked incident management (ATON)," in *Proc. ICPR*, Sep. 2000, pp. 664–667.
[55] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting objects, shadows and ghosts in video streams by exploiting color and motion information," in *Proc. 11th Int. Conf. Image Anal. Process.*, Palermo, Italy, Sep. 2001, pp. 360–365.
[56] I. Mikic, P. C. Cosman, G. T. Kogut, and M. M. Trivedi, "Moving shadow and object detection in traffic scenes," in *Proc. 15th Int. Conf. Pattern Recognit. (ICPR)*, Barcelona, Spain, vol. 1, Sep. 2000, pp. 321–324.
[57] P. Karpagavalli and A. V. Ramprasad, "Estimating the density of the people and counting the number of people in a crowd environment for human safety," in *Proc. Int. Conf. Commun. Signal Process.*, Melmaruvathur, India, Apr. 2013, pp. 663–667.
[58] J. Connell, A. W. Senior, A. Hampapur, Y.-L. Tian, L. Brown, and S. Pankanti, "Detection and tracking in the IBM peoplevision system," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Taipei, Taiwan, vol. 2, Jun. 2004, pp. 1403–1406.
[59] F. Pla, P. Ribeiro, J. Santos-Victor, and A. Bernardino, "Extracting motion features for visual human activity representation," in *Proc. Iberian Conf. Pattern Recognit. Image Anal.* Berlin, Germany: Springer, 2005, pp. 537–544.
[60] D. Hall *et al.*, "Comparison of target detection algorithms using adaptive background models," in *Proc. IEEE Int. Workshop Vis. Surveill. Perform. Eval. Tracking Surveill.*, Beijing, China, Oct. 2005, pp. 113–120.
[61] J. C. Nascimento, M. A. T. Figueiredo, and J. S. Marques, "Motion segmentation for activity surveillance?" in *Proc. ISR Workshop Syst., Decis. Control Robot. Monit. Surveill.*, Jun. 2005.
[62] D. Hall, "Automatic parameter regulation for a tracking system with an auto-critical function," in *Proc. 7th Int. Workshop Comput. Archit. Mach. Perception (CAMP)*, Palermo, Italy, Jul. 2005, pp. 39–45.
[63] H. S. Dadi, P. G. K. Mohan, and M. Latha, "Human tracking under severe occlusions," *i-Manager's J. Softw. Eng.*, vol. 12, no. 1, p. 29, Jan. 2017.
[64] B. C. Ko, M. Jeong, and J. Y. Nam, "Fast human detection for intelligent monitoring using surveillance visible sensors," *Sensors*, vol. 14, no. 11, pp. 21247–21257, Nov. 2014.
[65] P. Feng, W. Wang, S. M. Naqvi, and J. Chambers, "Adaptive retrodiction particle PHD filter for multiple human tracking," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1592–1596, Nov. 2016.
[66] L. Li, W. Huang, I. Y.-H. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection," *IEEE Trans. Image Process.*, vol. 13, no. 11, pp. 1459–1472, Nov. 2004.
[67] C. Guyon, T. Bouwmans, and E.-H. Zahzah, "Foreground detection based on low-rank and block-sparse matrix decomposition," in *Proc. ICIP*, Sep./Oct. 2012, pp. 1225–1228.
[68] S. E. Ebadi and E. Izquierdo, "Foreground segmentation with tree-structured sparse RPCA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 9, pp. 2273–2280, Sep. 2018.
[69] W. Hu, Y. Yang, W. Zhang, and Y. Xie, "Moving object detection using tensor-based low-rank and saliently fused-sparse decomposition," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 724–737, Feb. 2017.
[70] D. Berjón, C. Cuevas, F. Morán, and N. García, "Real-time nonparametric background subtraction with tracking-based foreground update," *Pattern Recognit.*, vol. 74, pp. 156–170, Feb. 2018.
[71] T. S. F. Haines and T. Xiang, "Background subtraction with Dirichletprocess mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 4, pp. 670–683, Apr. 2014.
[72] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1168–1177, Jul. 2008.
[73] J. W. Davis and M. A. Keck, "A two-stage template approach to person detection in thermal imagery," in *Proc. IEEE Workshops Appl. Comput. Vis. (WACV/MOTION)*, vol. 1, Jan. 2005, pp. 364–369.
[74] J. W. Davis and V. Sharma, "Background-subtraction using contour-based fusion of thermal and visible imagery," *Comput. Vis. Image Understand.*, vol. 106, no. 2, pp. 162–182, 2007.
[75] R. Miezianko, "Terravic research infrared database," in *Proc. IEEE OTCBVS WS Series Bench*, 2005. [Online]. Available: http://vcipl-okstate.org/pbvs/bench/
[76] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1778–1792, Nov. 2005.
[77] A. T. Nghiem, F. Bremond, M. Thonnat, and V. Valentin, "ETISEO, performance evaluation for video surveillance systems," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, London, U.K., Sep. 2007, pp. 476–481.
[78] S. Hwang, Y. Uh, M. Ki, K. Lim, D. Park, and H. Byun, "Real-time background subtraction based on GPGPU for high-resolution video surveillance," in *Proc. 11th Int. Conf. Ubiquitous Inf. Manage. Commun.*, 2017, Art. no. 109.
[79] Z. Zhang, O. P. Concha, and M. Piccardi, "Tracking people under heavy occlusions by layered data association," *Multimedia Tools Appl.*, vol. 74, no. 17, 7239-7259, Sep. 2015.
[80] S. Calderara, R. Melli, A. Prati, and R. Cucchiara, "Reliable background suppression for complex scenes," in *Proc. 4th ACM Int. Workshop Video Surveill. Sensor Netw.*, 2006, pp. 211–214.
[81] A. Nurhadiyatna, R. Wijayanti, and D. Fryantoni, "Extended Gaussian mixture model enhanced by hole filling algorithm (GMMHF) utilize GPU acceleration," in *Proc. Inf. Sci. Appl. (ICISA)*. Singapore: Springer, 2016, pp. 459–469.
[82] J. Li and Z. Miao, "Foreground segmentation for dynamic scenes with sudden illumination changes," *IET Image Process.*, vol. 6, no. 5, pp. 606–615, Jul. 2012.
[83] A. M. Hamad and N. Tsumura, "Background subtraction based on time-series clustering and statistical modeling," *Opt. Rev.*, vol. 19, no. 2, pp. 110–120, Mar. 2012.
[84] S. Blunsden and R. B. Fisher, "The BEHAVE video dataset: Ground truthed video for multi-person behavior classification," *Annu. BMVA*, vol. 4, nos. 1–12, p. 4, May 2010.
[85] T. Zhang, Z. Yang, W. Jia, B. Yang, J. Yang, and X. He, "A new method for violence detection in surveillance scenes," *Multimedia Tools Appl.*, vol. 75, no. 12, pp. 7327–7349, Jun. 2016.
[86] T. Zhang, W. Jia, B. Yang, J. Yang, X. He, and Z. Zheng, "MoWLD: A robust motion image descriptor for violence detection," *Multimedia Tools Appl.*, vol. 76, no. 1, pp. 1419–1438, Jan. 2017.
[87] F. Tiburzi, M. Escudero, J. Bescos, and J. M. Martinez, "A ground truth for motion-based video-object segmentation," in *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 17–20.
[88] A. Hernández-Vela, M. Reyes, V. Ponce, and S. Escalera, "Grabcut-based human segmentation in video sequences," *Sensors*, vol. 12, no. 11, pp. 15376–15393, Nov. 2012.
[89] A. Pereira, O. Saotome, and D. Sampaio, "Patch-based local histograms and contour estimation for static foreground classification," *EURASIP J. Image Video Process.*, vol. 2015, no. 1, p. 6, Dec. 2015.
[90] *PETS2001 Dataset*. Accessed: Dec. 20, 2018. [Online]. Available: http://www.cvg.reading.ac.uk/PETS2001/
[91] V. Mahadevan and N. Vasconcelos, "Background subtraction in highly dynamic scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, Jun. 2008, pp. 1–6.
[92] V. Mahadevan and N. Vasconcelos, "Spatiotemporal saliency in dynamic scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 171–177, Jan. 2010.
[93] *i-LIDS Dataset*. Accessed: Dec. 20, 2018. [Online]. Available: http://www.eecs.qmul.ac.uk/~andrea/avss2007_d.html

[94] W. Wahyono and K.-H. Jo, "Cumulative dual foreground differences for illegally parked vehicles detection," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2464–2473, Oct. 2017.
[95] K. Garg, A. Prakash, and T. Srikanthan, "Low complexity techniques for robust real-time traffic incident detection," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Yokohama, Japan, Oct. 2017, pp. 1–8.
[96] G. Lisanti, I. Masi, A. D. Bagdanov, and A. Del Bimbo, "Person re-identification by iterative re-weighted sparse ranking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1629–1642, Aug. 2015.
[97] S. Ding, L. Lin, G. Wang, and H. Chao, "Deep feature learning with relative distance comparison for person re-identification," *Pattern Recognit.*, vol. 48, no. 10, pp. 2993–3003, Oct. 2015.
[98] C. Benedek and T. Sziranyi, "Bayesian foreground and shadow detection in uncertain frame rate surveillance videos," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 608–621, Apr. 2008.
[99] C. Benedek and T. Szirányi, "Study on color space selection for detecting cast shadows in video surveillance," *Int. J. Imaging Syst. Technol.*, vol. 17, no. 3, pp. 190–201, 2007.
[100] M. Russell, J. J. Zou, G. Fang, and W. Cai, "Feature-based image patch classification for moving shadow detection," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.
[101] D. Mukherjee, Q. M. J. Wu, and T. M. Nguyen, "Gaussian mixture model with advanced distance measure based on support weights and histogram of gradients for background suppression," *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 1086–1096, May 2014.
[102] I. Kavasidis, S. Palazzo, R. Di Salvo, D. Giordano, and C. Spampinato, "An innovative web-based collaborative platform for video annotation," *Multimedia Tools Appl.*, vol. 70, no. 1, pp. 413–432, May 2014.
[103] C. Spampinato, S. Palazzo, and I. Kavasidis, "A texton-based kernel density estimation approach for background modeling under extreme conditions," *Comput. Vis. Image Understand.*, vol. 122, pp. 74–83, May 2014.
[104] Z. Zeng, J. Jia, D. Yu, Y. Chen, and Z. Zhu, "Pixel modeling using histograms based on fuzzy partitions for dynamic background subtraction," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 3, pp. 584–593, Jun. 2017.
[105] S. Brutzer, B. Höferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *Proc. CVPR*, Colorado Springs, CO, USA, Jun. 2011, pp. 1937–1944.
[106] S. Maity, A. Chakrabarti, and D. Bhattacharjee, "Block-based quantized histogram (BBQH) for efficient background modeling and foreground extraction in video," in *Proc. Int. Conf. Data Manage., Anal. Innov. (ICDMAI)*, Pune, India, Feb. 2017, pp. 224–229.
[107] J. D. Romero, M. J. Lado, and A. J. Méndez, "A background modeling and foreground detection algorithm using scaling coefficients defined with a color model called lightness-red-green-blue," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1243–1258, Mar. 2018.
[108] A. Vacavant, T. Chateau, A. Wilhelm, and L. Lequièvre, "A benchmark dataset for outdoor foreground/background extraction," in *Proc. Asian Conf. Comput. Vis.* Berlin, Germany: Springer, 2012, pp. 291–300.
[109] L. Maddalena and A. Petrosino, "The *3dSOBS+* algorithm for moving object detection," *Comput. Vis. Image Understand.*, vol. 122, pp. 65–73, May 2014.
[110] X. Liu, G. Zhao, J. Yao, and C. Qi, "Background subtraction based on low-rank and structured sparse decomposition," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2502–2514, Aug. 2015.
[111] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Changedetection.net: A new change detection benchmark dataset," in *Proc. CVPR Workshops*, Jun. 2012, pp. 1–8.
[112] M.-H. Yang, C.-R. Huang, W.-C. Liu, S.-Z. Lin, and K.-T. Chuang, "Binary descriptor based nonparametric background modeling for foreground extraction by using detection theory," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 4, pp. 595–608, Apr. 2015.
[113] S. Jiang and X. Lu, "WeSamBE: A weight-sample-based method for background subtraction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 9, pp. 2105–2115, Sep. 2018.
[114] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "CDnet 2014: An expanded change detection benchmark dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 393–400.
[115] K. Makantasis, A. Nikitakis, A. D. Doulamis, N. D. Doulamis, and I. Papaefstathiou, "Data-driven background subtraction algorithm for in-camera acceleration in thermal imagery," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 9, pp. 2090–2104, Sep. 2018.
[116] X. Zhao, Y. Chen, M. Tang, and J. Wang, "Joint background reconstruction and foreground segmentation via a two-stage convolutional neural network," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Hong Kong, Jul. 2017, pp. 343–348.
[117] L. A. Lim and H. Y. Keles. (Jan. 2018). "Foreground segmentation using a triplet convolutional neural network for multiscale feature encoding." [Online]. Available: https://arxiv.org/abs/1801.02225
[118] Y. Wang, Z. Luo, and P.-M. Jodoin, "Interactive deep learning method for segmenting moving objects," *Pattern Recognit. Lett.*, vol. 96, pp. 66–75, Sep. 2017.
[119] D. Zeng and M. Zhu, "Background subtraction using multiscale fully convolutional network," *IEEE Access*, vol. 6, pp. 16010–16021, 2018.
[120] M. Mandal, P. Saxena, S. K. Vipparthi, and S. Murala. (2018). "CANDID: Robust change dynamics and deterministic update policy for dynamic background subtraction." [Online]. Available: https://arxiv.org/abs/1804.07008
[121] T. Wang and Z. Zhu, "Real time moving vehicle detection and reconstruction for improving classification," in *Proc. IEEE Workshop Appl. Comput. Vis. (WACV)*, Breckenridge, CO, USA, Jan. 2012, pp. 497–502.
[122] A. Akula, R. Ghosh. S. Kumar, and H. K. Sardana, "Moving target detection in thermal infrared imagery using spatiotemporal information," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 30, no. 8, pp. 1492–1501, Aug. 2013.
[123] A. Akula, N. Khanna, R. Ghosh, S. Kumar, A. Das, and H. K. Sardana, "Adaptive contour-based statistical background subtraction method for moving target detection in infrared video sequences," *Infr. Phys. Technol.*, vol. 63, pp. 103–109, Mar. 2014.
[124] G. A. Bilodeau, A. Torabi, P.-L. St-Charles, and D. Riahi, "Thermal–visible registration of human silhouettes: A similarity measure performance evaluation," *Infr. Phys. Technol.*, vol. 64, pp. 79–86, May 2014.
[125] Z. Wu, N. Fuller, D. Theriault, and M. Betke, "A thermal infrared video benchmark for visual analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 201–208.
[126] *MULTIVISION Website*. Accessed: Jan. 1, 2019. [Online]. Available: http://atcproyectos.ugr.es/mvision/
[127] E. J. Fernandez-Sanchez, L. Rubio, J. Diaz, and E. Ros, "Background subtraction model based on color and depth cues," *Mach. Vis. Appl.*, vol. 25, no. 5, pp. 1211–1225, Jul. 2014.
[128] E. Fernandez-Sanchez, J. Diaz, and E. Ros, "Background subtraction based on color and depth using active sensors," *Sensors*, vol. 13, no. 7, pp. 8895–8915, Jul. 2013.
[129] M. I. Chacon-Murguia, H. E. Orozco-Rodríguez, and J. A. Ramirez-Quintana, "Self-adapting fuzzy model for dynamic object detection using RGB-D information," *IEEE Sensors J.*, vol. 17, no. 23, pp. 7961–7970, Dec. 2017.
[130] J. Stückler and S. Behnke, "Efficient dense 3D rigid-body motion segmentation in RGB-D video," in *Proc. BMVC*, 2013, pp. 1–11.
[131] B. Farou, M. N. Kouahla, H. Seridi, and H. Akdag, "Efficient local monitoring approach for the task of background subtraction," *Eng. Appl. Artif. Intell.*, vol. 64, pp. 1–12, Sep. 2017.
[132] M. Camplani and L. Salgado, "Background foreground segmentation with RGB-D Kinect data: An efficient combination of classifiers," *J. Vis. Commun. Image Represent.*, vol. 25, no. 1, pp. 122–136, Jan. 2014.
[133] L. Maddalena and A. Petrosino, "Towards benchmarking scene background initialization," in *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: Springer, 2015, pp. 469–476.
[134] M. De Gregorio and M. Giordano, "Background estimation by weightless neural networks," *Pattern Recognit. Lett.*, vol. 96, pp. 55–65, Sep. 2017.
[135] H.-C. Wang, Y.-C. Lai, W.-H. Cheng, C.-Y. Cheng, and K.-L. Hua, "Background extraction based on joint Gaussian conditional random fields," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 11, pp. 3127–3140, Nov. 2018. doi: 10.1109/TCSVT.2017.2733623.
[136] S. Javed, A. Mahmood, T. Bouwmans, and S. K. Jung, "Spatiotemporal low-rank modeling for complex scene background initialization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 6, pp. 1315–1329, Jun. 2018.
[137] S. Javed, T. Bouwmans, and S. K. Jung, "SBMI-LTD: Stationary background model initialization based on low-rank tensor decomposition," in *Proc. Symp. Appl. Comput.*, 2017, pp. 195–200.
[138] I. Kajo, N. Kamel, Y. Ruichek, and A. Malik, "SVD-based tensor-completion technique for background initialization," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 3114–3126, Jun. 2018.
[139] D. Avola, M. Bernardi, L. Cinque, G. L. Foresti, and C. Massaroni, "Adaptive bootstrapping management by keypoint clustering for background initialization," *Pattern Recognit. Lett.*, vol. 100, pp. 110–116, Dec. 2017.

[140] C. Cuevas, E. M. Yez, and N. García, "Labeled dataset for integral evaluation of moving object detection algorithms: LASIESTA" *Comput. Vis. Image Understand.*, vol. 152, pp. 103–117, Nov. 2016.
[141] Y. Chen, J. Wang, B. Zhu, M. Tang, and H. Lu, "Pixel-wise deep sequence learning for moving object detection," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.
[142] Y. Yao, P. Liu, X. Sun, and L. Zhang, "Moving object surveillance using object proposals and background prior prediction," *J. Vis. Commun. Image Represent.*, vol. 61, pp. 85–92, May 2019.
[143] S. Du and T. Ikenaga, "Local temporal coherence for object-aware keypoint selection in video sequences," in *Proc. Pacific Rim Conf. Multimedia*. Cham, Switzerland: Springer, 2017, pp. 539–549.
[144] C. Cuevas, R. Martinez, D. Berjón, and N. Garcia, "Detection of stationary foreground objects using multiple nonparametric background-foreground models on a finite state machine," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1127–1142, Mar. 2017.
[145] Z. Xu, B. Min, and R. C. C. Cheung. (2018). "A robust background initialization algorithm with superpixel motion detection." [Online]. Available: https://arxiv.org/abs/1805.06737
[146] B. Laugraud, S. Piérard, and M. Van Droogenbroeck, "LaBGen-P: A pixel-level stationary background generation method based on LaBGen," in *Proc. 23rd Int. Conf. Pattern Recognit.*, Dec. 2016, pp. 107–113.
[147] T. Minematsu, A. Shimada, and R. Taniguchi, "Background initialization based on bidirectional analysis and consensus voting," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Cancun, Mexico, Dec. 2016, pp. 126–131.
[148] M. Sultana, A. Mahmood, S. Javed, and S. K. Jung. (2018). "Unsupervised deep context prediction for background foreground separation." [Online]. Available: https://arxiv.org/abs/1805.07903
[149] D. Ortego, J. C. SanMiguel, and J. M. Martínez, "Rejection based multi-path reconstruction for background estimation in SBMnet 2016 dataset," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Cancun, Mexico, Dec. 2016, pp. 114–119.
[150] M. Camplani, L. Maddalena, G. M. Alcover, A. Petrosino, and L. Salgado, "A benchmarking framework for background subtraction in RGBD videos," in *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: Springer, 2017, pp. 219–229.
[151] T. Minematsu, A. Shimada, H. Uchiyama, and R.-I. Taniguchi, "Simple combination of appearance and depth for foreground segmentation," in *Proc. Int. Conf. Image Analysis Process.* Cham, Switzerland: Springer, 2017, pp. 266–277.
[152] G. Moyà-Alcover, A. Elgammal, A. Jaume-I-Capó, and J. Varona, "Modeling depth for nonparametric foreground segmentation using RGBD devices," *Pattern Recognit. Lett.*, vol. 96, pp. 76–85, Sep. 2017.
[153] L. Maddalena and A. Petrosino, "Background subtraction for moving object detection in RGBD data: A survey," *J. Imag.*, vol. 4, no. 5, p. 71, May 2018.
[154] R. Trabelsi, I. Jabri, F. Smach, and A. Bouallegue, "Efficient and fast multi-modal foreground-background segmentation using RGBD data," *Pattern Recognit. Lett.*, vol. 97, pp. 13–20, Oct. 2017.
[155] G. Yao, T. Lie, J. Zhong, P. Jiang, and W. Jia, "Comparative evaluation of background subtraction algorithms in remote scene videos captured by MWIR sensors," *Sensors*, vol. 17, no. 9, p. 1945, Aug. 2017.
[156] E. Yang, J. Gwak, and M. Jeon, "Multi-human tracking using part-based appearance modelling and grouping-based tracklet association for visual surveillance applications," *Multimedia Tools Appl.*, vol. 76, no. 5, pp. 6731–6754, Mar. 2017.
[157] R. B. Fisher, "The PETS04 surveillance ground-truth data sets," in *Proc. 6th IEEE Int. Workshop Perform. Eval. Tracking Surveill.*, May 2004, pp. 1–5.
[158] T. List, J. Bins, J. Vazquez, and R. B. Fisher, "Performance evaluating the evaluator," in *Proc. IEEE Int. Workshop Vis. Surveill. Perform. Eval. Tracking Surveill.*, Beijing, China, Oct. 2005, pp. 129–136.
[159] T. Bouwmans, C. Silva, C. Marghes, M. S. Zitouni, H. Bhaskar, and C. Frelicot, "On the role and the importance of features for background modeling and foreground detection," *Comput. Sci. Rev.*, vol. 28, pp. 26–91, May 2018.
[160] Y. Liu, G. Shi, Q. Cui, Y. Sheng, and G. Liu, "A method of personnel location based on monocular camera in complex Terrain," in *Proc. Chin. Conf. Biometric Recognit.* Cham, Switzerland: Springer, 2018, pp. 175–185.
[161] A. Costache, D. Popescu, C. Popa, and S. Mocanu, "Moving person detection using blob descriptors and support vector machine," in *Proc. 25th Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Maribor, Slovenia, Jun. 2018, pp. 1–5.
[162] T. Bouwmans, S. Javed, M. Sultana, and S. K. Jung. (2018). "Deep neural network concepts for background subtraction: A systematic review and comparative evaluation." [Online]. Available: https://arxiv.org/abs/1811.05255
[163] L. Wang and D. Sng. (2015). "Deep learning algorithms with applications to video analytics for a smart city: A survey." [Online]. Available: https://arxiv.org/abs/1512.03131
[164] D. Xie, L. Zhang, and L. Bai, "Deep learning in visual computing and signal processing," *Appl. Comput. Intell. Soft Comput.*, vol. 2017, Feb. 2017, Art. no. 1320780.
[165] M. Braham and M. Van Droogenbroeck, "Deep background subtraction with scene-specific convolutional neural networks," in *Proc. IEEE Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Bratislava, Slovakia, May 2016, pp. 1–4.
[166] K. Lim, W.-D. Jang, and C.-S. Kim, "Background subtraction using encoder-decoder structured convolutional neural network," in *Proc. 14th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug./Sep. 2017, pp. 1–6.
[167] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[168] R. Wang, F. Bunyak, G. Seetharaman, and K. Palaniappan, "Static and moving object detection using flux tensor with split gaussian models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 420–424.
[169] L. A. Lim and H. Y. Keles, "Foreground segmentation using convolutional neural networks for multiscale feature encoding," *Pattern Recognit. Lett.*, vol. 112, pp. 256–262, Sep. 2018.
[170] L. A. Lim and H. Y. Keles. (2018). "Learning multi-scale features for foreground segmentation." [Online]. Available: https://arxiv.org/abs/1808.01477
[171] W. Zheng, K. Wang, and F. Wang, "Background subtraction algorithm based on Bayesian generative adversarial networks," *Acta Autom. Sinica*, vol. 44, no. 5, pp. 878–890, 2018.
[172] W. Zheng, K. Wang, and F. Y. Wang, "A novel background subtraction algorithm based on parallel vision and Bayesian GANs," *Neurocomputing*, 2018.
[173] T. Bouwmans and B. Garcia-Garcia. (2019). "Background subtraction in real applications: Challenges, current models and future directions." [Online]. Available: https://arxiv.org/abs/1901.03577
[174] S. Li, D. Florencio, Y. Zhao, C. Cook, and W. Li, "Foreground detection in camouflaged scenes," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 4247–4251.
[175] S. Li, D. Florencio, W. Li, Y. Zhao, and C. Cook, "A fusion framework for camouflaged moving foreground detection in the wavelet domain," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3918–3930, Aug. 2018.
[176] C. Li, X. Wang, L. Zhang, J. Tang, H. Wu, and L. Lin, "Weighted low-rank decomposition for robust grayscale-thermal foreground detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 4, pp. 725–738, Apr. 2017.
[177] A. Zheng, T. Zou, Y. Zhao, B. Jiang, J. Tang, and C. Li, "Background subtraction with multi-scale structured low-rank and sparse factorization," *Neurocomputing*, vol. 328, pp. 113–121, Feb. 2019.
[178] A. Zheng, Y. Zhao, C. Li, J. Tang, and B. Luo, "Multispectral foreground detection via robust cross-modal low-rank decomposition," in *Proc. Pacific Rim Conf. Multimedia*. Cham, Switzerland: Springer, 2018, pp. 819–829.
[179] S. Yang, B. Luo, C. Li, G. Wang, and J. Tang, "Fast grayscale-thermal foreground detection with collaborative low-rank decomposition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 10, pp. 2574–2585, Oct. 2018.
[180] F. El Baf, T. Bouwmans, and B. Vachon, "Comparison of background subtraction methods for a multimedia application," in *Proc. 14th Int. Workshop Syst., Signals Image Process.*, Jun. 2007, pp. 385–388.
[181] *Underwater Change Detection Dataset*. Accessed: Mar. 12, 2019. [Online]. Available: http://underwaterchangedetection.eu/index.html
[182] D. D. Bloisi, L. Iocchi, A. Pennisi, and L. Tombolini, "ARGOS-venice boat classification," in *Proc. 12th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug. 2015, pp. 1–6.
[183] *MARDCT Dataset*. Accessed: Mar. 13, 2019. [Online]. Available: http://labrococo.dis.uniroma1.it/MAR/index.htm

[184] *CCT Dataset*. Accessed: Mar. 13, 2019. [Online]. Available: https://beerys.github.io/CaltechCameraTraps/
[185] S. Beery, G. Van Horn, and P. Perona, "Recognition in Terra incognita," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 472–489.
[186] X. P. Burgos-Artizzu, P. Dollár, D. Lin, D. J. Anderson, and P. Perona, "Social behavior recognition in continuous video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 1322–1329.
[187] S. Singh, S. A. Velastin, and H. Ragheb, "MuHAVi: A multicamera human action video dataset for the evaluation of action recognition methods," in *Proc. 7th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Boston, MA, USA, Aug./Sep. 2010, pp. 48–55.
[188] *MUHAVI-MAS Dataset*. Accessed: Mar. 15, 2019. [Online]. Available: http://velastin.dynu.com/MuHAVi-MAS/
[189] L. Sigal, A. O. Balan, and M. J. Black, "HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," *Int. J. Comput. Vis.*, vol. 87, nos. 1–2, pp. 4–27, 2010.
[190] S. Abdelhedi, A. Wali, and A. M. Alimi, "Toward a kindergarten video surveillance system (KVSS) using background subtraction based type-2 FGMM model," in *Proc. 6th Int. Conf. Soft Comput. Pattern Recognit. (SoCPaR)*, Tunis, Tunisia, Aug. 2014, pp. 440–446.
[191] S. Abdelhedi, A. Wali, and A. M. Alimi, "Fuzzy logic based human activity recognition in video surveillance applications," in *Proc. 2nd Int. Afro-Eur. Conf. Ind. Advancement (AECIA)*. Cham, Switzerland: Springer, 2016, pp. 227–235.
[192] *Edinburgh Ceilidh Overhead Video Data*. Accessed: Mar. 15, 2019. [Online]. Available: http://homepages.inf.ed.ac.uk/rbf/CEILIDHDATA/
[193] N. Avinash, M. S. S. Kumar, and S. M. Sagar, "Automated video surveillance for retail store statistics generation," in *Proc. 4th Int. Conf. Signal Image Process. (ICSIP)*. New Delhi, India: Springer, 2013, pp. 585–596.
[194] W. Huang, Q. Zeng, and M. Chen, "Motion characteristics estimation of animals in video surveillance," in *Proc. IEEE 2nd Adv. Inf. Technol., Electron. Automat. Control Conf. (IAEAC)*, Chongqing, China, Mar. 2017, pp. 1098–1102.
[195] E. Stergiopoulou, K. Sgouropoulos, N. Nikolaou, N. Papamarkos, and N. Mitianoudis, "Real time hand detection in a complex background," *Eng. Appl. Artif. Intell.*, vol. 35, pp. 54–70, Oct. 2014.
[196] H. Huang, X. Fang, Y. Ye, S. Zhang, and P. L. Rosin, "Practical automatic background substitution for live video," *Comput. Vis. Media*, vol. 3, no. 3, pp. 273–284, 2017.
[197] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[198] C. Lallier, E. Reynaud, L. Robinault, and L. Tougne, "A testing framework for background subtraction algorithms comparison in intrusion detection context," in *Proc. 8th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Klagenfurt, Austria, Aug./Sep. 2011, pp. 314–319.
[199] *OpenCV-Background Subtraction*. Accessed: Mar. 17, 2019. [Online]. Available: https://docs.opencv.org/3.3.0/db/d5c/tutorial_py_bg_subtraction.html
[200] D. H. Parks and S. S. Fels, "Evaluation of background subtraction algorithms with post-processing," in *Proc. IEEE 5th Int. Conf. Adv. Video Signal Based Surveill.*, Santa Fe, NM, USA, Sep. 2008, pp. 192–199.
[201] Y. Benezeth, D. Sidibé, and J. B. Thomas, "Background subtraction with multispectral video sequences," in *Proc. IEEE Int. Conf. Robot. Autom. Workshop Non-Classical Cameras, Camera Netw. Omnidirectional Vis. (OMNIVIS)*, HongKong, 2014, p. 6.
[202] *Scene*. Accessed: Mar. 17, 2019. [Online]. Available: http://scene.sourceforge.net/index.html
[203] A. Sobral, "BGSLibrary: An opencv c++ background subtraction library," in *Proc. IX Workshop de Visao Computacional*, 2013, vol. 2, no. 6, p. 7.
[204] A. Sobral and T. Bouwmans, "Bgs library: A library framework for algorithm's evaluation in foreground/background segmentation," in *Background Modeling and Foreground Detection for Video Surveillance*. Boca Raton, FL, USA: CRC Press, 2014, ch. 23.
[205] *BGSLibrary*. Accessed: Mar. 20, 2019. [Online]. Available: https://github.com/andrewssobral/bgslibrary
[206] A. Sobral, T. Bouwmans, and E. H. Zahzah, "Lrslibrary: Low-rank and sparse tools for background modeling and subtraction in video," in *Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*. Boca Raton, FL, USA: CRC Press, 2016.
[207] *LRSLibrary*. Accessed: Mar. 20, 2019. [Online]. Available: https://github.com/andrewssobral/lrslibrary
[208] L. Li, W. Huang, I. Y. H. Gu, and Q. Tian, "Foreground object detection from videos containing complex background," in *Proc. 11th ACM Int. Conf. Multimedia*, 2003, pp. 2–10.
[209] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "A self-adjusting approach to change detection based on background word consensus," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Waikoloa, HI, USA, Jan. 2015, pp. 990–997.
[210] L. Maddalena and A. Petrosino, "The SOBS algorithm: What are the limits?" in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Providence, RI, USA, Jun. 2012, pp. 21–26.
[211] A. Manzanera and J. C. Richefeu, "A new motion detection algorithm based on Σ–Δ background estimation," *Pattern Recognit. Lett.*, vol. 28, no. 3, pp. 320–328, 2007.
[212] S. Javed, A. Mahmood, T. Bouwmans, and S. K. Jung, "Background–foreground modeling based on spatiotemporal sparse subspace clustering," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5840–5854, Dec. 2017.
[213] X. Wang, L. Liu, G. Li, X. Dong, P. Zhao, and X. Feng, "Background subtraction on depth videos with convolutional neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Rio de Janeiro, Brazil, Jul. 2018, pp. 1–7.
[214] T. D. Bui, S. Ravi, and V. Ramavajjala, "Neural graph learning: Training neural networks using graphs," in *Proc. 11th ACM Int. Conf. Web Search Data Mining*, 2018, pp. 64–71.

RUDRIKA KALSOTRA received the bachelor's and master's degrees in computer applications from the University of Jammu, India, in 2013 and 2016, respectively. She is currently pursuing the Ph.D. degree in computer science with Shri Mata Vaishno Devi University, Katra, India. She is also doing her research work on moving object detection for automated video surveillance. She is also working on the background subtraction technique and deep learning algorithms. Her research interests include video analytics and image processing. She received the Gold Medal for academic excellence in bachelor's and master's degrees.

SAKSHI ARORA received the bachelor's degree in sciences and the master's degree in computer applications from the University of Jammu and the Ph.D. degree in computer science from Shri Mata Vaishno Devi University, Katra, India. She has over ten years of experience in teaching. She is currently an Assistant Professor with the School of Computer Science and Engineering, Shri Mata Vaishno Devi University. She has authored several research papers that are published at reputed journals and conferences. Her research interests include image processing, pattern recognition, and metaheuristic techniques for real-world optimization. She has been associated with various international conferences as a Paper Referee/ Program Technical Committee Member.