

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Pedestrian Detection and Tracking in Video Surveillance System: Issues, Comprehensive Review, and Challenges

Ujwalla Gawande, Kamal Hajari and Yogesh Golhar

Abstract

Pedestrian detection and monitoring in a surveillance system are critical for numerous utility areas which encompass unusual event detection, human gait, congestion or crowded vicinity evaluation, gender classification, fall detection in elderly humans, etc. Researchers' primary focus is to develop surveillance system that can work in a dynamic environment, but there are major issues and challenges involved in designing such systems. These challenges occur at three different levels of pedestrian detection, viz. video acquisition, human detection, and its tracking. The challenges in acquiring video are, viz. illumination variation, abrupt motion, complex background, shadows, object deformation, etc. Human detection and tracking challenges are varied poses, occlusion, crowd density area tracking, etc. These results in lower recognition rate. A brief summary of surveillance system along with comparisons of pedestrian detection and tracking technique in video surveillance is presented in this chapter. The publicly available pedestrian benchmark databases as well as the future research directions on pedestrian detection have also been discussed.

Keywords: pedestrian tracking, pedestrian detection, visual surveillance, pattern recognition, artificial intelligence

1. Introduction

The word surveillance, prefix sur is a French word means “over” and the root veiller means “to watch.” In distinction to surveillance, Steve Mann in [1] introduces the term “sousveillance.” Contrasting the word sur, sous meaning is “under,” i.e., it signifies that the camera is with human physically (ex. camera mounting on head). Surveillance and sousveillance both are used for continuous attentive observation of a suspect, prisoner, person, group, or ongoing behavior and activity in order to collect information. In order to improve conventional security systems, the use of surveillance system has been increasingly emboldened by government and private organizations. Currently, surveillance systems have been widely investigated and used effectively in several applications like (a) transport systems (railway stations, airports, urban and ruler motorway road networks), (b) government agencies (military base camps, prisons, strategic infrastructures, radar centers,

laboratories, and hospitals), (c) industrial environments, automated teller machine (ATM), banks, shopping malls, and public buildings, etc. The most of the surveillance systems at public and private places depend on the human operator observer, who detect any suspicious pedestrian activities in a video scene [2, 3]. The term pedestrian is a person who is walking or running on the street. In some communities, a person using wheelchair is also considered as pedestrians. The most challenging task for automatic video surveillance is to detect and track the suspicious pedestrian activity. For a real-time dynamic environment, the learning-based methods did not provide an appropriate solution for real-time scene analysis because it is difficult to obtain a prior knowledge about all the objects. Still, the learning-based methods are adopted due to their accuracy and robust nature. In the literature, several researchers use efficiently deep-learning (DL) based model for classification purpose in video surveillance over traditional approaches viz. perceptron model, probabilistic neural network (PNN), radial basis neural network (RBN), etc. Numerous learning-based techniques include artificial neural network (ANN), support vector machine (SVM), AdaBoost, etc. These techniques require the features such as histogram of oriented gradients (HOG), speeded-up robust features (SURF), local binary pattern (LBP), scale and invariant feature transform (SIFT), etc. to classify the type of object. Specifically, these features are represented by different deep learning algorithm versions such as the deep belief networks (DBN), recurrent neural network (RNN), generative adversarial networks (GANs), convolutional neural network (CNN), restricted Boltzmann machine (RBM), AlphaGo, AlphaZero, capsule networks bidirectional encoder representations for transformers (BERT), etc.

These variants of DL algorithms are used in many computer vision applications like face recognition, image classification, speech recognition, text-to-speech generation, handwriting transcription, machine translation, medical diagnosis, cars: drivable area, lane keeping, pedestrian and landmark detection for driver, digital assistants, ads, search, social recommendations, game playing, and content-based image retrieval. The advantage of DL approaches is its ability to learn complex scene features with very less processing of raw data and its capability of learning unlabeled raw data efficiently. Most recently, a new deep-learning technique called CNN have shown high performances over conventional methods in video processing research space. CNN can handle efficiently complex and large data.

During the past decade video surveillance systems have revolved from the simple video acquisition system to real-time intelligent autonomous systems.

Figure 1 shows a timeline chart of the evolution of video surveillance.

Visual surveillance systems come back into existence back in 1942. Primarily, closed-circuit television (CCTV) is used commercially as a security system, mainly for indoor environment. The main concerns of initial CCTVs were (1) voltage

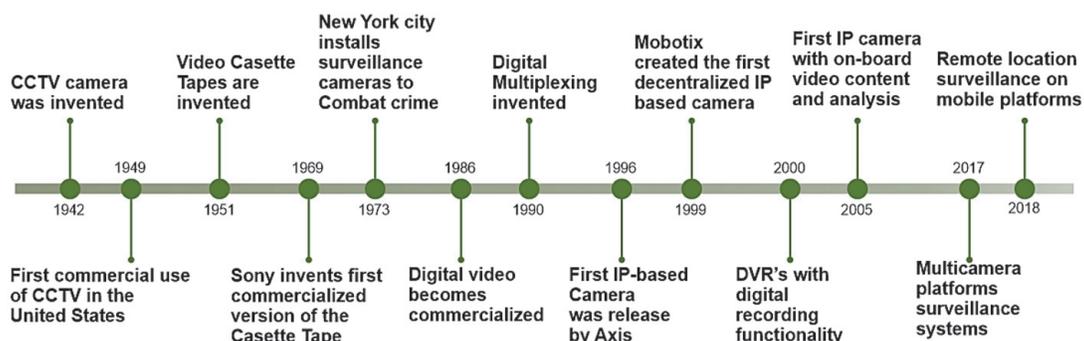


Figure 1.
Evolution of surveillance systems.

signals not openly transmitted in a distributed environment, (2) CCTV depends on strategic placements of cameras as per the geographical structure of workplace, (3) human observer is required for camera inputs to monitor the CCTV recorded footage [4]. The CCTV loses its primary advantage as an active, real-time medium, because the video footage can be used only after the fact or incident occurs, that can be used as a legal evidence or forensic tool. Next, in 1996, IP-based surveillance cameras were introduced by Axis, that overcomes the limitation of initial CCTV cameras such as (1) IP-based camera's transmits the raw images instead of voltage signals using the secure transmission channel of TCP/IP, (2) IP-camera comes along with the video analytics, i.e., camera itself can be used for analyzing the images, (3) Ethernet cable can be used as a medium for power supply instead of dedicated power supply, and (4) two-way bidirectional audio signals can be transmitted over a single dedicated network [5]. The recent surveillance system facilitates with remote location monitoring on handheld device like mobile phones.

The video surveillance systems can be categories based on a camera system, application and architecture. The camera system includes single camera, multi camera, fixed camera, moving camera and hybrid camera systems, etc. The application-based system includes object tracking and recognition, ID re-identification, customized event notification and alert based system, behavior analysis, etc. Finally, the architecture-based system includes standalone systems, cloud-based and distributed systems [6]. A general framework of automated visual surveillance system is shown in **Figure 2** [7–9]. Normally video surveillance system is based on multiple cameras, the videos from the multiple cameras are taken through the network and store in database.

The data need to be fused before incorporating the further processing. This can be done using data fusion techniques such as multi-sensory level, track to track and appearance to appearance [10–12]. After the data fusion following steps are performed. The traditional video surveillance system consists of various steps such

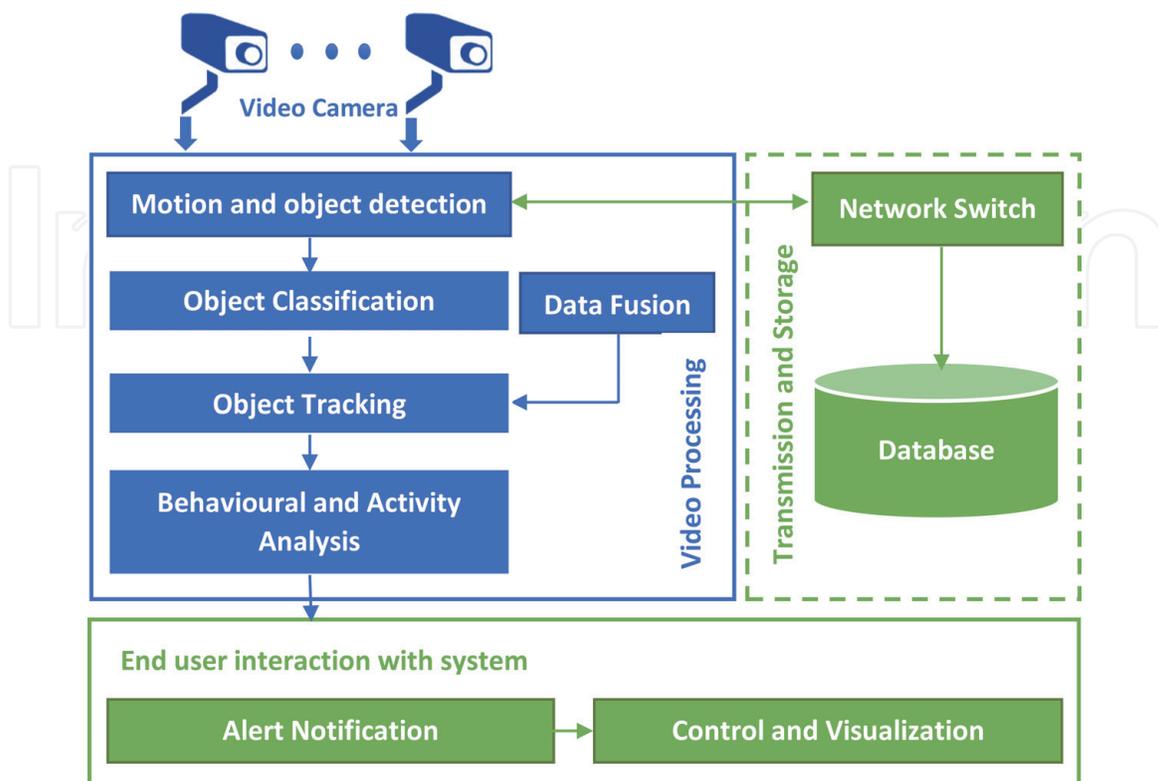


Figure 2.
A general framework of an automated visual surveillance system [7–9].

as (1) motion and object detection, (2) object classification, (3) object tracking, (4) behavior understanding and activity analysis, (5) pedestrian identification and (6) data fusion. Each stage of automated visual surveillance system is described as follows.

1.1 Motion and object detection

Object detection is the first step that deals with detecting instances of semantic objects of a certain class, such as humans, buildings, cars, etc. in a sequence of videos. The different approaches of object detection are frame-to-frame difference, background subtraction and motion analysis using optical flow techniques [13]. These approaches typically use extracted features and learning algorithms to recognize instances of an object category. The object detection process is divided into two categories. First, object detection, which include mainly three types of methods such as background subtraction, optical flow and spatiotemporal filtering. Second, object classification, use primarily visual features as shape based, motion based and texture-based method [13]. Motion detection is one of the problems in video surveillance, as it is not only responsible for the extraction of moving objects, but also critical to many applications including object-based video encoding, human motion analysis, and human machine interactions [14, 15].

After object detection, next step is motion segmentation. This step is used for detecting regions corresponding to moving objects such as humans or vehicles. It mainly focuses on detecting moving regions from video frames, and creating a database for tracking and behavior analysis. Motion detection is used for detecting a change in the position of an object, relative to its surroundings or a change in the surroundings, relative to an object. Motion detection can be achieved using electronic motion sensors, which detect the motion from the real environment.

1.2 Object tracking

Tracking of objects in a video sequence means identifying the same object in a sequence of frames using the object unique characteristics represented in the form of features. Generally, the detection process is always followed by tracking in video surveillance systems. Tracking is performed from one frame to another, using tracking algorithms such as kernel-based tracking, point based tracking and silhouette-based tracking [16].

1.3 Behavior and activity analysis

In some conditions, it is mandatory to analyze the behaviors of people and determine whether their behaviors are suspicious or not, such as the behavior of pedestrian at a crowded place (e.g. public market places and government offices, etc.). In this step the motion of objects is recognized from the video scene and generate the description of the action. Ahmed Elaiw et al. [80] proposed a critical analysis and modelling strategy of human crowds with the intention of selecting the most relevant scale out of three approaches, i.e., (1) microscopic, means pedestrian are individual detected based on the location, velocity and motion parameter is neglected, (2) mesoscopic, means pedestrian are detected based on position, velocity and depend on the distribution function and (3) macroscopic, mean the pedestrian are identified based on the average pedestrian quantity, moment of pedestrian. It can be used for efficient decision making in critical situations when

human crowd safety is important. Safety of human crowds depends upon the quantity and density of pedestrian move physically at different high crowded places.

1.4 Person identification

The last step is human identification. Human face and gait are the main biometric features that can be used for personal identification in visual surveillance systems after a behavior analysis [8].

The goal of this chapter is to discuss the issues and challenges involved in designing visual surveillance system. Again, group pedestrian detection and tracking methods used for moving and fixed camera into broad categories and give an informative analysis of relative methods in each category. The main contributions of this chapter are as follows:

- The comparative analysis of publicly available benchmark datasets of pedestrian with its use, specification and environment limitation
- Analyze issues and challenges of pedestrian detection and tracking in the video sequences captured by a moving and fixed camera
- Categorizing the methods of pedestrian detection and tracking in different ways based on the general concept of methods belonging to each category and described proposed improvements for each method

This chapter is organized into the following sections. Section 1 gives an introductory part, the importance of video surveillance system, recent advancement and general framework of video surveillance. Section 2, discusses different benchmark pedestrian datasets used to compare the different methods of pedestrian detection and tracking. Section 3, presents a detailed discussion on issues and challenges of pedestrian detection and tracking in video sequence. Section 4, groups the methods of pedestrian detection and tracking method for moving and fixed camera into different categories, describe their general concept with the improvements in each category. In Section 5, discusses possible future directions. Finally, the chapter concluded with a discussion in Section 6.

2. Pedestrian datasets reported in literature

The state-of-the-art methods for pedestrian detection and tracking method include adaptive local binary pattern (LBP), histogram of oriented gradient (HOG) into a multiple kernel tracker, spatiotemporal context information-based method using benchmark databases [10]. In this section we outlined the benchmark datasets that has been commonly used by the researchers. **Figure 3** shows a sample image of each pedestrian dataset. Next, we discuss each database with its specification, use and environmental constrain followed by comparative analysis.

2.1 Massachusetts Institute of Technology (MIT) pedestrian dataset

It is one of the first pedestrian datasets, fairly small and relatively well solved at this point. This data set contains 709 pedestrian images taken in city streets. Out of this 509 training and 200 test images of pedestrian in city scenes. Each image contains either a front or a back view with a relatively limited range of poses [11, 12].

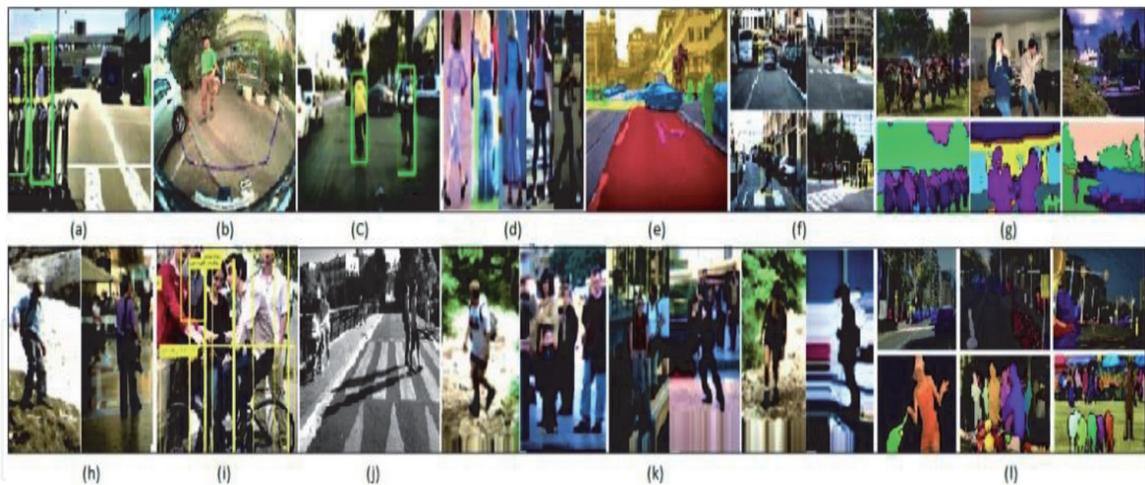


Figure 3.

Example of pedestrians dataset. (a) Caltech pedestrian dataset images consists of unique annotated pedestrians. (b) GM-ATCI rear-view pedestrians' dataset. (c) Tsinghua-Daimler Cyclist Detection Benchmark dataset images. (d) NICTA urban dataset. (e) ETH urban dataset. (f) TUD-Brussels dataset. (g) Microsoft COCO pedestrian dataset. (h) INRIA person static pedestrian detection datasets. (i) PASCAL object dataset. (j) CVC-ADAS collection of pedestrian datasets. (k) MIT pedestrian database images. (l) Mapillary vistas research dataset.

2.2 Caltech pedestrian dataset

The Caltech dataset consists of 640×480 resolution video taken from a vehicle driving through regular traffic in an urban environment. About 250,000 with a total of 350,000 bounding boxes and 2300 unique pedestrians were annotated for testing and training purpose. The annotation includes bounding boxes for each pedestrian walking on streets and detailed occlusion labels for each object captured in a video sequence in an urban environment. The annotation of pedestrians is used for validating the pedestrian detection and tracking algorithm accuracy [10].

2.3 General Motors-Advanced Technical Center (GM-ATCI) pedestrian dataset

GM-ATCI dataset is a rear-view pedestrians database captured using a vehicle-mounted standard automotive rear-view display camera for evaluating rear-view pedestrian detection. In total, the dataset contains 250 clips duration of 76 min and over 200K annotated pedestrian bounding boxes. The dataset has been captured at different locations, including: indoor and outdoor parking lots, city roads and private driveways. This dataset was collected in both day and night scenarios, with different weather and lighting conditions [15].

2.4 Daimler pedestrian dataset

The pedestrian images captured from a vehicle-mounted calibrated stereo camera rig in an urban environment. This dataset contains tracking information and a large number of labeled bounding box with a float disparity map and a ground truth shape image. The training set contains 15,560 pedestrian samples with 6744 label pedestrian and testing set contains more than 21,790 images with 56,492 pedestrian labels [15].

2.5 National Information and Communication Technology Australia (NICTA) pedestrian dataset

It is a large-scale urban dataset collected in multiple cities and countries. The dataset contains around 25,551 unique pedestrians of humans, allowing for a dataset

of over 50 K images with mirroring and annotation for validating detection and tracking algorithm accuracy [16].

2.6 Swiss Federal Institute of Technology (ETH) pedestrian dataset

It is an urban dataset captured from a stereo rig mounted on a stroller. Observing a traffic scene from inside a vehicle. The database is used for pedestrian detection and tracking from moving platforms in an urban scenario. Dataset consists of traffic agents such as different cars and pedestrians. One can predict their further motion, or even interpret their intention. At the same time, one needs to stay clear of any obstacles, remain on the assigned road, and read or interpret any traffic signs on the side of the street. On top that, a human is able to assess the situation, when close to a school or pedestrian crossing, one ideally will adapt one's driving behavior [17].

2.7 TUD-Brussels pedestrian dataset

This dataset consists of pairs recorded in a crowded urban setting from a moving platform with an onboard camera and challenging automotive safety scenario in urban environment [18].

2.8 National Institute for Research in Computer Science and Automation (INRIA) pedestrian dataset

INRIA is currently one of the most popular static pedestrian detection datasets. It contains moving people with significant variation in appearance, pose, clothing, background, illumination, coupled with moving cameras and backgrounds. Each pair shows two consecutive frames [19].

2.9 PASCAL visual object classes (VOC) 2007 and 2012 dataset

This is static object dataset with diverse object views and poses. The goal of visual object classes challenge is to recognize objects from a number of visual object classes in realistic scenes. The 20 object classes that have been selected are (1) person, (2) animal, (3) vehicle [20].

2.10 Microsoft Common Object in Context (COCO) 2018 dataset

The COCO is recent dataset created by Microsoft [22]. The dataset designed to spur object detection research with a focus on detecting objects in context. The annotations include different instances of segmentations for objects belonging to 80 categories of object, stuff segmentations for 91 categories, key point annotations for person instances, and five image label per image. The different COCO 2018 dataset challenges are (1) object detection with segmentation masks on the image, (2) panoptic segmentation, (3) person key point estimation, and (4) dense pose detection. **Figure 3(g)** shows the sample images of MS COCO dataset.

2.11 Mapillary vistas research dataset

The Mapillary vistas panoptic segmentation targets the full perception stack for scene segmentation in street-images [22]. Panoptic segmentation solves both stuff and thing classes, unifying the typically distinct semantic and instance segmentation tasks efficiently. **Figure 3(l)** shows a sample image of Mapillary vistas research datasets. The comparative analysis of recently utilized pedestrian database with its

Data source	Purpose	Image or video clips	Annotation	Environment	Ref.	Year
MIT	City street pedestrian segmentation, detection and tracking	709 pedestrian images 509 training and 200 test images	No annotated pedestrian	Day light scenario	[11] [12]	2000, 2005
Caltech pedestrian dataset	Detection and tracking of pedestrian walking on the street	250,000 frames (in 137 approximately minute long segments)	350,000 bounding boxes and 2300 unique pedestrians were annotated	Urban environment	[10]	2012
GM-ATCI	Rear view pedestrian segmentation, detection and tracking	250 video sequences	200K annotated pedestrian bounding boxes	Dataset was collected in both day and night scenarios, with different weather and lighting conditions	[13]	2015
Daimler	Detection and tracking of pedestrian	15,560 pedestrian samples, 6744 negative samples	2D bounding box overlap criterion and float disparity map and a ground truth shape image	Urban environment	[15]	2016
NICTA 2016	Segmentation, pose estimation, learning of pedestrian	25,551 unique pedestrians, 50,000 images	2D ground truth image	Urban environment	[16]	2016
MS COCO 2018	Object detection, segmentation, keypoint detection, DensePose detection	300,000, 2 million instances, 80 object categories	5 captions per image	Urban environment	[22]	2018
Mapillary vistas dataset 2017	Semantic understanding street scenes	25,000 images, 152 object categories	Pixel-accurate and instance-specific human annotations for understanding street scenes	Urban environment	[22]	2017
MS COCO 2017	Recognition, segmentation, captioning	328,124 images, 1.5 million object instances	Segmented people and objects	Urban environment	[22]	2017
MS COCO 2015	Recognition, segmentation, captioning	328,124 images, 80 object categories	Segmented people and objects	Urban environment	[22]	2015
ETH	Segmentation, detection, tracking	Videos	Dataset consist of other traffic agents such as different cars and pedestrians	Urban environment	[17]	2010
TUD-Brussels	Detection, tracking	1092 image pairs	1776 annotated pedestrian	Urban environment	[18]	2009

Data source	Purpose	Image or video clips	Annotation	Environment	Ref.	Year
INRIA	Detection, segmentation	498 images	Annotations are marked manually	Urban environment	[19]	2005
CVC-ADAS	Detection, tracking	60,000 frames	7900 annotated pedestrians	Urban environment	[20]	2009
PASCAL VOC 2012	Detection, classification, segmentation	11,530 images, 20 object classes	27,450 ROI annotated 6929 segmentations	Urban environment	[21]	2012

Table 1.
Recently used pedestrian databases by the researchers.

application for video surveillance system is shown in **Table 1**. The comparison is performed in terms of application of dataset, size of dataset, dataset creation environment scenarios and type of annotation details used for testing, training and validation of detection and tracking algorithm performance. These datasets used by the researchers for testing the performance of their respective pedestrian detection and tracking algorithm.

3. Issues and challenges of pedestrian detection and tracking

The moving object is a nonrigid thing that moves over time in image sequences of a video captured by a fix or moving the camera. In video surveillance system the region of interest is a human being that needs to be detected and tracked in the video [23]. However, this is not an easy task to do due to the many challenges and difficulties involved. These challenges occur at three different levels of pedestrian detection. Video acquisition, human detection and its tracking. The challenges in acquiring video are, viz. illumination variation, abrupt motion, complex background, shadows, object deformation, etc. Human detection and tracking challenges are varied poses, occlusion, crowd density area tracking, etc. Each issues and challenges are represented here in this section.

3.1 Problems related to camera

Many factors related to video acquisition systems, acquisition methods, compression techniques, stability of cameras (or sensors) can directly affect the quality of a video sequence. In some cases, the device used for video acquisition might cause limitation for designing object detection and tracking (e.g., when color information is unavailable, or when the frame rate is very low). Moreover, block artifacts (as a result of compression) and blur (as a result of camera's vibrations) reduce the quality of video sequences [36]. Noise is another factor that can severely deteriorate the quality of image sequences. Besides, different cameras have different sensors, lenses, resolutions and frame rates producing different image qualities. A low-quality image sequence can affect moving object detection algorithms. **Figure 4** shows an example of each challenge.

3.1.1 Camera motion

When dealing with detecting moving objects in the presence of moving cameras, the need for estimating and compensating the camera motion is evitable. However,

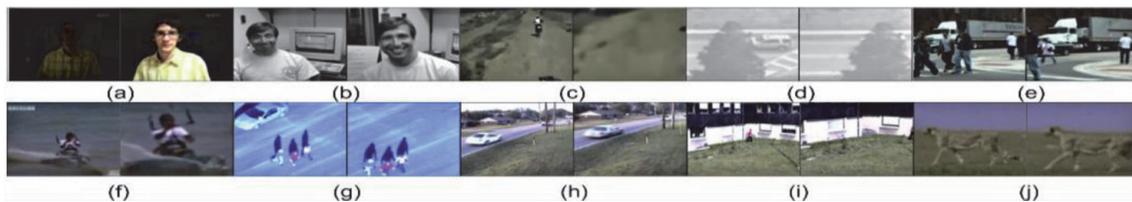


Figure 4.

Issues and challenges. (a) An example of illumination variation challenge (David indoor in the Ross dataset [39]). (b) An example of appearance change challenge (Dudek in the Ross dataset [39]). (c) An example of abrupt motion challenge (Motocross in the Kalal dataset [50]). (d) An example of occlusion challenge (car in the Kalal dataset [50]). (e) An example of freely motion of camera in the Michigan University dataset [10]. (f) An example of dynamic background challenge (Kitesurf in the Zhang dataset [60]). (g) An example of shadow challenge (pedestrian 4 in the Kalal dataset [50]). (h) An example of panning in camera in the CDNET database [10]. (i) An example of zooming in camera in the CDNET database [10]. (j) An example of nonrigid moving object in a video sequence [67].

it is not an easy task to do because of possible camera's depth changes and its complex movements. Many works elaborated an easy scenario by considering simple movements of the camera, i.e., pan tilt zoom (PTZ) cameras. This limited movement allows using a planar homography in order to compensate camera motions, which results in creating a mosaic (or a panorama) background for whole frames of the video sequence [37].

3.1.2 Nonrigid object deformation

In some cases, different parts of a moving object might have different movements in terms of speed and orientation. For instance, a walking dog when wags its tail or a moving tank when rotates its turret. When dealing with detecting such moving objects, most algorithms, different moving objects. It produces an enormous challenge, especially for nonrigid objects and in the presence of moving cameras. In Hou et al. [40], articular models have been proposed for moving nonrigid objects to handle nonrigid object deformation. In these models, each part of an articulated object is allowed to have different movements. It can be concluded that local features of a moving object along with updating background models are more efficient for dealing with this challenge.

3.2 Challenges in video acquisition

3.2.1 Illumination variation

The lighting conditions of the scene and the target might change due to the motion of light source, different times of day, reflection from bright surfaces, whether in-outdoor scenes, partial or complete blockage of the light source by other objects, etc. The direct impact of these variable results in background appearance changes, which causes false positive detections for the methods based on background modeling. Thus, it is essential for these methods to adapt their model to this illumination variation. Meanwhile, because the object's appearance changes under illumination variation, appearance-based tracking methods may not be able to track the object in the sequence [23–28]. Thus, it is required for these methods to use features which are invariant to illumination.

3.2.2 Presence of abrupt motion

Sudden changes in the speed and direction of the object's motion or sudden camera motion are another challenge of video acquisition that affects the object detection and tracking. If the object or the camera moves very slowly, the temporal

differencing methods may fail to detect the portions of the object coherent to background [31]. Meanwhile, a very fast motion produces a trail of the ghost detected region. So, if this object's motions or camera motions are not considered, the object cannot correctly be detected correctly by methods based on background modeling. On the other hand, for tracking-based methods, prediction of motion becomes hard or even impossible; as a result, the tracker might lose the target. Even if the tracker does not lose the target, the unpredictable motion can introduce a greater amount of error in algorithms [32].

3.2.3 Complex background

The background may be highly textured, especially in natural outdoor environments where high variability of textures is presented in outdoor scenes. Moreover, the background may be dynamic, like it may contain movement (e.g., a fountain, clouds in movement, traffic lights, trees waggle, water waves, etc.). These need to be considered as background in many moving object detection algorithms. Such movements can be periodic or nonperiodic [34].

3.2.4 Shadows

The presence of shadows in video image sequences complicates the task of moving object detection. Shadows are created due to the occlusion of the object by the light source. If the object does not move during the sequence, resulted shadow is considered as static and can effectively be incorporated into the background. However, a dynamic shadow, caused by a moving object, has a critical impact for accurately detecting moving objects, since it has the same motion properties as the moving object and is tightly connected to it. Shadows can be often removed from images of the sequence using their observed properties such as color, edges and texture or applying a model based on prior information such as illumination conditions and moving object shape [35, 47, 48]. However, dynamic shadows are still difficult to be distinguished from moving objects, especially for outdoor environment where the background is usually complex.

Next, human detection and tracking issues and challenges are discussed in brief. It includes varying poses, occlusion, crowd density area tracking, etc.

3.3 Challenges in human detection and tracking

3.3.1 Pedestrian occlusion

The object may be occluded by other objects in the scene. In this case, some parts of the object can be camouflaged or just hidden behind other objects (partial occlusion) or the object can be completely hidden by others (complete occlusion). As an example, consider the target to be a pedestrian walking on the sidewalk. It may be occluded by trees, cars in the street, other pedestrians, etc. Occlusion severely affects the detection of objects in background modeling methods, where the object is completely missing or separated into unconnected regions [33]. If occlusion occurs, the object's appearance model can change for a short time, which can cause some of the object tracking methods.

3.3.2 Pose variation: moving object appearance changes

In real scenarios, most objects can occur in 3D space, but we have the projection of their 3D movement in a 2D plane. Hence, any rotation in the direction of third

Challenge	Proposed methodology	Advantage	Identifies gap	Observation	Ref.	Year
Problems related to camera	1. Heterogeneous feature-based technique	1. Detecting human movements in low resolution conditions	1. Low-resolution humans' images may not cover all targets with different sizes	1. Region of Interest method can be used in outdoor environment	[53]	2017
	2. Color-based appearance model and the APSO based framework	2. Deal with low resolution video sequences, the techniques based on fusion can be used to better detect moving objects	2. Human detection accuracy is less in complex background	2. Appearance model, the object is significantly fast	[52]	2015
	3. Pyramidal structure model	3. Fast detect moving objects	3. Not tested for real time scenarios	3. Detect the moving object with detection accuracy are 80%	[51]	2014
	4. Blur-driven tracker (BLUT) framework	4. Can robustly track severely blurred targets	4. Misclassification in detection and tracking in complex background	4. Effectively tracks the blurred objects without deblurring	[50]	2011
	5. Cascade particle filter in tracking	5. Detection for low frame rate videos	5. Tracker not able to distinguish between different targets of video sequence	5. Efficient multi-target tracker is required	[49]	2008
Camera motion	1. Image stabilization techniques based on motion compensation of features	1. Efficiently detect moving object and motion blur issue resolve	1. Motion detection fails in long video. It computationally complex	1. Adaptive particle filter framework uses PCA to reduce SIFT feature	[63]	2014
	2. 3D motion models	2. Can be efficiently used to compensate camera vibrations	2. It cannot be applied to real-time systems due to the slowness of the SIFT computation	2. 3D camera motion model requires fast feature extraction	[62]	2011
	3. Multi plane representation of the 3D scene	3. Computational complexity is decreased	3. It is computationally very complex	3. It can be used for image analysis	[61]	2009
	4. Adaptive motion model	4. More accurately detects moving objects	4. Camera motion detection time is more	4. Efficiently resolve the issue of motion blur	[60]	2009
Nonrigid object deformation	1. Target object as a combination of different segments having different movements	1. Segments with motion consistency	1. Computationally intensive	1. Eigen tracker is used to track and recognize gestures	[59]	2017
	2. Model of articulated objects composed of rigid bodies	2. Accurately produce a depth image representing different poses of a moving object	2. Dense articulated real-time tracker requires initial object pose. Tracker not able to measure the similarity of the object accurately for small object	2. Accuracy of tracker can be improved by more color features	[58]	2005
	3. View-based eigenspace representation	3. Produced good results for handling tracking		3. Tracker not able to detect the small objects	[57]	1994
				[56]	2017	
				[55]	2000	
				[54]	1998	
				[66]	2017	
				[65]	2015	
				[64]	1998	

Challenge	Proposed methodology	Advantage	Identifies gap	Observation	Ref.	Year
Illumination variation	1. Conditional background update scheme	1. Evaluate rapidly scene changes adaptively	1. Computation time required for this model is more	1. Results: TPF—7.36 ms and Speed FPS—136	[23]	2017
	2. Local representation model	2. More efficient to detect moving objects	2. Final segmentation results consist of large noise patches	2. Post processing can reduce large noise patches	[24]	2015
	3. 2D-Cepstrum approach	3. Provided good robustness to illumination variations	3. This method is computationally intensive	3. Color and texture information retain	[25]	2010
	4. Adaptive local binary pattern	4. Tolerant to illumination variations	4. This method requires a nonmoving camera	4. Texture features having FPS—15 fps	[26]	2006
	5. Bayesian framework	5. Not sensitive to illumination variation	5. If the colors changes are very fast then, this method fails to detect object	5. Dynamic texture coefficient is used for texture variations with 5 fps	[27]	2006
	6. Color modelling approach with B-spline curves	6. Adapt to irregular illumination variations and abrupt changes of brightness	6. Speed of visual tracker is less. Visual tracker can detect only single target object	6. Visual tracker speed improved using good workstation	[28]	2001
Presence of abrupt motion	1. Kernelized correlation filter tracker based on swarm intelligence method	1. Effectively handle abrupt motion tracking in videos	1. Kernelized correlation filter tracker computational complexity is more	1. A unified framework track smooth or abrupt motion	[31]	2017
	2. Hamiltonian Markov Chain Monte Carlo (MCMC) based tracking algorithm	2. Effective in handling different type of abrupt motions	2. Tracker does not handle abrupt motion in scale and position	2. Tracker can handle smooth and abrupt motion	[30]	2012
	3. Bayesian filter tracking framework	3. Effective against abrupt motions	3. Tracker not suitable in significant motion	3. Bayesian filtering solve the local-trap problem	[29]	2012
	4. Wang-Landau Monte Carlo sampling method	4. Efficiently deals with the abrupt motions	4. Tracker does not fully consider the abrupt changes	4. Tracking algorithm efficiently handle abrupt motions	[28]	2008
Complex background	1. Adaptive background model	1. Highly complex backgrounds for detecting moving objects	1. Less accurate in background regions with large re-projection errors	1. Adaptive to any object detection methods	[42]	2017
	2. Dynamic texture modelling methods	2. Effectively used for moving object detection	2. It is computationally very complex	2. Operates on dynamic texture sequence	[41]	2017
	3. Principal features based on statistical characteristics	3. Effectively overcome the complex background	3. Wrongly absorb a foreground object with background	3. Fusion of information solved the issue	[40]	2005
	4. Auto-regressive model		4. The model is very complex		[39]	2004
					[38]	2003
					[37]	1995

Challenge	Proposed methodology	Advantage	Identifies gap	Observation	Ref.	Year
	5. Active contour model	<ol style="list-style-type: none"> Effectively handle dynamic background Can detect active segments 	5. Output with noise patches	<ol style="list-style-type: none"> Filters operates with linear prediction model It detects complex background object with noise patches 		
Shadow	<ol style="list-style-type: none"> Shadow elimination algorithm using HSV color space and texture features Modified Gaussian mixture model 	<ol style="list-style-type: none"> Can effectively distinguish shadow regions Can handle a highly dynamic environment object detection for moving camera 	<ol style="list-style-type: none"> Long video accuracy reduces due to texture variation Misclassification in shadow detection result if complex scene 	<ol style="list-style-type: none"> Discrimination in image get difficult due to texture variation This algorithm works in real time with good accuracies 	[44] [43]	2017 2016
Pedestrian occlusion	<ol style="list-style-type: none"> Histogram of oriented gradient (HOG) into a multiple kernel tracker Spatiotemporal context information-based method Maintaining appearance models Tracking was achieved by evolving the contour from frame to frame Appearance model based on the filter responses from a steerable pyramid 	<ol style="list-style-type: none"> Effectively handle occlusions in different conditions of moving cameras Tracker is able to distinguish the object in occlusions effectively Occlusions can be more efficiently handled Minimize some energy function Overcome changing appearance and occlusion problems 	<ol style="list-style-type: none"> Histogram of oriented gradient (HOG) into a multiple kernel tracker Spatiotemporal context information method is complex Occlusion fail for long sequences with varying lighting conditions Offline object tracking not possible for all types of objects Tracked object move with its background. Tracker, fail when object is occluded 	<ol style="list-style-type: none"> Effectively handle occlusions in moving cameras Spatiotemporal context to analyze occlusion Tracker deals with complex real scenario Nonrigid object tracking derived by Bayesian framework Motion-based tracker with RMS error rate is 0.1–1.1 	[36] [35] [34] [33] [32]	2017 2007 2006 2004 2003
Pose variation—moving object appearance changes	<ol style="list-style-type: none"> Trainable model which uses the optical flow Wandering-stable-lost framework model 	<ol style="list-style-type: none"> Good performance in handling appearance changes Adaptive and significant to appearance changes 	<ol style="list-style-type: none"> Difficult to identify ambiguous motion pattern of object Sensitive to lighting changes. Multiple cameras required to cope with self-occlusion 	1. CNN, learn with synthetic video with Mean—69.7	[27] [26] [25] [24]	2017 2006 2006 2004

Challenge	Proposed methodology	Advantage	Identifies gap	Observation	Ref. Year
	3. Covariance matrix and Lie algebra 4. Low-dimensional subspace representation	3. Adaptively track moving objects under their appearance changes for moving camera 4. Efficiently adapts online to changes in the appearance of the moving objects	3. Covariance tracker is computationally intensive 4. Tracker occasionally drift from a target object	2. Models appearance using a mixture with 180 angles 3. Tracker handle illumination changes 4. Robust object tracking. RMS error is 5.07 pixels per frame	

Table 2.
Challenges of pedestrian detection and tracking with related reference works.

axis may change the object appearance [29]. Tracking algorithm performance gets affected due to variation in pose. Same pedestrian looks different in consecutive frames, if the pose changes continuously. Moreover, the objects themselves may have some changes in their pose and appearance like facial expressions, changing clothes, wearing a hat, etc. Also, the target can be a nonrigid object, where its appearance may change over time. In many applications, the goal is tracking humans or pedestrians, which makes tracking algorithms vulnerable in this challenging case [30]. **Table 2** summarizes the comparative analysis of methodologies with its advantage, identified gaps and observation for handling these challenging issues in a video surveillance system.

4. Pedestrian detection and tracking

In video-based surveillance, one of the key tasks is to detect the presence of pedestrians in a video sequence, i.e., localizing all subjects that are human [45, 68]. This problem corresponds to determining regions, typically the smallest rectangular bounding boxes in the video sequence that enclose humans. In most of the surveillance systems, human behavior has been recognized using analysis of the trajectories, positions of persons and historical or prior knowledge about the scene. **Figure 5** shows some examples of pedestrian detection and tracking. Haritaoglu et al. [46] describe a combined approach of shape analysis and body tracking, and model different appearances of a person. This has been designed for outdoor environment using a single camera. The system detects and tracks groups of people and monitors the behaviors, even in the presence of partial occlusion. However, the performance is mainly based on the detected trajectories of the concerned objects in video. Furthermore, the results are not sufficient for semantic recognition of dynamic human activities and event analysis in some cases. The advanced automatic video surveillance system consists of many features such as, motion detection [69, 70],



Figure 5.

Example of pedestrians detection and tracking. (a) Detecting pedestrians outdoors, walking along the street. (b) ADAS pedestrian detection. (c) Pedestrian detector is based on the aggregate channel feature detector. (d) Real-time vehicle and pedestrian detection of road scenes. (e) Pedestrian action prediction is based on the analysis of human postures in context of traffic. (f) Pedestrian detection based on hierarchical co-occurrence model (g) Cross-modal deep representations for robust pedestrian detection. (h) Pedestrian detection OpenCV. (i) Object tracking with dlib C++ library. (j) Multiple object tracking with Kalman tracker. (k) Multi-Class Multi-Object Tracking using Changing Point Detection. (l) Pedestrian tracking using Deep-Occlusion Reasoning method.

human behavior analysis, detection and tracking [71–73]. Human tracking is quite challenging, since humans may vary in intra-class variability in shape, appearance due to different viewing perspectives and other visual properties.

Krahnstoeber et al. [75] designed a real-time control system of active cameras for a multiple-camera surveillance system. Hence, various researchers shifted focus from static fixed camera-based pedestrian detection to moving dynamic multi-camera-based pedestrian detection. Pedestrian tracking has been done by stationary cameras using a shape-based method [76], which detects and compares the human-body shape in consecutive frames. The cameras have been calibrated using a common site-wide metric coordinate system described in [77, 78]. Funahasahi et al. [73] developed a system for tracking the human head and face parts by means of a hierarchical tracking method using a stationary camera and a PTZ camera. The recent surveillance system focuses on human tracking by detection as described in [72–75]. Andriluka et al. [76–78] combined the initial estimate of the human pose across frames in a tracking-by-detection framework. Sapp et al. [79] coupled locations of body joints within and across frames from an ensemble of tractable sub-models. Wu and Nevatia [80] proposed an approach for detection and tracking of partially occluded people using an assembly of body parts.

The tracking of humans becomes more challenging under moving cameras than in static cameras as discussed in Section 2. Many effective pedestrian tracking techniques used in static camera, such as background subtraction and modeling [80] and a constant ground plane assumption, makes the task more difficult. Instead of using background modeling-based methods to extract the human information, human detectors are widely used to detect the human in the video. Therefore, the challenge is to successfully detect the humans in moving cameras, and then apply the tracking techniques to detected humans. However, human detectors may effectively extract human, still have some limitations viz. human detectors may produce false or miss human detection, when humans are partially or fully occluded, the detections can fail and the tracking can be unreliable until the human reappear in the frames. It is observed that, many of the researcher works on many of challenges of pedestrian detection and tracking, but still complete and reliable solution to all the challenges like discussed. Most of the algorithms of pedestrian detection and tracking were tested in indoor and outdoor environment. Attempts were also made to estimate the accuracy of the system based on detection rate, time and computational complexity. From the performance evaluation of algorithms presented in authors, it is observed that, deep learning based pedestrian detection and tracking approaches can be efficient choice for real-time environment [45, 65]. There is still a scope of improvement in existing approaches of pedestrian detection and its tracking in surveillance system.

5. Conclusions

This chapter describes and reviews the methodologies, strategies and steps involved in video surveillance. It also addresses the challenges, issue, available databases, available solutions and research trends for human detection and tracking in video surveillance system. Based on the literature survey, most of the available techniques proposed by the earlier researchers can perform object detection and tracking either within single camera view or across multiple cameras. However, most of them failed to encounter trade-off problem between accuracy and speed. Although the accuracy of the trackers is very good, they are often impractical because of their high computational requirements and vice versa. Thus, to achieve an optimal trade-off, adaptive object detection and tracking method, it is essential

to achieve a real-time and reliable surveillance system. It is due to this reason that the main aim of this paper is to provide a valuable insight into the related areas of the related research topic in video surveillance and to promote new research.

IntechOpen

Author details

Ujwalla Gawande^{1*}, Kamal Hajari¹ and Yogesh Golhar²

1 Department of Information Technology, Yeshwantrao Chavan College of Engineering, Nagpur, Maharashtra, India

2 Department of Computer Science and Engineering, GH Rasoni Institute of Engineering and Technology, Nagpur, Maharashtra, India

*Address all correspondence to: ujwallgawande@yahoo.co.in

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Minsky M, Kurzweil R, Mann S. IEEE International Symposium on Technology and Society, Toronto, Ontario, Canada, 27th–29th June; The Society of Intelligent Veillance. 2013. pp. 13-17
- [2] Foresti GL, Micheloni C, Snidaro L, Remagnino P, Ellis T. Active video-based surveillance system: The low-level image and video processing techniques needed for implementation. In: IEEE Signal Processing Magazine, Vol. 22(2). March 2005. pp. 25-37
- [3] Gawande U, Golhar Y. Biometric security system: A rigorous review of unimodal and multimodal biometrics techniques. *International Journal of Biometrics (IJBM)*. April 2018;**10**(2)
- [4] Cornett B. Intro to Surveillance Camera Technologies. Available at: <http://www.ezwatch.com>
- [5] Alexandr L. IP Video Surveillance. An Essential Guide. Alexandr Lytkin. ISBN 978-5-600-00033-9. 4th April 2012
- [6] Zafeiriou S, Zhang C, Zhang Z. A survey on face detection in the wild: Past, present and future. *International Journal of Computer Vision Image Understand*. Sep. 2015;**138**:1-24
- [7] Teddy K, Lin W. A survey on behavior analysis in video surveillance applications. *Video Surveillance*. 2011. pp. 281-291. Available at: <http://www.intechopen.com/books/videosurveillance>
- [8] Gawande U, Golhar Y, Hajari K. Biometric-based security system: Issues and challenges, intelligent techniques in signal processing for multimedia security. In: *Studies in Computational Intelligence*. Vol. 660. Cham: Springer; 2017. pp. 151-176
- [9] Gawande U, Zaveri M, Kapur A. A novel algorithm for feature level fusion using SVM classifier for multibiometrics-based person identification. *Applied Computational Intelligence and Soft Computing*. July 2013;**2013**:1-11
- [10] Dollar P, Wojek C, Schiele B, Perona P. Pedestrian detection: An evaluation of the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. April 2012;**34**(4): 743-761
- [11] Papageorgiou C, Poggio T. A trainable system for object detection. *International Journal of Computer Vision*. June 2000;**38**(1):15-33
- [12] MIT Pedestrian Dataset. Center for Biological and Computational Learning at MIT and MIT. 2005. Available at: <http://cbcl.mit.edu/software-datasets/PedestrianData.html> [Accessed: 22 September 2018]
- [13] Levi D, Silberstei S. Tracking and motion cues for rear-view pedestrian detection. In: *18th IEEE Intelligent Transportation Systems Conference, Spain, 15th–16th Sept. 2015*. pp. 664-671
- [14] Li X, Flohr F, Yang Y, Xiong H, Braun M, Pan S. A new benchmark for vision-based cyclist detection. In: *IEEE Intelligent Vehicles Symposium, Sweden, 19th–22nd. June 2016*. pp. 1028-1033
- [15] Campbell D, Petersson L. GOGMA: Globally-optimal Gaussian mixture alignment. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, IEEE. 2016*
- [16] Pellegrini S, Ess A, Van Gool L. Wrong turn–No dead end: A stochastic pedestrian motion model. *International*

Workshop on Socially Intelligent Surveillance and Monitoring (SISM'10), (CVPR), San Francisco, CA, USA, 13th–18th June. 2010

[17] Christian W, Stefan W, Schiele B. Multi-cue onboard pedestrian detection. In: IEEE Conference on Computer Vision and Pattern Recognition, Miami, Florida, USA, 20–25 June. 2009

[18] Navneet D. Finding people in images and videos [PhD thesis]. Inria Grenoble-Rhône-Alpes; 2006

[19] Everingham M, Van Gool L, Winn WC, Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*. June 2015;111(1):98-136

[20] CVC-ADAS Pedestrian dataset. 2012. Available at: <http://adas.cvc.uab.es/site/> [Accessed: 22 September 2018]

[21] Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft coco: Common objects in context. In: *European Conference on Computer Vision*. Springer; 2014, 2015. pp. 740-755

[22] Yun K, Lim J, Choi JY. Scene conditional background update for moving object detection in a moving camera. *Pattern Recognition Letters*. 2017;88:57-63

[23] St-Charles PL, Bilodeau GA, Bergevin R. Subsense: A universal change detection method with local adaptive sensitivity. *IEEE Transactions on Image Processing*. 2015;24(1):359-373

[24] Cogun F, Cetin AE. Object tracking under illumination variations using 2D spectrum characteristics of the target. *IEEE International Workshop on Multimedia Signal Processing*. 2010: 521-526

[25] Heikkila M, Pietikainen M. A texture-based method for modelling the

background and detecting moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2006;28(4):657-662

[26] Shen C, Lin X, Shi Y. Moving object tracking under varying illumination conditions. *Pattern Recognition Letters*. 2006;27(14):1632-1643

[27] Lee YB. A real-time color-based object tracking robust to irregular illumination variations. In: *IEEE International Conference on Robotics and Automation*, 21st–26th. May 2001. pp. 1659-1664

[28] Tokmakov P, Alahari K, Schmid C. Learning motion patterns in videos. In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*. 2017. pp. 531-539

[29] Balan A, Black MJ. An adaptive appearance model approach for model based articulated object tracking. In: *IEEE Conference on Computer Vision and Pattern Recognition*. June 2006. pp. 758-765

[30] Porikli F, Tuzel O. Covariance tracking using model update based on lie algebra. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR' 1*. June 2006. pp. 728-735

[31] Lim J, Ross DA, Lin RS, Yang MH. Incremental learning for visual tracking. *Advances in Neural Information Processing Systems*. 2004:793-800

[32] Kwo J, Lee KM. Tracking of abrupt motion using Wang-Landau Monte Carlo estimation. In: *European Conference on Computer Vision*. Oct. 2008. pp. 387-400

[33] Zhou X, Lu Y, Lu J, Zhou J. Abrupt motion tracking via intensively adaptive Markov-chain Monte Carlo sampling. *IEEE Transactions on Image Processing*. 2012;21(2):789-801

- [34] Wang F, Lu M. Hamiltonian Monte Carlo estimator for abrupt motion tracking. In: International Conference on Pattern Recognition, ICPR. Nov. 2012. pp. 3066-3069
- [35] Zhang H, Zhang J, Wu Q, Qian X, Zhou T, Hengcheng FU. Extended kernel correlation filter for abrupt motion tracking. *KSII Transactions on Internet and Information Systems*. 2017; **11**(9):4438-4460
- [36] Jepson AD, Fleet DJ, El-Maraghi TF. Robust online appearance models for visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2003; **25**(10):1296-1311
- [37] Yilmaz A, Li X, Shah M. Contour-based object tracking with occlusion handling in video acquired using mobile cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2004; **26**(11):1531-1536
- [38] Senior A, Hampapur A, Tian YL, Brown L, Pankanti S, Bolle R. Appearance models for occlusion handling. *Image and Vision Computing*. 2006; **24**(11):1233-1243
- [39] Pan J, Hu B. Robust occlusion handling in object tracking. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2007. pp. 1-8
- [40] Hou L, Wan W, Lee KH, Hwang JN, Okopal G, Pitton J. Robust human tracking based on DPM constrained multiple-kernel from a moving camera. *Journal of Signal Processing Systems*. 2017; **86**(1):27-39
- [41] Delagnes P, Benois J, Barba D. Active contours approach to object tracking in image sequences with complex background. *Pattern Recognition Letters*. 1995; **16**(2):171-178
- [42] Monnet A, Mittal A, Paragios N, Ramesh V. Background modelling and subtraction of dynamic scenes. In: *Ninth IEEE International Conference on Computer Vision*. 2003. pp. 1305-1312
- [43] Li L, Huang W, Gu IY, Tian Q. Statistical modelling of complex backgrounds for foreground object detection. *IEEE Transactions on Image Processing*. 2004; **13**(11):1459-1472
- [44] Chetverikov D, Péteri R. A brief survey of dynamic texture description and recognition. *Computer Recognition Systems*. 2005:17-26
- [45] Arashloo SR, Amirani MC, Noroozi A. Dynamic texture representation using a deep multi-scale convolutional network. *Journal of Visual Communication and Image Representation*. 2017; **43**:89-97
- [46] Minematsu T, Uchiyama H, Shimada A, Nagahara H, Taniguchi RI. Adaptive background model registration for moving cameras. *Pattern Recognition Letters*. 2017:86-95
- [47] Xia H, Shuxiang S, Liping H. A modified Gaussian mixture background model via spatiotemporal distribution with shadow detection. *Signal, Image and Video Processing*. 2016; **10**(2): 343-350
- [48] Song R, Liu M. A shadow elimination algorithm based on HSV spatial feature and texture feature. In: *International Conference on Emerging Internetworking, Data Web Technologies*. 2017. pp. 585-591
- [49] Treptow A, Zell A. Real-time object tracking for soccer-robots without color information. *Robotics and Autonomous Systems*. 2004; **48**(1):41-48
- [50] Hua C. A noise-insensitive object tracking algorithm. In: *Conference on Computer Vision*. Nov. 2007
- [51] Unger M, Asdsch M, Hosten P. Enhanced background subtraction using global motion compensation and

- mosaicing. In: International Conference on Image Processing, ICIP. Oct. 2008. pp. 2708-2711
- [52] Li Y, Ai H, Yamashita T. Tracking in low frame rate video: A cascade particle filter with discriminative observers of different life spans. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2008;**30**(10):1728-1740
- [53] Spampinato C, Chen-Burger YH, Nadarajan G, Fisher RB. Detecting, tracking and counting fish in low quality unconstrained underwater videos. *VISAPP*. 2008;**2**:514-519
- [54] Wu Y, Ling H, Yu J, Li F, Mei X, Cheng E. Blurred target tracking by blur-driven tracker. In: International Conference on Computer Vision. Nov. 2011. pp. 1100-1107
- [55] Dollár P, Appel R, Blondie S, Perona P. Fast feature pyramids for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2014;**36**(8):1532-1545
- [56] Zhang X, Hu W, Xie N, Bao H, Maybank S. A robust tracking system for low frame rate video. *International Journal of Computer Vision*. 2015; **115**(3):279-304
- [57] Chen HK, Zhao XG. heterogeneous features fusion-based low-resolution human detection method for outdoor video surveillance. *International Journal of Automation and Computing*. 2017; **14**(2):136-146
- [58] Irani M, Anandan P. A unified approach to moving object detection in 2d and 3d scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1998;**20**(6):577-589
- [59] Sawhney HS, Guo Y, Kumar R. Independent motion detection in 3d scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2000;**22**(10):1191-1199
- [60] Zhou D, Frémont V, Quost B, Dai Y, Li H. Moving object detection and segmentation in urban environments from a moving platform. *Image and Vision Computing*. 2017;**68**:76-87
- [61] Available from: <http://www.cvpapers.com/datasets.html>
- [62] Wang J. Representing moving images with layers. *IEEE Transactions on Image Processing*. 1994;**3**(5):625-638
- [63] Xiao J, Shah M. Motion layer extraction in the presence of occlusion using graph cut. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2005;**27**(10):1644-1659
- [64] Chen T, Lu S. Object-level motion detection from moving cameras. *IEEE Transactions on Circuits and Systems for Video Technology*. 2017;**27**(11): 2333-2343
- [65] Hayman E, Eklundh JO. Statistical background subtraction for a mobile observer. In: *IEEE International Conference on Computer Vision, ICCV*. 2003. pp. 67-74
- [66] Shen Y. Video stabilization using principal component analysis and scale invariant feature transform in particle filter framework. *IEEE Transactions on Consumer Electronics*. 2009;**55**(3): 1714-1721
- [67] Li-Fen T, Qi P, Si-Dong Z. A moving object detection method adapted to camera jittering. *Journal of Electronics and Information Technology*. 2014; **35**(8):1914-1920
- [68] Black MJ, Jepson AD. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. *International Journal of Computer Vision*. 1998;**26**(1):63-84
- [69] Athanasiou SP. Systematic survey on object tracking methods in video. *International Journal of Advanced*

Research in Computer Engineering and Technology. 2012;**1**(8):242-247

[70] Chaaraoui A, Climent-Perez P. A review on vision techniques applied to human behavior analysis for ambient-assisted living. *Expert System Application*. Sep. 2012;**39**(12): 10873-10888

[71] Hu W, Tan T, Wang L, Maybank S. A survey on visual surveillance of object motion and behaviors. *IEEE Transaction on Systems, Man, and Cybernetics, Part C*. 2004;**34**:334-352

[72] Haritaoglu I, Harwood D, Davis L. Real-time surveillance of people and their activities. *IEEE Transaction on Pattern Analysis and Machine Intelligence*. 2000;**22**:809-830

[73] Hu W, Tan T, Wang L. A survey on visual surveillance of object motion and behaviors. *IEEE Transaction on Systems, Man, and Cybernetics, Part C*. Aug. 2004;**34**(3):334-352

[74] Tsai D-M, Lai S-C. Independent component analysis-based background subtraction for indoor surveillance. *IEEE Transaction on Image Processing*. Jan. 2009;**18**(1):158-167

[75] Stauffer C, Grimson WEL. Learning patterns of activity using real-time tracking. *IEEE Transaction on Pattern Analysis and Machine Intelligence*. Aug. 2000;**22**(8):747-757

[76] Yilmaz A. Object tracking: A survey. *ACM Computing Survey*. Dec. 2006; **38**(4):1-45

[77] Wu D, Shao L. Silhouette analysis-based action recognition via exploiting human poses. *IEEE Transaction on Circuits Systems and Video Technology*. Feb. 2013;**23**(2):236-243

[78] Li L, Huang W, Gu IY-H, Tian Q. Statistical modeling of complex backgrounds for foreground object

detection. *IEEE Transaction on Image Processing*. Nov. 2004;**13**(11):1459-1472

[79] Krahnstoever N, Yu T, Lim S-N, Patwardhan K, Tu P. Collaborative real-time control of active cameras in large scale surveillance systems. In: *Proceeding Workshop, France*. Oct. 2008. pp. 1-12

[80] Elaiw A, Al-Turki Y, Alghamdi M. A critical analysis of behavioural crowd dynamics—From a modelling strategy to kinetic theory methods. *MDPI Symmetry Journal*. July 2019;**11**(851): 1-11