New Trends on Moving Object Detection in Video Images Captured by a moving Camera: A Survey

Mehran Yazdi, Thierry Bouwmans

New Trends on Moving Object Detection in Video Images Captured by a moving Camera: A Survey

Mehran Yazdi

Lab. of Signal and Image Proc., Faculty of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran.

Email: yazdi@shirazu.ac.ir

Lab. MIA, Univ. La Rochelle, France.

Thierry Bouwmans

Lab. MIA, University of La Rochelle, La Rochelle, France.

Email: thierry.bouwmans@univ-lr.fr

Corresponding author: Mehran Yazdi (yazdi@shirazu.ac.ir)

Abstract: This paper presents a survey on the latest methods of moving object detection in video sequences captured by a moving camera. Although many researches and excellent works have reviewed the methods of object detection and background subtraction for a fixed camera, there is no survey which presents a complete review of the existing different methods in the case of moving camera. Most methods in this field can be classified into four categories: modelling based background subtraction, trajectory classification, low rank and sparse matrix decomposition, and object tracking. We discuss in details each category and present the main methods which proposed improvements in the general concept of the techniques. We also present challenges and main concerns in this field as well as performance metrics and some benchmark databases available to evaluate the performance of different moving object detection algorithms.

Keywords: *Moving object detection; Moving camera; Background subtraction; Motion compensation.*

1. Introduction

In the field of computer vision, detection of moving objects from a video sequence, which is based on representing moving objects by a binary mask in each frame, is an important issue and interested in many vision based applications such as action recognition [1], traffic controlling [2], industrial inspection [3], human behavior identification [4], and intelligent video surveillance [5]. In many of these applications, a moving camera is inherently utilized. For example, in most intelligent video surveillance systems, we use camera movement techniques such as pan-tilt-zoom (PTZ) to better focus and track the targets [6]. Recently, progress in drone technology with using relatively cheap drones with advanced imaging capabilities promises vast future commercial applications [7]. Here, the camera may operate with various degrees of movement and autonomy. Besides, with advances in camera phone technology for mobile phones, more and more people are interested to capture video sequences with their mobile phone capable to detect and track the moving objects [8]. Here, the camera may have free movements. Even in capturing outdoor scenes by a fixed camera, the camera cannot completely be considered as stationary due to the non-controlled environment [9]. Here, we are facing jitter problems in camera or camera shake problems. Thus, the increasing use of moving cameras along with growing interests in detecting moving objects make it essential to develop robust methods of moving object detection for moving cameras.

In the simple case of a fixed camera, the only changes between consecutive frames are caused by moving objects. However, all these changes are not due to the objects of interest (targets) for a user or a desired application. Concerning an indoor scene, even under a controlled environment, shadow regions and illumination source changes may occur and are undesired for moving object detection [10]. For an outdoor scene, because typically the environment is not controllable, many undesired changes such as branch movement, cloud movement and illumination variations can cause serious problems for moving object detection [9]. Many previous works [11,12] have addressed the moving object detection in video sequences captured by a fixed camera in the presence or absence of undesired changes in the scene. To do that, the principal idea is to create a stable background

modelling and then to apply a background subtraction technique, namely to subtract current frame from background to detect moving objects.

For the case of moving camera, it is important that the method of moving object detection considers not only all problems arise in a fixed camera but also certain difficulties due to compensation of camera motion. This is why a simple background subtraction with a naïve motion compensation model cannot efficiently be applied for a moving camera. Indeed, inaccuracy in motion compensation, which is highly possible for a free movement of the camera, causes the background modelling to fail creating a good model for background and foreground pixels [10].

For detecting moving objects in the case of a moving camera, one strategy is to differentiate the movements caused by moving objects from those caused by the camera. There are two main categories of solution. One is based on background modelling [13,14] which tries to create an appropriate background for each frame of the sequence by using a motion compensation method. Another one is trajectory classification [15,16] in which long term trajectories are computed for feature points using an appropriate tracker and next a clustering approach is used to differentiate the trajectories belonging to the same objects from those of background.

Another strategy is to extend background subtraction methods based on low rank and sparse matrix decomposition developed for the case of static cameras [17,18,19,20,21,22] for the case of a moving camera [23,24]. The principal idea is that if certain coherency exists between a set of image frames, low rank representation of the matrix formed by these frames contains this coherency and sparse representation of this matrix contains outliers. Since the moving objects give intensity changes, which are different from the background and cannot be fitted into the low-rank model of the background, they can be considered as outliers for the low rank representation. Thus, sparse representation of the frames contains the moving objects in these frames. However, it is true based on the assumption that the background is the same for all frames, i.e. the camera is static. Although, this technique cannot directly be applied for the case of a moving camera, where the background changes between frames, a transformation can be integrated into the model in order to compensate for the background motion caused by the moving camera [23,24]. This transformation can be an

2D parameter transform in which the parameters can be adjusted (e.g. using the affine transform for PTZ motions [25] or the perspective transform for free motions [26]).

Object tracking strategy can also be considered as moving object detection although its objective is different. Indeed, in object tracking, typically we mark an object as our desired object (target) and then try to localize it in the next frames of the video sequence. To do that, target information such as histogram, color, texture, statistics etc., is extracted from current frame and then the best candidate in the next frame is obtained using a model of similarity or an appropriate classifier. Finally, the characteristics of the target will be updated to be used for next frames [12].

We will present the methods proposed in each strategy in details and compare their advantages and disadvantages. We will describe different aspects of moving object detection with focus on the case of moving camera. We will also introduce some benchmark video datasets used in the related works and the metrics used to evaluate the performance of the implemented algorithms.

As we have exhaustively searched the publications that presented surveys of different moving object detection methods, nearly all of them have focused on the case of a fixed camera [27,28,29,30,31], where background image pixels maintain their position in the corresponding frames throughout a video sequence. Although the methods introduced in this domain can be applied successfully to the special case of automated surveillance, where the cameras mounted on a fixed platform, they cannot directly be extended for the cases of moving camera such as video taken by mobile phones, hand held cameras or cameras are mounted on a moving platform where the background image pixels do not maintain their position throughout the video sequence. Most of the review publications in this regard have focused on presenting primitives for detecting moving objects in video and methodologies specifically for tracking objects [32,12]. For instance, Shantaiya *et al.* in [32] reviewed the works done under the general term of object detection in video and categorized them as featured based, template based, classifier based and motion based with no constraints on camera motion. In the literature survey [12], it has been introduced various segmentation methods relevant to tracking objects in video and categorized object tracking into point tracking, kernel tracking and silhouette tracking and compared the methods in each category.

In another work, Deori and Thounaojam [33] divided object tracking methods into contour based, feature based and region based. In [34], Parekh *et al.* also focused on tracking objects by dividing it into three steps of object detection, object classification and object tracking and compared the methods proposed in each steps. In all these surveys, no constraints on camera motion were imposed. A survey on moving object detection for mobile camera has been introduced in [35], however, Sanap *et al.* briefly discussed some categories and did not clearly compare the methods. In [36], although Shahre and Shende elaborated the moving object detection in presence of static and dynamic background, their main concern has been the case of fixed camera and the case of moving camera was discussed shortly by considering the complexities arise in this case without categorizing the methods.

The goal of this survey is to group moving object detection methods for moving camera into broad categories and give general concepts for representative methods in each category. The main contributions of this survey are as follows:

- We categorize the methods in different ways based on the general concept of methods belonging to each category and describe proposed improvements with respect to this general concept.
- We extensively study the whole challenges facing accurate detection of moving objects in the video sequences captured by a moving camera.
- We introduce performance metrics and a list of existing benchmark datasets for evaluating the performance of different proposed methods in the case of moving cameras.

The paper is organized as follows. In the next section, we discuss different aspects regarding the detection of moving objects in a video sequence. In section 3, we group the methods of moving object detection for moving camera into different categories, describe their general concept and introduce main methods which brought the improvements in each category. We then introduce in section 4 some benchmark datasets used to compare the different methods of moving object

detection. Section 5 presents some performance metrics used to evaluate the algorithms for moving detection for moving camera. Finally, we terminate the paper with a discussion and conclusion.

2. Challenges to moving object detection in video

The objective of moving object detection is to take a video sequence from a fixed/moving camera and outputs a binary mask representing moving objects for each frame of the sequence. However, this is not an easy task to do due to many challenges and difficulties involved when a camera is used to capture a video sequence of moving objects. Here, we present these challenging issues in details.

2.1. Moving object definition

The definition of a moving object in video sequences is a challenging issue in computer vision domain. The general idea of moving object detection is to represent a set of connected pixels of an image in the video sequence having a coherence motion over time (temporal aspect) and having a semantic similarity over image space (spatial aspect) [12]. It means that we should consider the spatio-temporal relationships of pixels to well detect a moving object. However, many methods consider only temporal aspect of pixels to detect moving objects. Moving objects could be people, animals, all kind of vehicles such as cars, trucks, air planes, ships, etc. For these examples, usually the time-dependent position is relevant and we can represent them by moving pixels or moving bounding box over image sequences of a video [12]. In contrary, for some other examples such as hurricanes, forest fires, oil spills and regions on the water due to the kiteboarder's motion, we consider these regions as "stuff" [37] in computer vision which are irrelevant for moving object detection and should be considered as background. Some objects may have complex shape such as hand, fingers which cannot be well defined by a simple geometric shape representation. We consider them as non-rigid moving objects [38]. Some others such as floating clouds and swaying trees are not the desired objects for being detected and should be considered as background in computer vision [12]. Therefore, a moving object is a rigid or non-rigid thing that

moves over time in image sequences of a video captured by a fix or moving camera and is targeted to be detected and localized (or tracked) in the video. Target(s) can be a single moving object or multiple moving objects for being detected and tracked in videos. Hereafter, depending on environment in which the detection is performed and the end use for which the detection is desired, the motion and appearance of the object should be modelled.

2.2. Challenging difficulties

There are many difficulties in developing a robust method to detect moving objects. In this section, some major challenging difficulties in detecting moving objects in video sequences captured by moving cameras are reviewed. Many of these difficulties are related to both fixed and moving cameras and few of them are specifically related to moving cameras. We will elaborate here most of these difficulties and present the methods which tried to address and handle each challenging case.

2.2.1. Illumination variation

Lighting conditions of the scene and the target might change due to motion of light source, different times of day, reflection from bright surfaces, weather in outdoor scenes, partial or complete blockage of light source by other objects etc. The direct impact of these variations is that the background appearance changes which causes false positive detections for the methods based on background modelling. 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. Thus, it is required for these methods to use features which are invariant to illumination. Figure 1 shows an example of this challenge in a video sequence.

Fig. 1: An example of illumination variation challenge (David indoor in the Ross dataset [39])

Many authors have tried to handle the consequences related to illumination variations. In early related works, different color spaces were used to handle this challenge. For instance, in [40], a color modelling approach that includes intensity information in HSV (Hue-Saturation-Value) color space using B-spline curves for real time tracking under various illumination conditions was proposed. The proposed tracker is able to be adapted to irregular illumination variations and abrupt changes of brightness. Then combination of color and texture was used by many works to tackle this problem. For instance, Shen *et al.* [41] suggested a method for tracking non-rigid moving objects under varying illumination. In this work, the Bayesian framework along with the combination of a specially designed color model and texture model through a level set partial differential function were used. The color and texture models, which have been used in this approach, can extract robust information which is not sensitive to illumination variations. The proposed algorithm is able to discriminate between the temporal variations caused by object motion and illumination changes. Later, local features were used to make moving object detection algorithms more robust to illumination variation. In this regard, Heikkila and Pietikainen [42] proposed an algorithm based on adaptive local binary pattern which can be well tolerant to illumination variations for indoor and outdoor scenes. Cogun and Cetin [43] proposed an 2D-Cepstrum approach for tracking the objects under illumination variations and moving cameras.

They extracted the Cepstral domain features of a target region and then used them to track moving objects based on the covariance matrix. In practice, the approach provided good robustness to illumination variations. The combination of local features was also used in many works. For instance, St-Charles *et al.* [44] showed that a local representation model using spatio-temporal information of color and texture is more efficient to detect moving objects for moving cameras under various conditions including illumination variations. Recently, Yun *et al.* [45] proposed a moving object detection algorithm based on scene conditional background update scheme being able to evaluate rapidly scene changes and adaptively build a mew background model. This algorithm can effectively deal with illumination variations especially in the presence of moving cameras. It can be concluded that local features of a moving object along with updating background models are more efficient for dealing with this challenge.

2.2.2. Moving object appearance changes

In real scenarios, most objects can move in 3D space, but we only have the projection of their 3D movement on a 2D plane (sequence images), so any rotation in the direction of third axis (along the camera's line of sight) may change the object appearance (for example, a car front view is different from its side view). Moreover, the objects themselves may have some changes in their appearance like facial expressions, changing clothes, wearing a hat, etc. Also the target can be a non-rigid object, where its appearance may change over time. In many applications, the goal is tracking humans or pedestrians, which makes tracking algorithms vulnerable to this challenging case. Figure 2 shows an example of this challenge.

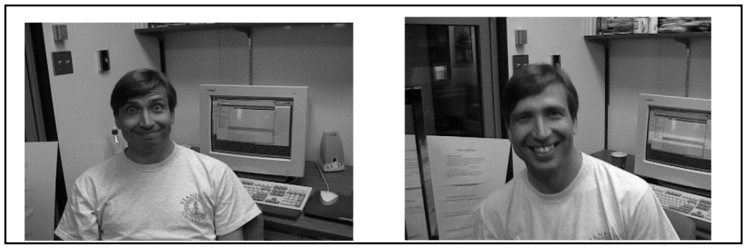

Fig. 2: An example of appearance change challenge (Dudek in the Ross dataset [39])

Different methods are proposed to deal with this challenge. Early methods were based on learning representation of moving objects. For instance, Lim *et al.* [46] presented a tracking method that incrementally learns a low-dimensional subspace representation. The proposed method efficiently adapts online to changes in the appearance of the moving objects. In another work, Porikli *et al.* [47] described objects using a covariance matrix and applied an update mechanism using Lie algebra which can adaptively track moving objects under their appearance changes for moving camera. Some other methods considered appearance changes due to the presence of non-rigid moving objects. For instance, Balan and Black [48] proposed a robust, adaptive, appearance model based on the Wandering-Stable-Lost framework for tracking of articulated objects (human body parts). They have modeled the appearance using a mixture model that includes an adaptive template, frame-to-frame matching and an outlier process. Their tracking results for a walking sequence (including 180 degree turns) depicts the ability of this algorithm to deal with significant appearance changes. Recently, in [49], the authors suggested to learn motion patterns in videos to handle appearance changes due to moving objects even in the presence of moving cameras. The approach is based on a trainable model which uses the optical flow as input features to a fully convolutional network for separating independent objects from camera motion. Their results show a good performance in handling appearance changes.

It seems that the success of the methods proposed to handle this challenge is highly dependent on using a good representation of moving objects and applying an efficient learning strategy.

2.2.3. Presence of abrupt motion

Sudden changes in the speed and direction of the object's motion or sudden camera motion are another challenges in object detection and tracking. If the object or camera moves very slowly, the temporal differencing methods may fail to detect the portions of the object coherent to background. Meanwhile, a very fast motion produces a trail of ghost detected region. So, if these object's motions or camera motions are not considered, the object cannot correctly be detected by

methods based on background modelling. On the other hand, for tracking based methods, prediction of motion becomes hard or even impossible and as a result, the tracker might lose the target. Even if the tracker does not lose the target, the unpredictable motion can introduce a great amount of error in some algorithms. An example of this challenge is shown in Fig. 3.

Fig. 3: An example of abrupt motion challenge (Motocross in the Kalal dataset [50])

There have been some works that consider this problem. In an early work, Kwo and Lee [51] proposed an algorithm based on the Wang-Landau Monte Carlo sampling method that efficiently deals with the abrupt motions. They have integrated the Wang-Landau algorithm into the Markov chain Monte Carlo based tracking method. It is shown that their method can accurately track the objects with drastically changing motions. Then, in [52], a sampling-based tracking scheme is proposed and tries to handle abrupt motion problem in the Bayesian filtering framework. They have introduced the stochastic approximation Monte Carlo (SAMC) sampling method into the Bayesian filter tracking frame framework. Their results show that this method is effective against abrupt motions and is also computationally efficient. Wang and Lu [53] introduced a Hamiltonian Markov Chain Monte Carlo (MCMC) based tracking algorithm for handling abrupt motion. They have integrated the Hamiltonian Dynamics into traditional MCMC tracking method. Their results show that this method is effective in handling different type of abrupt motions. Recently, Zhang *et al.* [54] proposed an extended kernelized correlation filter tracker based on swarm intelligence method which can effectively handle abrupt motion tracking in videos.

Due to the nature of this challenge which is an unpredictable event, the methods are so various and there is no robust solution that works in all conditions

2.2.4. 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 in the sidewalk, it may be occluded by trees, cars in the street, other pedestrians, etc. Occlusion severely affects the detection of objects in background modelling methods where the object is completely missing or separated into unconnected regions. If occlusion occurs, the object's appearance model can change for a short time which can cause some difficulties for the object tracking methods. Figure 4 shows this challenge.

Fig. 4: An example of occlusion challenge (Car in the Kalal dataset [50])

There are several works that try to handle occlusions. Early attempts were focused on appearance models. For instance, In the approach proposed by Jepson *et al.* [55], the EM algorithm in addition to an appearance model based on the filter responses from a steerable pyramid was exploited to overcome changing appearance and occlusion problems. To handle occlusion, in [56], tracking was achieved by evolving the contour from frame to frame by minimizing some energy function evaluated in the contour vicinity defined by a band. Senior *et al.* [57] suggested that by maintaining appearance models of moving objects over time, occlusions can be more efficiently handled. Pan and Hu [58] proposed an algorithm that progressively analyzes the occlusion situation using the spatio-temporal context information which is also checked by reference object and motion constraints. By this approach, the tracker is able to distinguish the object in occlusions effectively. Recently, occlusion problem was addressed in [59] for moving cameras. The method

integrates a deformable part model based on histogram of oriented gradient (HOG) into a multiple kernel tracker based on mean shift. This approach can effectively handle occlusions in different conditions of moving cameras.

It seems that occlusion can be more efficiently handled if a good appearance modelling of moving objects is updated appropriately.

2.2.5. Complex background

The background may be highly textured, especially in natural outdoor environments where high variability of textures present in outdoor scenes. Moreover, the background may be dynamic, namely some regions of the background may contain movement (e.g., a fountain, clouds in movement, traffic lights, trees waggle, water waves (see Fig. 5), etc.), which should be considered as background in many moving object detection algorithms.

Fig. 5: An example of dynamic background challenge (Kitesurf in the Zhang dataset [60])

Such movements can be periodic or non-periodic. In [14], adjustable polygons were introduced as a novel active contour model, which is a set of active segments that can fit any object shape. Then, Monnet *et al.* [61] directly addressed the problem of dynamic background by proposing an auto-regressive model to predict the behavior of such backgrounds for effectively detect moving objects. In [62], first the principal features based on statistical characteristics of image pixels were extracted and then a Bayesian decision rule was used to classify image pixels into moving objects and background. Due to time varying adaptation of background features, the algorithm can effectively overcome the complexity of background. A new energy based on textural characteristics of objects was also proposed to make the algorithm able to perform well in complex

backgrounds. For modelling and classification of a dynamic background, dynamic texture modelling methods [63,64] can also be utilized. Parameters of such a background model can be effectively used as a cue for moving object detection. Recently, Minematsu *et al.* [65] suggested an adaptive background model registration based on homography motion estimation to handle highly complex backgrounds for detecting moving objects for moving cameras.

In general, detection of moving objects in complex background requires a good modelling of background and a precise estimation of camera motions.

2.2.6. Shadow

The presence of shadows in video image sequences complicates the task of moving object detection. Shadows occur due to the block of illumination from the light source by objects. 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 object 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. However, dynamic shadows are still difficult to be distinguished from moving objects, especially for outdoor environment where the background is usually complex. Figure 6 shows an example of this challenge.

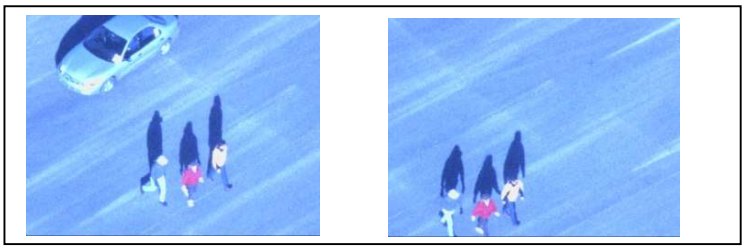

Fig. 6: An example of shadow challenge (Pedestrain4 in the Kalal dataset [50])

Fortunately, there have been a lot of surveys written to compare various shadow removal methods in the literature. An early and good comprehensive survey of moving shadow detection algorithms and the suggestion of quantitative metrics to evaluate their performance were done by Prati *et al.* [66]. In other survey, Sanin *et al.* [67] followed the work done in [66] but with comparing recent publications and using a more comprehensive set of test sequences. Recently, in their survey, Al-Najdawi *et al.* [68] brought an exhaustive comparison on cast shadow detection algorithms based on object/environment dependency and implementation domain. Recent algorithms have been compared in a survey written by Tiwari *et al.* [69] who categorized the shadow detection algorithms in different scenarios such indoor/outdoor scenes, fixed/moving cameras, and umbra/penumbra shadows. More recent works were proposed in [70] and [71]. In [70], the authors introduced a modified Gaussian mixture model to handle a highly dynamic environment in other to overcome shadow problem for moving object detection for moving camera. Their algorithm works in real time and achieves good accuracies. Song *et al.* [71] proposed a shadow elimination algorithm using HSV colour space and texture features. Indeed, HSV features can effectively distinguish shadow regions if an appropriate threshold is used. In this approach, Otsu's thresholding algorithm was used for accurately eliminating shadows in the presence of camera motion.

It is worthy to mention that a good segmentation of moving objects and use of appropriate features of each segmented region (including shadow) can work better to eliminate the shadow cast.

2.2.7. 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 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 (see Fig. 7).

Fig. 7: An example of low quality image challenge (Car1 in the BLUT dataset [72])

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 confuse moving object detection algorithms if they are not well designed and able to deal with such video qualities. In [73], a method for object tracking for soccer robots was proposed when the color information is not available and the camera is a low quality black and white one. It exploits Haar-like features and AdaBoost algorithm to get a colorless representation of the ball and the tracking is done using a particle filter. It shows that even by a low quality camera, we can still detect and track the objects. Hua *et al.* [74] proposed a pixel wise object tracking algorithm that is not sensitive to noise. This approach is based on a data grouping algorithm that adds reliability evaluation into the K-means clustering. Then, by considering the fact that a noisy data is typically far from any cluster center and taking into account the distance of nearest cluster centers, low reliability is assigned to noisy data and noisy data are ignored. In [75], the authors used a modified affine transformation model to handle partial lens distortion to better detect moving objects. In [76] Li *et al.* proposed a cascade particle filter in tracking and detection for low frame rate videos. The algorithm needs a complex detector which is computationally expensive. For underwater videos, where the image quality is very low, Spampinato *et al.* [77] proposed a time variant feature extraction technique that can be adapted to different water conditions and efficiently detect moving fishes. In [78], a Blur-driven tracker

(BLUT) framework is presented, that is used to track motion-blurred objects. This algorithm uses the blur information without deblurring. It uses the motion information inferred by the blurs to guide the sampling process in the particle filter based tracking. The results show that this method can robustly track severely blurred targets. In videos with high frame rates, Dollar *et al.* [79] extracted features in pyramidal structure in order to fast detect moving objects. Recently, in [80], a tracking system for low frame rate videos was proposed which is used a dominant color based appearance model and the APSO based framework search to track moving objects. When we deal with low resolution video sequences, the techniques based on fusion can be used to better detect moving objects. In [81], a heterogeneous feature based technique was proposed for detecting human movements in low resolution conditions which can be utilized for outdoor video sequences.

There are many challenging aspects related to fix/moving cameras and for each aspect, different techniques with different degrees of success have been proposed. There is no common solution that can be applied in all cases and the techniques vary based on the camera condition and its applications.

2.2.8. Camera motion

When dealing with detecting moving objects in present of moving cameras, the need for estimating and compensating the camera motion is evitable, however it is not an easy task to do because of possible camera's depth changes and its complex movements. Many works elaborated an ease scenario by considering simple movements of the camera, i.e. PTZ [82,83,84] (see Figs. 8 and 9). 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. A homography is an invertible transformation that relates points in two views [85].

Fig. 8: An example of panning in camera in the CDNET database [86]

Fig. 9: An example of zooming in camera in the CDNET database [86]

In PTZ, the optical center of the camera is fixed, which is the case for many applications such as intelligent camera surveillance and monitoring. When the camera is displaced, the optical center of the camera moves and the 3D scene captured by the camera cannot be considered as planar and the images belonging to different planes create parallax and require a complex transformation and registration. Many works also ease this scenario to estimate camera motion by imposing various constraints assuming the existence of a dominant plane [87,88,89,90]. This case is very common in aerial video sequences or for cameras mounted on a moving platform. Difficulties arise when camera moves freely in any directions with various depth changes such as handheld cameras (see Fig. 10).

Fig. 10: An example of freely motion of camera in the Michigan University dataset [91]

Here, we need to use multiple homography transforms to create a multi plane representation of the 3D scene [92,93,94] which is computationally very complex. By assuming that the camera's parameters are known or can be estimated [95], the homography can be relaxed and its computational complexity is decreased. Another camera movement considered by many works for moving object detection is camera shaking or jitter. This situation is common for handheld cameras where the handshakes unconsciously or videos captured by unstable cameras (e.g. vibrated by an external source such as wind). To handle this camera motion, some methods use image stabilization techniques based on motion compensation of feature points to achieve a stable background and then apply background subtraction methods to detect moving objects [96,97]. For the case of handheld cameras, camera motion models (eg., 3D motion models [98] or [99]) can be efficiently used to compensate camera vibrations.

In general, the camera motion is not a simple issue. The simple case is PTZ cameras, in which the platform, where the camera is mounted, is stable. For handheld cameras, stability is a challenging task but the camera movement is still usually limited. However, for airborne cameras, stability is an important challenge, where the movement of the camera is rapid and in a wide range. For handling camera motion in most methods, a pre-information about the motion of camera is required to apply an adaptive motion model based on this pre-information in order to effectively compensate the camera motion and more accurately detect moving objects.

2.2.9. Non-rigid 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 (see Fig. 11). When dealing with detecting such moving objects, most algorithms detect different parts as different moving objects. It produces an enormous challenge especially for non-rigid objects and in the presence of moving cameras.

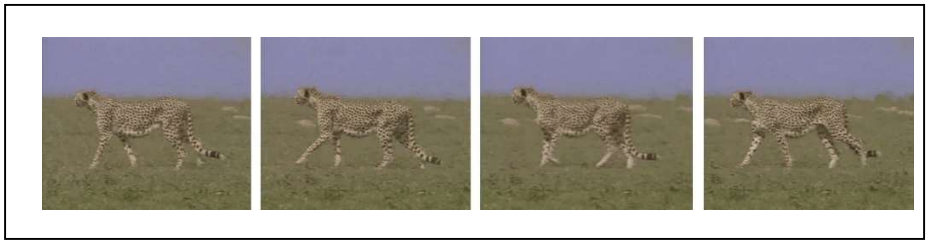

Fig. 11: An example of non-rigid moving object in a video sequence [100].

One strategy followed by many works to overcome this challenge is to use articular models for moving non-rigid objects. In these models, each part of an articulated object is allowed to have different movements. In an early work by Black and Jepson [38], a view-based eigenspace representation was proposed which used optical flow for motion estimation of each articulate part of an object. The mentioned representation allows unifying articulate parts as one moving object. This approach produced good results for hand tracking. In another method [101], a complete model of articulated objects composed of rigid bodies was proposed and applied for tracking and pose estimation of manipulated robots. The success of this approach was mostly due to use of depth sensors in order to accurately produce a depth image representing different poses of a moving object. Another strategy is based on the fact that if different moving parts belong to one object, they should have certain similarity which can be served to unify them. Hereby, a segmentation before tracking can be used to overcome this challenge. In a recent work done in [102], the authors considered a target object as a combination of different segments having different movements. Then, tracking is performed for each segment and certain weights are assigned to each segment such that higher weights are given to segments with motion consistency. Finally an across class similarity measure based on colour histogram allows unifying adjacent moving segments.

Most methods to handle this challenge either use a pre-information about the articulated object to adopt an appropriate moving object detection approach or first segment a moving object

and then merge different moving segments based on a similarity criterion. Table 1 presents an overview of the studied challenges and related works.

Table 1. Challenges for moving camera and related reference works

Challenge	**Representative works**
Illumination variation	Lee *et al.* [40] (2001) Shen *et al.* [41] (2006) Heikkila and Pietikainen [42] (2006) Cogun and Cetin [43] (2010) St-Charles *et al.* [44] (2015) Yun et al. [45] (2017)
Moving object appearance change	Lim *et al.* [46] (2004) Porikli *et al.* [47] (2006) Balan and Black [48] (2006) Tokmokov *et al.* [49] (2017)
Presence of abrupt motion	Kwo and Lee [51] (2008) Zhou *et al.* [52] (2012) Wang and Lu [53] (2012) Zhang *et al.* [54] (2017)
Occlusion	Jepson *et al.* [55] (2003) Yilmaz *et al.* [56] (2004) Senior *et al.* [57] (2006) Pan and Hu [58] (2007) Hour *et al.* [59] (2017)
Complex background	Delagnes *et al.* [14] (1995) Monnet *et al.* [61] (2003) Li *et al.* [62] (2004) Chetverikov and Péteri [63] (2005) Arashloo *et al.* [64] (2017) Minematsu *et al.* [65] (2017)
Shadow	Prati *et al.* [66] (2003) Sanin *et al.* [67] (2012) Al-Najdawi *et al.* [68] (2012) Tiwari *et al.* [69] (2016) Xia *et al.* [70] (2016) Song *et al.* [71] (2017)
Problems related to camera	Treptow and Zell [73] (2004) Hua *et al.* [74] (2007) Unger et al. [75] (2008) Li *et al.* [70] (2008) Spampinato *et al.* [77] (2008) Wu *et al.* [78] (2011)

	Dollar et al. [79] (2014) Zhang *et al.* [80] (2015) Chen *et al.* [81] (2017)	
Camera motion	PTZ	Rowe and Blake [82] (1996) Ren *et al.* [83] (2003) Avola *et al.* [84] (2017)
	Plane+parallax	Irani *et al.* [87] (1997)) Irani and Anandan [88] (1998) Sawhney *et al.* [89] (2000) Zhou *et al.* [90] (2017)
	Free motion	Wang and Adelson [92] (1994) Xiao and Shah [93] (2005) Chen and Lu [94] (2017)
	Vibration	Shen *et al.* [96] (2009) Wang *et al.* [98] (2009) Koh *et al.* [99] (2011) Li-Fen *et al.* [97] (2014)
Non-rigid object deformation		Black and Jepson [38] (1998) Schmidt *et al.* [101] (2015) Lin *et al.* [102] (2017)

2.3. Post processing

Nearly in all methods of moving object detection, post processing is inevitable due to imperfectness in algorithms such as motion estimation, background modelling, segmentation, classification etc. Even if the estimated background produced by a background subtraction method in the case of static camera is almost perfect, object regions may have a similar color or texture to the background results in producing holes in binarized output images. For enhancing the results in such cases, Hoynck *et al.* [103] suggested a segmenting step based on neighboring relations and similarity measurements between adjacent segments. Another problem faced by background subtraction methods, which severely affects the results of moving object detection, is the noise. Although the noise introduced by cameras can be reduced using an appropriate threshold, random noise patterns may appear after the thresholding and should be removed by applying a post processing step. Parks and Fels [104] compared different post processing techniques for noise

reduction and suggested that morphological operators produce better results than other techniques such as median filter. Shadow and reflections are also serious problems since their region pixels are wrongly considered as moving objects after applying most moving object detection methods. Although it is better to handle these problems by designing an effective algorithm at the time of detection, post processing steps can be well applied to remove these false errors in the output results. For instance, Kartika and Mohamed [105] suggested a post processing technique based on adaptive thresholding and shadow detection in HSV color space to correct these false errors. When a trajectory based moving object detection is used, some over-segmentation errors may be produced during clustering trajectories. To correct such errors, Brox and Malik [15] proposed a merging step based on the mutual fit of motion models of trajectories. In this regard, in [106] a linking step was suggested based on region convolutional neural network which can link mini-trajectories (tracklets) disconnected over time. Sometimes discontinuities might occur in the boundary of detected moving objects. In order to overcome this problem, a linking process can be added to detection algorithm to make the boundaries of moving objects continuous. In [107], a linking technique based on morphological operators was used to associate moving object segments and achieve a unified moving object.

In order to remove local jitter and annoying shaking motion in videos, a technique of video stabilization can be performed. For instance, in [108] by using a matching score filtering, low frequency components were kept for preserving global motion and removing unwanted shakes in video sequences.

In general, it is suggested to use morphological operations and connected component labeling as common post processing techniques to refine the edge of resulting moving object silhouette and remove false errors.

2.4. Real time aspects

Although most methods for moving object detection work offline by analyzing recorded video sequences, nowadays the need for real-time moving object detection has received a

considerable attention, especially in some domains such as sociology, criminology, suspect behavior detection and tracking, traffic accident detection, crowd tracking and vehicle and robot navigation. In general, the need for a real-time system imposes very low computational time and minimal hardware with low memory requirements. We review some methods, which proposed a real time platform for moving object detection and tracking in video sequences as follows.

In early work done in [109], a new method for real-time tracking of non-rigid objects seen from a moving camera was introduced. The central computational module is based on the mean shift iterations which find the most probable target position in the current frame. The dissimilarity between the target model (its color distribution) and the target candidate is expressed by a metric derived from the Bhattacharyya coefficient. The capability of the tracker to handle in real-time partial occlusions, significant clutter and target scale variations is demonstrated for several image sequences. Then challenges for outdoor environment was investigated by [110]. Ponga and Bowden [110] presented a variety of probabilistic models for tracking small-area targets which are common objects of interest in outdoor visual surveillance scenes. The authors address the problem of using both appearance and motion models in classifying and tracking objects, when detailed information of the object's appearance is not available. The approach relies upon motion, shape cues and color information to help in associating objects temporally within a video stream. The results show that the system can successfully track multiple people moving independently and is able to maintain trajectories in the presence of occlusions and background clutter. Later, Yang *et al.* [111] presented a real-time system for multiple objects tracking in dynamic scenes. The unique characteristic of the system is its ability to cope with long duration and complete occlusion without a prior knowledge about the shape or motion of objects. The system produces good segment and tracking results at a frame rate of 15-20 fps for image size of 320×240, as demonstrated by extensive experiments performed using video sequences under different conditions of indoor and outdoor environments with long-duration and complete occlusions in changing background. Tracking in a controlled environment was studied by Heinemann *et al.* [112]. They presented a method for detecting and tracking the ball in a RoboCup scenario with robust performance. They

use Haar-like features trained by an AdaBoost algorithm to get a colorless representation of the ball. Tracking is performed by a particle filter. It is shown that the proposed algorithm is able to track the ball in real-time with 25 fps even in a cluttered environment. The problem of online training was investigated by Grabner *et al.* [113]. They proposed a novel on-line AdaBoost feature selection algorithm for object tracking. The distinct advantage of this method is its capability of on-line training. By using fast computable features (e.g. Haar-like wavelets, orientation histograms, local binary patterns), the algorithm can run in real-time. The performance of the algorithm is evaluated using sequences that contain changes in brightness, view-point and further appearance variations. The results show that the proposed tracker can cope with all these variations and the algorithm is at least as good as those presented in the according publications. The problem of hardware implementation was also studied in some works. In [114], for tracking moving objects in real-time without delay and loss of image sequences, a particle filter algorithm was implemented specifically for an electronic circuit for the goal of object tracking. This circuit was designed by VHDL (VHSIC Hardware Description Language) and implemented in an FPGA (Field Programmable Gate Array). The tracking in a camera network system was also investigated by some researches. In [115], an automated surveillance system was proposed. This algorithm is deployed in a variety of real-world scenarios ranging from railway security to law enforcement. This algorithm has been actively used in several surveillance-related projects funded by different government and private agencies. It is shown that the proposed algorithm can detect and classify targets and seamlessly track them across multiple cameras. It also generates a summary in terms of key frames and the textual description of trajectories for a monitoring officer for final analysis and response decision. Current system limitations include the inability to detect camouflaged objects, handling large crowds, and operating in rain and extreme weather conditions. In [116], a probabilistic framework for robust real-time visual tracking of previously unseen objects from a moving camera was derived. The tracking problem is handled using a bag of pixels representation and comprises a rigid registration between frames, segmentation and online appearance learning. The registration compensates for rigid motion; segmentation models any residual shape

deformation and the online appearance learning provides continual refinement of both object and background appearance models. Tracking in night is a challenging task, especially for real time applications. In [117], a novel real time object detection algorithm was proposed for night-time visual surveillance. The algorithm is based on contrast analysis. In the first stage, the contrast in local change over time is used to detect potential moving objects. Then motion prediction and spatial nearest neighbor data association are used to suppress false alarms. Experiments on real scenes show that the algorithm is effective for night-time object detection and tracking. Another problem is the shadows that occasionally remain connected to the silhouette after motion detection. They can cause the tracker to lose the target resulting in disjointed trajectories. A real-time tracking algorithm should be robust in different scenarios of moving camera. In [118], a novel approach for global target tracking based on mean shift technique was proposed. The proposed method represents the model and the candidate in terms of background weighted histogram and color weighted histogram, respectively, which can obtain precise object size adaptively with low computational complexity. Experimental results on various tracking videos and its application to a tracking and pointing subsystem show that the proposed method can successfully cope with different situations such as camera motion, camera vibration, camera zoom and focus, high-speed moving object tracking, partial occlusions, target scale variations, etc. Danelljan *et al.* [119] investigated the contribution of color in object tracking for real-time systems. The results suggest that color attributes provide the superior performance for visual tracking. They further propose an adaptive low dimensional variant of color attributes for real-time aspect. Both quantitative and attribute based evaluations were performed on 41 challenging benchmark color sequences for some of them the tracker must deal with camera motion. Furthermore, they have shown that the proposed approach outperforms state-of-the-art tracking methods while running at more than 100 frames per second. Recently in [120], real-time multiple object tracking was addressed by using a region based convolutional neural network (RCNN) for object detection and by creating a regression network for generic object tracking. Their approach can be effectively used for autonomous navigation. More recently, Minaeian *et al.* [121] performed keypoint tracking by using local motions for

moving objects in the case of moving camera mounted on aerial vehicles. Their work achieved good results in real tome applications. A benchmark for testing object tracking methods on higher frame video datasets in the presence of moving camera was proposed in [122] which can be effectively to evaluate the performance of methods in terms of speed and accuracy. Table 2 shows, in an overview table, which platform is used in each method and if it deals with moving cameras or not.

Table 2: Scenarios and platforms used in each work for real-time moving object detection.

Reference paper	**Non-static camera**	**Implemented Platform**
Comaniciu *et al.* [109] (2000)	Yes	PC
Ponga and Bowden [110] (2003)	Not mentioned	PC
Yang *et al.* [111] (2005)	Not mentioned	PC
Heinemann *et al.* [112] (2006)	Yes	Soccer robots
Grabner *et al.* [113] (2006)	Not mentioned	PC
Cho *et al.* [114] (2006)	Not mentioned	FPGA
Shah *et al.* [115] (2007)	Not mentioned	PC
Bibby and Reid [116] (2008)	Yes	PC
Huanga *et al.* [117] (2008)	Yes	PC
Li *et al.* [118] (2010)	Yes, vibrations	PC
Danelljan *et al.* [119] (2014)	Yes	PC
Agarwal and Suryavanshi [120] (2017)	Yes	PC
Minaeian *et al.* [121] (2018)	Yes	On board system

3. Moving object detection methods

Contrary to vast algorithms proposed for the moving object detection in the case of fixed camera over last decade, few algorithms directly focused their work on the case of moving camera and most of them have been recently published since this subject became lately very attractive and

showed many future promising applications. We can broadly classify all efforts in this field in four categories which will be detailed as follows.

3.1. Modelling based Background subtraction

One of the most common techniques used to detect moving objects in a video sequence captured by a moving camera is background subtraction based on background modelling. The general concept of these techniques is shown in Fig. 12. As can be seen, first a background model is initialized using a set of first frames of the sequence. Typically the model is created based on statistical features extracted from the background. Next, some feature points are extracted from the current frame and then their correspondences in the background are found. However, the correspondences for all pixels of the image (each frame considered as a static image) are not computed because it is computationally expensive and arise many errors in the result. After that, the transform matrix between current frame and background is computed using previously found correspondences. Now the pixels of the current frame's background can be easily computed by applying the inverse transform. While the current frame and its background are available, an appropriate classifier can determine which pixel belongs to background and foreground. Finally, the statistical model of the background is updated and will be used for the next frame. In each step of this general technique, different algorithms have been proposed to improve the performance of the results. We classify and review these algorithms as follows.

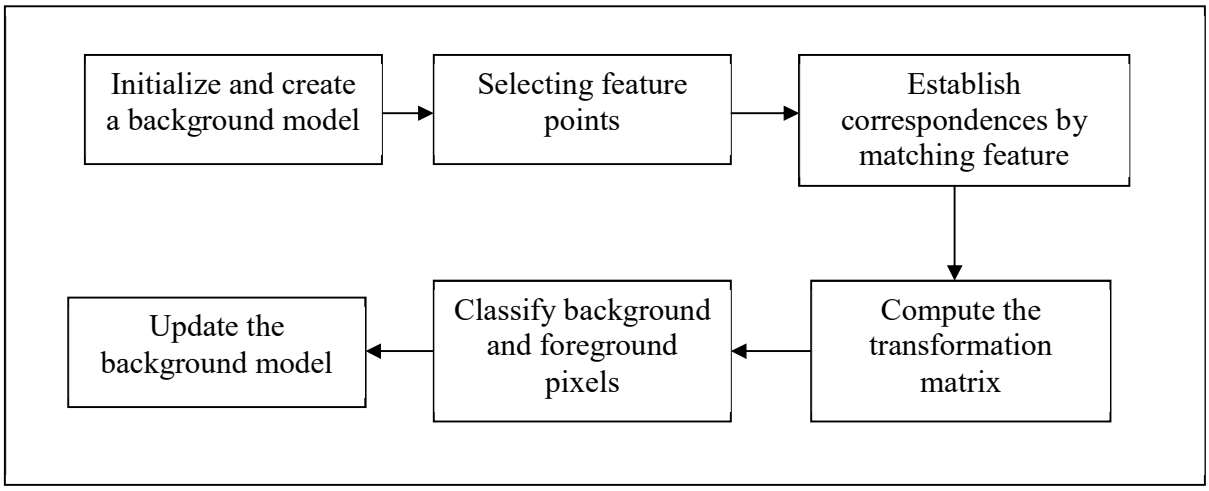

Fig. 12: Flowchart of modelling based background subtraction techniques

-*Background model creation*: Wren *et al.* [13] used a simple Gaussian model, which was not robust to a complex background. Hayman and Eklundh [95] used a mixture of Gaussian (MoG) which was more appropriate for representing a complex background. Zivkovic and Van der Heijden [123] proposed an adaptive MoG, which was capable to change the parameters of the model for new scenes. Yi *et al.* [124] suggested that a double Gaussian model can be well represent most outdoor scenes. Recently, Cuevas *et al.* [125] changed the statistical approach and used a spatio-temporal model based on kernel. They showed that this model can be easily adapted to new changes in the background and is also robust to complex backgrounds.

-*Feature point selection*: Zivkovic and Van der Heijden [123] used Laplacian gradient to find the interesting points in the current frame, however it is very sensitive to noise, which is present most of the time in the frame. Setyawan *et al.* [126] proposed to use the Harris corner detector to extract the feature points. This approach is less sensitive to the noise and robust to illumination changes and served commonly by other methods.

-*Correspondence matching*: Unger *et al.* [75] proposed the optical flow for global motion estimation and concluded that this leads to better results for unconstraint camera motion. Setyawan *et al.* [126] used Kanade-Lucas-Tomasi (KLT) tracker, which makes use of spatial intensity

information to direct the search for the position that yields the best match. It is faster than traditional techniques for examining far fewer potential matches between the frames.

-*Transformation matrix*: Jin *et al.* [127] used a multi-layer homography transform which works well for complex changes between frames and a free moving camera. Lenz *et al.* [128] proposed a complex homography to create an 3D background, when a stereo camera was used which is, by the way, computationally expensive. Setyawan *et al.* [126] suggested that a simple 2D projection can be efficient, when we are facing a slow motion of the camera. Minematsu *et al.* [129] also suggested a planar tomography, when the camera moves with no depth changes and no significant parallax occurs between frames. Viswanath *et al.* [130] concluded that a direct linear transform can be efficient, when the camera motion is just PTZ.

-*Classifier*: Maddalena and Petrosino [131] used a non-parametric model based on neural network for classifying background and foreground pixels, which needs however a long training phase. Yi *et al.* [124] suggested using a simple thresholding approach to speed up the whole process. Cuevas *et al.* [125] used the Bayesian classifier, which can statistically minimize the errors.

Some works use another pipeline strategies for modelling of background. Here, we present some recent related works. Zhu and Elgammal [132] proposed a new multilayer representation of background and foreground objects by estimating their motion and appearance model. Then, based on the collection of probability maps produced by each representation, the pixel-wise segmentation for the current frame is created by multi-label Graph cut. In [133], Zhou and Maskell used a background subtraction scheme and a motion compensation approach to address moving object detection for video of urban areas taken by moving airborne cameras. Despite of its complexity, their approach can effectively detect moving object, even when heavy parallax is present among frames of a video sequence. Recently, a background subtraction method based on a coarse-to-fine thresholding scheme was proposed by [134], which has low complexity and is robust against free motion of cameras. In another recent work, Gong *et al.* [135] proposed a background subtraction based on codebook modelling which uses full colour information to detect moving objects, when

the camera was mounted on a moving vehicle. One difficult case of moving cameras for detecting moving objects is when the camera is mounted on aerial vehicles. Recently, Minaeian *et al.* [121] addressed this case by estimating camera motion using extracted background keypoints and then segmenting foreground to detect moving objects using local motions. Their results demonstrate promising performance for videos taken by a camera on board of unmanned aerial vehicle. However, their approach needs camera setup parameters for robust estimation of movements.

Also recently, due to the availability of large annotated video datasets (e.g., CDnet [136]), deep Convolutional neural networks (CNNs), which require exhaustive training, have been considered for moving object detection for moving cameras. For instance, Babaee *et al.* [137] proposed a single CNN which is learned from data to select meaningful features and estimates an appropriate background model from video. Their approach is capable to be used in real time applications for moving object detection under various video scenes.

In brief, the techniques of modelling based background subtraction have better performance for smooth movement of the cameras and specific case of PTZ. They are also popular for real time applications due to their moderate complexity.

3.2. Trajectory classification

Trajectory classification has been used in many works to obtain moving objects in the video sequences captured by a moving camera. The general concept of this technique is shown in Fig. 13. Preliminary, the interesting points are chosen from the first image of the sequence. These points can be simply the pixels on a grid mesh fitted into the image or feature points extracted by a typical feature point detector. We rarely try to find the trajectories for all pixels of the image since it is a time consuming task and also highly sensible to the noise. Then, for each point, a trajectory representing its continuous displacements in adjacent frames of the sequence is obtained. To do that, we can use different motion models such as an 2D affine transformation for PTZ camera motion or a complex homography considering depth variances as the camera can freely moves into

3D scenes. Finally, a clustering approach is used to classify the trajectories into background and foreground, in which the moving objects can be distinguished.

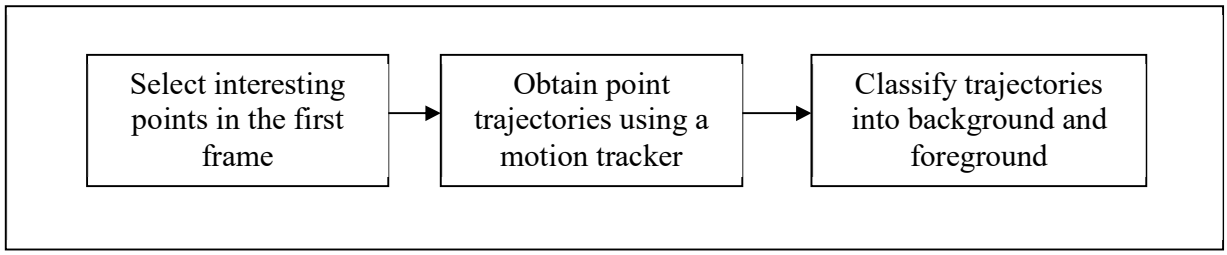

Fig. 13: General concept of trajectory classification technique

Many methods have tried to improve this technique by proposing different algorithms. Sheikh *et al.* [138] used a scale space forming by the principles trajectory bases for the background and then considered the trajectories that are not in this space as belonging to moving objects. Although the method worked based on using only 2D image quantities, the homography constraints were well applied to 3D scenes to handle freely moving cameras. The results showed a good performance of the algorithm while it is computationally time consuming. Brox and Malik [15] used the flow optic to compensate long term motion in a video sequence to obtain the point trajectories. In their work, a spectral clustering using spatial regularity was used to classify background and foreground trajectories. Although the method can handle a large set of sequence frames and partially occluded objects, it fails to obtain a dense segmentation of the objects. Yin *et al.* [16] also used the flow optic to compensate the camera motions and then applied PCA (principal component analysis) to reduce the number of atypical trajectories. For the clustering, they used the watershed transform to differentiate foreground and background trajectories. Although the unlabeled pixels are handled using a label inference, it fails to produce a dense segmentation of moving objects. Recently, Singh *et al.* [139] tackled trajectory detection for freely moving cameras in videos taken when a person wearing the camera. In their work, instead of using complex and complicated models, they use point tracking using optical flow and bag of words classifier to detect trajectory of moving objects. Their approach work well for first person action recognition. More recently, Zhang *et al.* [140] proposed a pre-trained CNN based approach by using learning adaptive discriminate features to

detect moving object trajectories in an unconstrained video over time. They first divide the video sequence into different shots and then detect tracklets (short trajectories) in each shot and finally link tracklets belonging to each moving objects in consecutive shots. Their approach work well for multi face tracking even in unconstrained movement of cameras. However, their approach requires different training data depends on type of moving objects.

In general, the technique of trajectory segmentation suffers from producing a dense segmentation of moving object and being very sensible to precise trajectory localization.

3.3. Low rank and sparse matrix decomposition

The Low rank and sparse decomposition is currently considered to be one of the leading techniques for video background modelling, which consists of segmenting the moving objects from the static background. To do that, Principal Component Pursuit (PCP) [141] is commonly applied to exactly obtain low rank and sparse representations of an observed matrix formed by a set of observed frames of the video via an optimization process defined as follow [17]:

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1 \text{ w.r.t. } A - L - S = 0$$

Where $A \in \mathbb{R}^{m \times n}$ is the observed matrix formed by storing each frame $f \epsilon \mathbb{R}^{M \times N}$ of a set of n frames as a column. $m = M \times N$, L is a low rank matrix ($rank(L) \ll m, n$) and S is a sparse matrix representing outliers. $\|\cdot\|_*$ and $\|\cdot\|_1$ are the nuclear norm and the ℓ_1–norm respectively and λ is a regularization parameter. Here, the low rank representation contains coherent parts in frames, i.e. background information, and the sparse representation contains outliers related to background i.e. moving objects. Although this scheme provides good performance in moving object detection of static cameras, it cannot directly be applied for the case of moving camera where the coherency of background for the consecutive frames does not hold. A common solution is to embed a global motion compensation model into matrix decomposition optimization. Figure 14 shows the general concept of this strategy.

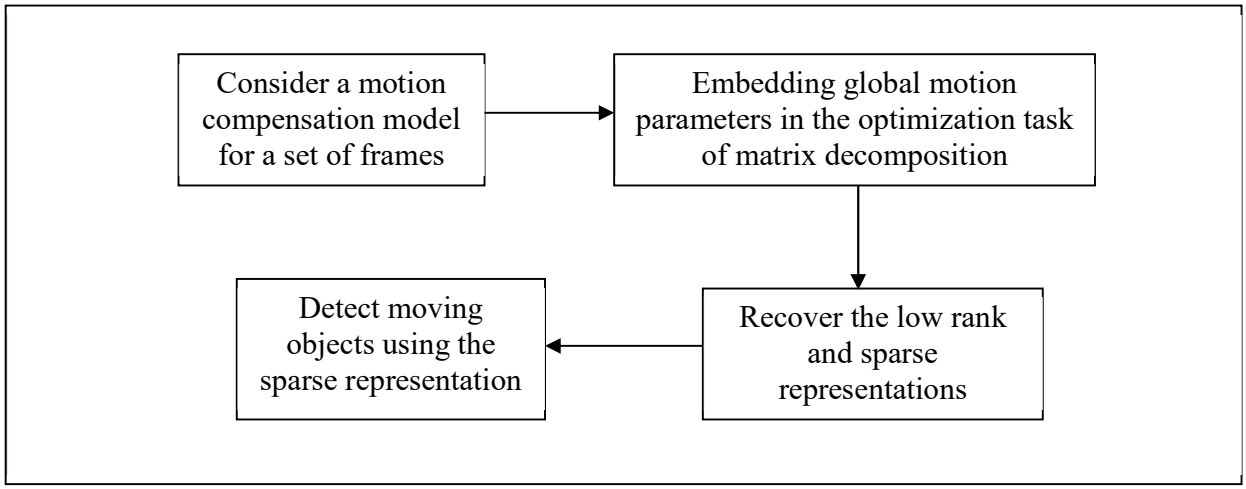

Fig. 14: General concept of low rank and sparse decomposition technique for moving camera

To apply this technique, instead of directly decomposing the observation matrix A, we use its aligned background by applying the transformation matrix τ containing the global motion of the camera and then decompose the transformed observation matrix $A°\tau$ into low rank matrix L and sparse matrix S (i.e. $A°\tau = L + S$). To find L, S and τ, the bellow energy function should be minimized.

$$\min_{L,S,\tau} \|A°\tau - L - S\|_F + \lambda\|S\|_1 \text{ w.r.t. } A°\tau - L - S = 0$$

Where $\|\cdot\|_F$ is the Frobenius norm. To compute the parameters in the energy function, this minimization task can be divided into three sub-optimizations and an iterative solution can be attained.

$$\tau^t = \min_{\tau} \|A°\tau - L^{t-1} - S^{t-1}\|_F^2$$

$$L^t = \min_{rank(L)\ll k} \|A°\tau^t - L - S^{t-1}\|_F^2$$

$$S^t = \min_{S} \|A°\tau^t - L^t - S\|_F^2 + \lambda\,\|S\|_1$$

τ can be computed by WLSM (Weighted Least Squares Minimization), L by SVD (Singular Value Decomposition) and S iteratively by above equations. Meanwhile, some methods have recently tried to produce better results by modified this general concept. Zhou *et al.* [18] utilized a different formula for the optimization by considering the noise in image sequences.

$$\min_{L,S,\tau} \frac{1}{2} \|P_{S^{\perp}}(A^{\circ}\tau - L)\|_F + \alpha\|L\|_* + \beta\|S\|_1 + \gamma\|Dvec(S)\|_1 \text{ w.r.t. } A^{\circ}\tau = L + S + \varepsilon$$

Where $P_{S^{\perp}}(x)$ is the complementary orthogonal projection of the matrix x onto the linear space of matrices supported by *S*. α, β and γ are the controlling parameters, *D* is the nod-edge incidence matrix of a graph of all pixels in the sequence and detected edges and ε denotes Gaussian noise.

In order to reduce the false detections, Chen *et al.* [142] used a spatio-temporal coherency of consecutive frames in the optimization formula as bellow.

$$(L^*, S^*) = \min_{L,S,\Delta\tau} (\|L\|_* + \lambda\|S\|_1) \text{ w.r.t. } A^{\circ}\tau + \sum J_i \Delta\tau_i \varepsilon_i^T = L + S$$

Where *Ji* is the Jacobian of the *i*th background patch and the 2D transform. τ_i and ε_i denote the standard basis. They served the low rank decomposition technique differently for moving object detection from a moving camera. They use low rank prior obtained from the last set of frames to track the background in the next set of frames of a sequence using low rank coherency. When the backgrounds were aligned for the set of frames, The Robust PCA is used to decompose the aligned observation matrix into low rank and sparse matrices, where we can extract the moving objects and low rank background prior for the next set of frames.

Ebadi *et al.* [23] used Block-PCP assuming that the zero elements of the matrix *S* appear in block by applying this formula:

$$\min_{rank(L) \ll k,S,\tau} \|A^{\circ}\tau - L - S\|_F + \lambda \sum \left\|mat(S_j)\right\|_{2,1} \text{ w.r.t. } A^{\circ}\tau = L + S$$

Where $\ell_{2,1}$-norm is the ℓ_1-norm of the vector formed by tacking the ℓ_2-norms of the columns of the considered matrix. $mat(\cdot)$ is a mapping operator transferring the underlying vector into a matrix.

Rodriguez and Wohlberg [24] proposed a fully incremental PCP algorithm for video background modelling while being able to handle camera jitter. They incrementally solve this optimization formula.

$$\min_{L^*,S,\mathcal{T}} \frac{1}{2} \|A - \mathcal{T}(L^*) - S\|_F + \lambda\|S\|_1 \text{ w.r.t. } A = \mathcal{T}(L^*) + S$$

Where L* is the aligned low rank representation and $\mathcal{T}$ is a set of rigid transformations. Later, Chau and Rodríguez [143] modified the algorithm in [24] and proposed an algorithm that continuously estimates the transformation $\mathcal{T}$ for more efficiently aligning previous L* with new observed frames. Their algorithm can better handle panning and camera motions in detection moving objects. Recently, Gao *et al.* [144] addressed the moving object detection in the presence of moving cameras and noise using a new robust PCA. In their work, first a robust registration is performed among frames of a video sequence and then a parametric low rank background component is created using optimal low rank matrix estimator [145]. Next, foreground and noise are decoupled from sparse component using a total variation based regularization model. This framework can effectively detect moving objects in the case of moving cameras. More recently, Thomas *et al.* [146] provided a low-rank representation of a target by calculating the union of subspaces taking all frames of a video into account and also by computing the sparse residue from the reference video in the case of low motion of camera. Although its computational complexity is high, their algorithm can well detect moving objects in a complex background.

In general, these approaches are quite effective to detect moving objects in a video sequence captured by a moving camera, however, they need that a certain pre-defined number of frames are collected before applying these algorithms, so they cannot be suitable for real-time applications. Moreover, they use an optimization phase which is complex and usually needs to be relaxed with certain constraints.

3.4. Object tracking

The aim of object tracking is to associate different regions belonging to the same object in consecutive frames of a video sequence. In other words, tracking is the localization of a moving object in frames of the sequence, so we can consider it as a process of moving object detection. The general concept of this technique is shown in Fig. 15. First, a number of connected pixels in the first image of the sequence are marked as the desired object (target). We can represent the target

by a bounding box, a silhouette, a contour, a skeleton, or simply by a center point. To do that, an operator can manually select the target or we can use an object detection method from single still image which can be based on saliency detection or object segmentation and recognition. Then, appearance features are extracted from the target by using the neighboring pixels in the case of center point representation or by using the interior pixels in other representations (i.e., bonding box, silhouette, etc.). The extracted appearance features can be colors, texture, edges, geometric information, frequency coefficients, simply the pixel gray values, or a combination of all of them which form a feature space [147,148,149,150,151]. Other features such as colour histogram [152] and histogram of oriented gradients (HOG) [153] can also be used for appearance modelling. The aim of HOG is to describe a moving object by a set of local histogram. These histograms count occurrence of gradient orientation in a local part of the object. Although HOG provides good local information of an object, it is sensitive to illumination changes. Various other kinds of local features such as GLOH [154], SURF [155] and SIFT [156] can be used to represent a moving object. A comparison of local feature descriptors can be found in [157,158]. SIFT (shift invariant feature transform) is especially more desired since it generates local features which are robust to changes in scale, noise, illumination and local geometric distortion [159]. To produce this feature, some keypoints should be detected. The performance of SIFT for object tracking deteriorates as false matched keypoints increase and it is not suitable for small object tracking. Appearance features can be also updated in time using convolutional neural network (CNN) [160]. CNNs can be also trained to learn descriptors encoding local spatial-temporal features [161]. Indeed, CNN aims to learn features in an adaptive, hierarchical and distributed representation way. Such a representation can be robust to significant appearance variations [162].

For the aim of matching, statistical models for the feature space can also be used. Moreover, subspace techniques such as Principal Component Analysis (PCA) or Independent Component Analysis (ICA) can be used to reduce the high dimensional feature space. The choice of object representation and appearance features is highly dependent on application domain. Next, a matching strategy should be used to determine the location of the target in the next frame. It is

based on serving a similarity/correlation function which determines which candidate in the next frame is the most similar to the target based on appearance features extracted from candidates and the target. Once the target is detected in the next frame, the target's features will be updated and the procedure will be repeated for the next frames.

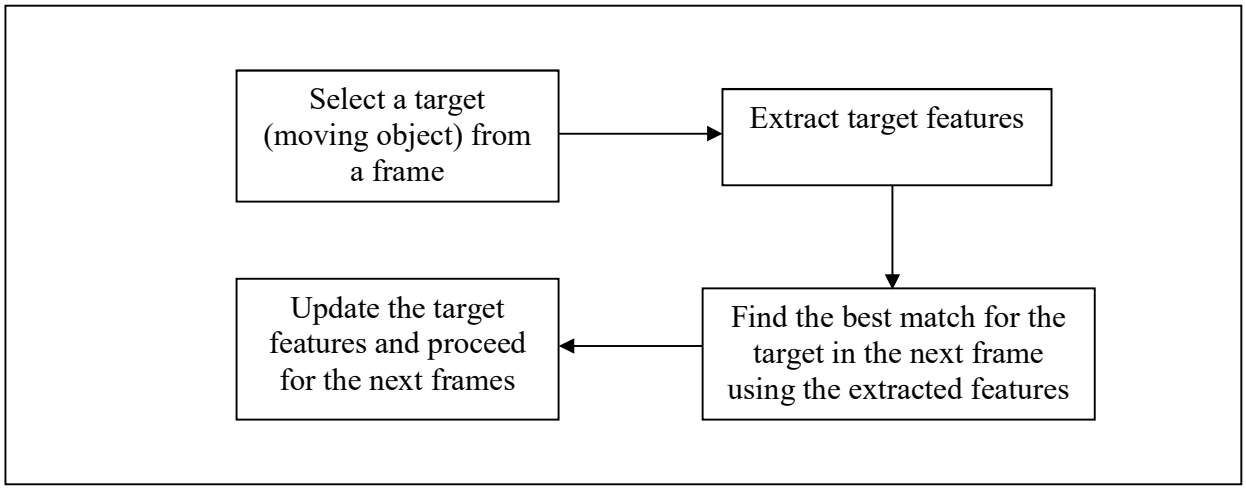

Fig. 15: General concept of object tracking for moving camera

Various methods have been proposed in each step of general scheme mentioned in Fig. 4. For the moving object selection step, in many applications the region of the desired object in the first image of the sequence is marked manually as a bounding box by a user [12]. Meanwhile, other methods used automatic object detection algorithms to find the moving objects in the first image of the sequence [12]. Background subtraction methods such as Gaussian Mixture Models have been widely used for moving object detection [163]. However, these methods fail in the presence of complex background. Optical flow is also used to find the coherent motion vectors which lead to detect moving objects [164], but it is highly sensitive to noise and needs high computational time. Bagherzadeh and Yazdi [165] used an object detection technique based on saliency map which is enough accurate to localize the objects but does not give the accuracy in their boundary.

In the feature extraction step, color, texture, silhouette, contour and shape are usually used in many methods. Color is relatively invariant to pose changes and simple to be computed whereas texture

is very sensitive to noise and difficult to be represented and these features are often applied for simple rigid objects. Silhouette, contour and shape are applied for complex non-rigid objects but are very sensitive to accurate determination of the region of the object.

In the matching step, some methods use the statistics of appearance features (color and texture) of the object for the matching [166]. Template based approaches use simple geometric shape, contour or silhouette of the object for the matching [167]. Because a simple template matching fails when the objects change their pose, some other methods use a deformable template matching by applying the parameterized transformations such as affine [168]. Sometimes, it is suitable to combine appearance features with shape representation for tracking. A hybrid template matching is also used by Liu *et al.* [169].

Some works have used different strategies for moving object tracking in moving camera. Comaniciu *et al.* [170] used a kernel based method named mean shift to create the histogram of a moving object as an appearance feature and a similarity measurement based on Bhattacharyya distance to find the best match for the moving object in the next frames. In [171], a learning model to classify pixel blocks containing the object was used and then the blocks already classified have been served to learn the object behavior in the next frames. Bagherzadeh, and Yazdi [172] used the saliency map to extract the appearance features in frequency domain and then applied the regularized least squared classifier to classify pixels belonging to a moving object. Another popular strategy from moving object tracking in moving camera is tracking-by-detection [173,174,175]. This tracking strategy, involves repetitively applying a detection algorithm in images of video sequence and then using a matching technique to associate the detected objects in consecutive frames. This strategy works well in challenging problems such as uncalibrated moving cameras, background changing, and especially occlusion. However, matching step or object association is difficult and the obtained results are a discrete set of responses and usually yield false positives and missing detections. However, some recent works [176,177] use information from future frames to better detect moving objects in current frame.

Although the currently used hand-craft features such as texture, color, intensity and shape for moving object detection and tracking in presence of moving camera produce good results, new trends are toward using more descriptive features. Indeed, it is more effective to exploit target specific representations through a learning process rather than using a fixed set of pre-defined features [178]. Deep neural networks, especially convolutional neural networks (CNN) have recently proposed for moving object tracking which effectively use category specific features for tracking and show some promising results even in the case of complex moving camera [179,180,181]. CNNs have been used for tracking usually in two strategies; one is to use them as a feature extractor incorporating with a good classifier [182] and the other is to use a unified deep structure for object tracking [183]. In [182], Wang and Ying firstly proposed deep learning tracker (DLT) to offline learn genetic features from auxiliary natural images. Although their work performed well, when a significant temporal changes of a moving object occur the approach fails to effectively learn these changes. In [184], Wang *et al.* Improved CNN architecture to learn hierarchical features for model-free object tracking which can handle temporal variations and do efficiently online tracking. Next, Wang *et al.* [185] took advantage of CNN features extracted from different layers and used a selection method by adding two pre-defined convolutional layers to filter out noisy, irrelevant or redundant features. Zhai *et al.* [186] used a Bayesian classifier as a loss layer in CNN tracker and updated the network parameters in online tracker. By doing this, appearance variations of a moving object over time were taken into account. More recently, an exhaustive comparison of tracking methods based on deep learning has been done in [187]. In general, the CNNs for tracking is trained in a simple an effective way which provides good features for object tracking. However, training of a robust CNN requires a considerable large number of annotated samplers and learning process is very time consuming.

In general, object tracking methods work very well for complex motions of the camera where other strategies for moving object detection fail to be robust and accurate. However, these methods highly depend on well initial selection of the moving object in the first frames of the sequence. In addition, they do not provide accurate information on the silhouette of tracked moving object in

the sequence. Meanwhile, multiple object tracking is also a challenging problem especially for freely camera motion.

In Table 3, we provide a general comparison of four mentioned categories and present main improving methods in each category.

Table 3. Comparative study on different moving object detection methods

Category	Negative points	Positive points	Representative works
Background modelling	• Not good for freely camera motion • Accuracy is highly dependent on background model	• Moderate in complexity • Good for real time applications • Providing good object's silhouette	Zivkovic and Van der Heijden [123] (2006) Yi *et al.* [124] (2013) Cuevas *et al.* [125] (2015) Setyawan *et al.* [126] (2014) Unger *et al.* [75] (2008) Jin *et al.* [127] (2008) Lenz *et al.* [128] (2011) Minematsu *et al.* [129] (2015) Viswanath and Behera [130] 2015) Maddalena and Petrosino [131] (2008) Zhu and Elgammal [132] (2017) Zhou and Maskell [133] (2017) Wu *et al.* [134] (2017) Gong *et al.* [135] (2017) Minaeian *et al.* [121] (2018) Babaee *et al.* [137] (2017)
Trajectory classification	• Very sensible to noise • Providing no information on object's silhouette • Accuracy is highly dependent on motion tracker model	• Providing good object's trajectory over time • Moderate in complexity	Sheikh *et al.* [138] (2009) Brox and Malik [15] (2010) Yin *et al.* [16] (2015) Singh *et al.* [139] (2017) Zhang *et al.* [140] (2017)
Low rank and sparse representation	• require a collection of frames • Not suitable for real-time applications • High in complexity	• Good accuracy	Zhou *et al.* [18] (2013) Chen *et al.* [142] (2016) Ebadi *et al.* [23] (2015) Rodriguez and Wohlberg [24] (2015)

		• Providing good object's silhouette	Chau and Rodríguez [143] (2017) Gao *et al.* [144] (2017)
Object tracking	• Providing no information on object's silhouette • Require good initial selection of the object	• Performing well for all camera motion • Moderate in complexity	Stauffer and Grimson [163] (1999) Chauhan and Krishan [164] (2013) Bagherzadeh and Yazdi [165] (2014) Liu *et al.* [169] (2011) Comaniciu *et al.* [170] (2003) Babenko *et al.* [171] (2011) Bagherzadeh and Yazdi [172] (2015) Breitenstein *et al.* [173] (2009) Chen *et al.* [176] (2017) Wang and Yung [182] (2013) Wang *et al.* [183] (2016) Zhai *et al.* [186] (2016)

4. Video databases for moving cameras

Usually, the performance of a new moving object detection algorithm should be evaluated and compared with that of the state of the art methods using same databases. Although many benchmark datasets for evaluation of moving object detection algorithms for fixed cameras are available and can be easily found in the Internet [136,188,188,190,191], a few of those for moving camera are available and hereby some authors have preferred to report their comparative results based on using their own dataset. Here, we introduce some available benchmark datasets for moving camera which commonly used by some authors.

- **Hopkins dataset**: it includes 155 video sequences introduced originally by Tron and Vidal [192] for testing motion segmentation algorithms. Most of sequences contain camera motions like rotation and translation and moving objects are mostly people and cars. Some of the sequences were manually annotated ground truth for a subset of frames provided by

Brox and Malik [15]. A characteristic of most sequences is that they are short and the objects are always in movement. This dataset has been used in [193,194,138,195].

- **PV dataset**: It is provides by [196] for the goal of evaluating motion estimation algorithms. It contains video sequences for moving cameras, mostly rotation and translation, and the moving objects are multiple cars and persons. It was used by many authors for comparing moving camera background subtraction algorithms [138,197,193,194].
- **Smartphone dataset**: It was recently introduced by Zamalieva and Yilmaz [198] for testing moving object detection algorithms for the sequences captured by smartphone cameras. It contains a set of image sequences taken by a smartphone camera in rotation and translation. Multiple moving objects were captured with challenges such as depth variations, shadows and reflections.
- **CDnet** [136]: This dataset was proposed originally for comparing the performance of change detection algorithms mostly for fixed cameras, however four videos (one indoor and three outdoor) captured by vibrating cameras with corresponding ground-truth are commonly used for evaluating algorithms for the case of camera jitter. A new corresponding expanded dataset including many PTZ sequences was also presented in [199].
- **Further datasets**: Apart from the datasets mentioned above, there are a large number of datasets publicly available in the Internet, in which some are related to the main topics handled in this paper. Here, we list some of these sites and links.
    - CVonline [200]: This site contains links to different datasets in various fields of image processing. The sequences are categorized by their application (e.g. action recognition, human activities, medical, etc.) or their contents (e.g. fingerprints, face, objects, etc.). It includes in total 407 links to different dataset divided into 19 categories.
    - MOT Benchmark [201]: The provided datasets by this site are sequences mostly in unconstrained environments filmed with both static and moving cameras. Tracking

and evaluation are done in image coordinates. All sequences have been annotated with high accuracy, strictly following a well-defined protocol. For some others, camera calibration is available, enabling tracking in world coordinates.

- CVL [202]: This site provides some datasets in various fields of image processing. Some of sequences contain walking pedestrians captured by moving cameras.

Recently, Dubuisson and Gonzales [203] provided an excellent review on datasets available for visual tracking. The authors categorized the datasets based on their difficulties and specific applications. Some of the introduced datasets contain the video sequences captured by moving cameras.

5. Performance metrics

In order to evaluate and compare various moving object detection algorithms for moving camera, quantitative metrics which are faire and consistent should be used. Meanwhile, to apply these metrics, the established ground truth of the sequences on which the experiments have been done should be available. Contrary to fixed camera case, a few ground truths are available for databases for moving camera and due to this unavailability, some authors provided the qualitative comparisons based on visual aspects [204,205].

For moving object detection methods such as background modelling and low rank matrix decomposition, the results are typically binary images corresponding to sequence frames where white pixels represent the detected moving objects. Therefore, related to established ground truth frames, the usual metrics for the binary classification can be applied which are based on following parameters.

- **True Positive (TP)**: also known as hit, is the number of moving object detected pixels corresponding to detected pixels in the ground truth.
- **False Positive (FP)**: also known as false alarm, is the number of moving object detected pixels corresponding to non-detected pixels in the ground truth.

- **True Negative (TN)**: also known as correct rejection, is the number of moving object non-detected pixels corresponding to non-detected pixels in the ground truth.
- **False Negative (FN)**: also known as miss, is the number of moving object non-detected pixels corresponding to detected pixels in the ground truth.

Hereby, three most utilized metrics are driven as follows:

- **Precision** (also known as positive predictive value) measures the percentage of all detected pixels which belongs to moving object.

$$Precision = \frac{TP}{TP + FP}$$

- **Recall** (also known as sensitivity) measures the percentage of all pixels belonging to moving object which is correctly detected.

$$Recall = \frac{TP}{TP + FN}$$

- **F1-measure** (also known as F1-score) measures the weighted average of the Precision and Recall.

$$F1 - measure = 2\frac{Precision \times Recall}{Precision + Recall}$$

Other related metrics used also by some authors are as follows.

- **Specificity** (also known as true negative rate) measures the percentage of all background pixels which is correctly non-detected.

$$Specificity = \frac{TN}{TN + FP}$$

- **Accuracy** measures the percentage of all image pixels which is correctly detected and rejected.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}$$

For moving object tracking methods which represent object by bounding box, other metrics are used as follows

- **Corloc** (correct localization) is defined as the percentage of frames correctly localized according to the PASCAL criterion [206];

$$\frac{area(b_P \cap b_{gt})}{area(b_P \cup b_{gt})} > 0.5$$

  Where b_p is the predicted box and b_{gt} is the ground-truth box. Corloc is usually reported in percentages.
- **IOU** (intersection over union) is simply the ratio of overlap area of predicted box and ground-truth box over their union area.

Some authors prefer to use the curve presentation to show the performance of algorithms. Two most utilized curves are Receive Operating Characteristic (**ROC**) curve and Precision-Recall (**PR**) curve. ROC curve is created by plotting the Recall against the false positive rate (i.e. 1-Specificity). ROC curve shows indeed how well the algorithm can be expected to achieve in general the detection of moving objects for different datasets. PR curve is created by plotting the Precision against the Recall. PR curve is more useful in practice for the problems, where the detected pixels (true or false) are more interesting than non-detected ones (true or false). PR curve shows indeed how meaningful is to obtain positive results by applying the algorithm for the given dataset.

For the methods such as trajectory classification and object tracking, the obtained results are typically trajectories or tracks indicating the position of central point of detected moving object region or its bounding box in each frame of the sequence. If the ground truth is available, the performance evaluation consists of using above mentioned metrics to determine how well the trajectory or track points are in match with their corresponding ground truth. However, in a rigorous manner, it is preferable to combine different aspects of trajectory classification or tracking performance (e.g. timeliness, trajectory accuracy, continuity, data association, false detection, etc.) into a single metric. Based on their preliminary works done in [207], Ristic *et al.* recently proposed a consistence metric for measuring the distance between the established ground truth track and the detected track obtained by using any tracker algorithm [208]. Some authors added other performance metrics to evaluate moving object detection algorithms. For example, Nascimento and Marques [209] proposed a framework to evaluate the moving object detection algorithms by

considering the ambiguous situations such as region splitting and merging to bring a better interpretation of false detections.

The essential issue in performance evaluation, which still arises for the case of moving camera, is how we can generate the ground truth for large datasets of video. Although many tools for grand truth generation have been proposed [210,211,212] for the case of video surveillance datasets, where the camera is fixed, and some of them may be extended to the datasets for moving cameras, this issue is still considered an important open problem with grand challenge.

6. Concluding remarks

In this paper, various aspects of moving object detection with focusing on moving camera have been studied. We have divided the moving object detection methods into four broad categories based on the technique they have used, namely, methods modelling the background, methods classifying trajectories, methods based on matrix decomposition and methods tracking moving objects. We have provided extensive reviews on difficulties and challenging tasks, real-time concern of moving object detection methods and main methods tried to improve the techniques in each category. Moreover, we have introduced benchmark datasets available for moving camera case and the common performance metrics used in most related works to compare various moving object detection algorithms.

Camera motion is recently very common in real-world and many videos are taken by unconstrained movement of cameras. So, the moving object detection algorithms are facing with more challenging difficulties and require more improvements in existing methods like using combinational features or trying newest strategies like deep convolutional neural networks. Here we elaborate some trends in each aforementioned categories for moving object detection in the presence of moving cameras.

- Modelling based background subtraction: The methods in this category are more popular due to their simplicity while having good performance in real time applications. The basic concept in this category is to create a good background

model even in the presence of freely moving camera which is a real challenge in this category. New trends are toward maintaining an up-to-date model of the background using multi model distributions (e.g., [213]) and by perfectly estimating camera motion in every instance of the video using multi transformational models (e.g., [214]). Meanwhile, deep neural networks have provided promising results in moving object detection in videos [215], and it seems they might also be successful for moving cameras.

- Trajectory classification: New trends in this category are toward using more efficient keypoint detectors and feature descriptors (e.g., 3D-SIFT [216], HOG3D [217] and Extended SURF [218]) for tracking feature points and using their trajectories as cues for further video analysis such as the action recognition. However, camera motion, especially freely motion, still poses a significant challenge in this category.
- Low rank and sparse matrix decomposition: The methods in this category have provided excellent results for moving object detection in presence of static cameras. However, real challenge is when unconstrained movements of the camera is involved. New trends are toward integrating a better camera motion model into the optimization constraint for low rank and sparse decomposition.
- Object tracking: The methods in this category can better handle camera motion, meanwhile, new trends are toward applications with more challenging scenarios. In long time moving object tracking the methods try to incorporate learning strategies and motion prediction models. Moreover, multi-object tracking for moving cameras is another application that rises recently many attentions by exploiting multiple statistical models or using more complex prediction filtering models. Deep neural networks have been also incorporated into visual tracking due to availability of enormous annotated datasets for their training.

We can summarize that the accurate prediction of camera motion plays an important role in any moving object detection algorithm. We are confident that this survey bring rich information about different aspects of moving object detection in the case of moving camera and can help and encourage all researchers willing to work in this topic.

7. Acknowledgements

The authors would like to acknowledge the Region Poitou-Charente of France and Shiraz University which provided funds for this work and the stay of Prof. Mehran Yazdi at the Lab. MIA of the Univ. La Rochelle.

8. References

[1] S. Wu, O. Oreifej, and M. Shah, Action recognition in videos acquired by a moving camera using motion decomposition of Lagrangian particle trajectories, IEEE International Conference on Computer Vision, (Nov. 2011), 1419-1426.

[2] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, Statistic and knowledge-based moving object detection in traffic scenes, IEEE Intelligent Transportation Systems, (2000), 27-32.

[3] E. N. Malamas, E. G. Petrakis, M. Zervakis, L. Petit, and J. D. Legat, A survey on industrial vision systems, applications and tools, Image and vision computing, 21(2) (2003), 171-188.

[4] W. Hu, T. Tan, L. Wang, and S. Maybank, A survey on visual surveillance of object motion and behaviors, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 34(3), (2004), 334-352.

[5] J. S. Kim, D. H. Yeom, and Y. H. Joo, Fast and robust algorithm of tracking multiple moving objects for intelligent video surveillance systems, IEEE Transactions on Consumer Electronics, 57(3), (2011), 1165-1170.

[6] J. Yang, X. Xie, and Y. Wang, Design of video surveillance and tracking system based on attitude and heading reference system and PTZ camera, In AIP Conference Proceedings, 1834(1), (2017), 040016…..

[7] P. Chen, Y. Dang, R. Liang, W. Zhu, and X. He, Real-Time Object Tracking on a Drone With Multi-Inertial Sensing Data, IEEE Transactions on Intelligent Transportation Systems, 19(1), (2018), 131-139.

[8] P. Dames, P. Tokekar, and V. Kumar, Detecting, localizing, and tracking an unknown number of moving targets using a team of mobile robots, The International Journal of Robotics Research, 36(13-14), (2017), 1540-1553.

[9] B. Risse, M. Mangan, L. Del Pero, and B. Webb, Visual Tracking of Small Animals in Cluttered Natural Environments Using a Freely Moving Camera, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2017), 2840-2849.

[10] L. Leal-Taixé, A. Milan, K. Schindler, D. Cremers, I. Reid, and S. Roth, Tracking the trackers: an analysis of the state of the art in multiple object tracking, arXiv preprint arXiv:1704.02781, (2017).

[11] T. Bouwmans, Traditional and recent approaches in background modelling for foreground detection: An overview, Computer Science Review, 11, (2014), 31-66.

[12] A. Yilmaz, O. Javed, and M. Shah, Object tracking: A survey, ACM Computing Surveys (CSUR), 38(4), (2006), p.13.

[13] C.R. Wren, A. Azarbayejani, T. Darrell, and A.P. Pentland, Pfinder:Real-time tracking of the human body, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), (1997), 780–785.

[14] P. Delagnes, J. Benois, and D. Barba, Active contours approach to object tracking in image sequences with complex background, Pattern Recognition Letters, 16(2), (1995) 171-178.

[15] T. Brox and J. Malik, Object Segmentation by Long Term Analysis of Point Trajectories, European conference on computer vision (ECCV), (2010), 282-295.

[16] X. Yin, B. Wang, W. Li, Y. Liu, and M. Zhang, Background subtraction for moving camera based on trajectory-controlled segmentation and label inference, KSII Trans. on Internet and Information Systems, 9(10), (2015), 4092-4107.

[17] T. Bouwmans and E. Zahzah, Robust PCA via Principal Component Pursuit: A Review for a Comparative Evaluation in Video Surveillance, Special Issue on Background Models Challenge, Computer Vision and Image Understanding, (CVIU), 122, (2014), 22–34.

[18] X. Zhou, C. Yang, and W. Yu, Moving object detection by detecting contiguous outliers in the low-rank representation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(3), (2013), 597-610.

[19] T. Bouwmans, A. Sobral, S. Javed, S. Jung, and E. Zahzah, Decomposition into Low-rank plus Additive Matrices for Background/Foreground Separation: A Review for a Comparative Evaluation with a Large-Scale Dataset, Computer Science Review, 23, (2017), 1-71.

[20] T. Bouwmans, N. Aybat, and E. Zahzah, Handbook on Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing, CRC Press, Taylor and Francis Group, (May 2016).

[21] S. Javed, A. Sobral, T. Bouwmans, and S. Jung, OR-PCA with Dynamic Feature Selection for

Robust Background Subtraction, ACM Symposium On Applied Computing, (SAC), (2015), 86-91.

[22] A. Sobral, T. Bouwmans, and E. Zahzah, Double-constrained RPCA based on Saliency Maps for Foreground Detection in Automated Maritime Surveillance, (ISBC), Workshop conjunction with (AVSS), (2015), 1-6.

[23] S. E. Ebadi, V. G. Ones, and E. Izquierdo, Efficient Background subtraction with low-rank and sparse matrix decomposition, IEEE International Conference on Image Processing, (ICIP), (Sept. 2015), 4863-4867.

[24] P. Rodriguez and B. Wohlberg, Translational and rotational jitter invariant incremental principal component pursuit for video background modelling, IEEE International Conference on Image Processing, (ICIP), (Sept. 2015), 537-541.

[25] S. Wu, T. Zhao, C. Broaddus, C. Yang, and M. Aggarwal, Robust pan, tilt and zoom estimation for PTZ camera by using meta data and/or frame-to-frame correspondences, 9th International Conference on Control, Automation, Robotics and Vision, (ICARCV), (2006), 1-7.

[26] Y. Wu, X. He, and T. Q. Nguyen, Moving Object Detection With a Freely Moving Camera via Background Motion Subtraction, IEEE Transactions on Circuits and Systems for Video Technology, 27(2), (2017), 236-248.

[27] K. A. Kinjal and D. G. Thakore, A survey on moving object detection and tracking in video surveillance system, International Journal of Soft Computing and Engineering (IJSCE), 2(3), (2012), 2231-2307.

[28] A. Ramya and P. Raviraj, A survey and comparative analysis of moving object detection and tracking, International Journal of Engineering Research & Technology (IJERT), 2(10), (2013), 3616-3621.

[29] T. Bouwmans, Recent advanced statistical background modelling for foreground detection: A systematic survey, In Recent Patents on Computer Science, 4, (2011), 147-176.

[30] A. Sobral and A. Vacavant, A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos, Computer Vision and Image Understanding, 122, (May 2014), 4-21.

[31] V. Sharma, N. Nain, and T. Badal, A survey on moving object detection methods in video surveillance, International Bulletin of mathematical research, 2(1), (2015), 2019-218.

[32] S. Shantaiya, K. Verma, and K. Mehta, A survey on approaches of object detection, International journal of Computer Applications, 65(18), (2013), 14-20.

[33] B. Deori and D. M. Thounaojam, A survey on moving object tracking in videos, International Journal of Information Theory, 3(3), (2014), 31-46.

[34] H. S. Parekh, D. G. Thakore, and U. K. Jaliya, A survey on object detection and tracking methods, International Journal of Innovative Research in Computer and Communication Engineering, 2(2), (2014), 2970-2978.

[35] V. A. Sanap, M. B. Kadu, and R. P. Labade, Survey on moving object detection, International Journal of Modern Trends in Engineering and Research, 2(11), (2015), 285-289.

[36] D. Shahre and R. Shende, A survey on moving object detection in static and dynamic background for automated video analysis, International Journal for Scientific Research & Development, 1(10), (2013), 2050-2054.

[37] E. H. Adelson, On seeing stuff: the perception of materials by humans and machines, In Human Vision and Electronic Imaging, 6(4299), (2001), 1-12.

[38] M. J. Black and A. D. Jepson, Eigentracking: Robust matching and tracking of articulated objects using a view-based representation, International Journal of Computer Vision, 26(1), (1998), 63-84.

[39] Incremental learning for robust visual tracking project website. http://www.cs.utoronto.ca/~dross/ivt/ (2007)

[40] Y. B. Lee, B. J. You, and S. W. Lee, A real-time color-based object tracking robust to irregular illumination variations, IEEE International Conference on Robotics and Automation, (2001), 1659-1664.

[41] C. Shen, X. Lin, and Y. Shi, Moving object tracking under varying illumination conditions, Pattern recognition letters, 27(14), (2006), 1632-1643.

[42] M. Heikkila and M. Pietikainen, A texture-based method for modelling the background and detecting moving objects, IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(4), (2006), 657-662.

[43] F. Cogun and A. E. Cetin, Object Tracking under Illumination Variations using 2D-Cepstrum Characteristics of the Target, IEEE International Workshop on Multimedia Signal Processing (MMSP), (Oct. 2010), 521-526.

[44] P. L. St-Charles, G. A. Bilodeau, and R. Bergevin, Subsense: A universal change detection method with local adaptive sensitivity, IEEE Transactions on Image Processing, 24(1), (2015), 359-373.

[45] K. Yun, J. Lim, and J. Y. Choi, Scene conditional background update for moving object detection in a moving camera, Pattern Recognition Letters, 88, (2017), 57-63.

[46] J. Lim, D. A. Ross, R. S. Lin, and M. H. Yang, Incremental learning for visual tracking, Advances in neural information processing systems, (2004), 793-800.

[47] F. Porikli, O. Tuzel, and P. Meer, Covariance tracking using model update based on lie algebra, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'), 1, (June 2006), 728-735.

[48] A. O. Balan and M. J. Black, An adaptive appearance model approach for model-based articulated object tracking, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (June 2006), 758-765.

[49] P. Tokmakov, K. Alahari, and C. Schmid, Learning motion patterns in videos, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2017), 531-539.

[50] Zdenek Kalal's website. http://personal.ee.surrey.ac.uk/Personal/Z.Kalal/ (2011)

[51] J. Kwo and K. M. Lee, Tracking of abrupt motion using Wang-Landau Monte Carlo estimation, European Conference on Computer Vision, (Oct. 2008), 387-400.

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

[53] F. Wang and M. Lu, Hamiltonian Monte Carlo estimator for abrupt motion tracking, International Conference on Pattern Recognition (ICPR), (Nov. 2012), 3066-3069.

[54] H. Zhang, J. Zhang, Q. Wu, X. Qian, T. Zhou, and F. U. Hengcheng, Extended kernel correlation filter for abrupt motion tracking, KSII Transactions on Internet & Information Systems, 11(9), (2017), 4438-4460.

[55] A. D. Jepson, D. J. Fleet, and T. F. El-Maraghi, Robust online appearance models for visual tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(10), (2003), 1296-1311.

[56] A. Yilmaz and X. Li, M. Shah, Contour-based object tracking with occlusion handling in video acquired using mobile cameras, IEEE Transactions on pattern analysis and machine intelligence, 26(11), (2004), 1531-1536.

[57] A. Senior, A. Hampapur, Y. L. Tian, L. Brown, S. Pankanti, and R. Bolle, Appearance models for occlusion handling, Image and Vision Computing, 24(11),(2006), 1233-1243.

[58] J. Pan and B. Hu, Robust occlusion handling in object tracking, IEEE Conference on Computer Vision and Pattern Recognition, (June 2007), 1-8.

[59] L. Hou, W. Wan, K. H. Lee, J. N. Hwang, G. Okopal, and J. Pitton, Robust human tracking based on DPM constrained multiple-kernel from a moving camera, Journal of Signal Processing Systems, 86(1), (2017), 27-39.

[60] Zhang dataset. http://www4.comp.polyu.edu.hk/~cslzhang/CT/CT.htm (2012)

[61] A. Monnet, A. Mittal, N. Paragios, and V. Ramesh, Background modelling and subtraction of dynamic scenes, Ninth IEEE International Conference on Computer Vision, (Oct. 2003), 1305-1312.

[62] L. Li, W. Huang, I. Y. Gu, and Q. Tian, Statistical modelling of complex backgrounds for foreground object detection, IEEE Transactions on Image Processing, 13(11), (2004), 1459-1472.

[63] D. Chetverikov and R. Péteri, A brief survey of dynamic texture description and recognition, Computer Recognition Systems, (2005), 17-26.

[64] S. R. Arashloo, M. C. Amirani, and A. Noroozi, Dynamic texture representation using a deep multi-scale convolutional network, Journal of Visual Communication and Image Representation, 43, (2017), 89-97.

[65] T. Minematsu, H. Uchiyama, A. Shimada, H. Nagahara, and R. I. Taniguchi, Adaptive background model registration for moving cameras, Pattern Recognition Letters, (2017), 86-95.

[66] A. Prati, I. Mikic, N. M. Trivedi, and R. Cucchiara, Detecting moving shadows: algorithms and evaluation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(7), (2003), 918-923.

[67] A. Sanin, C. Sanderson, and B. C. Lovell, Shadow detection: A survey and comparative evaluation of recent methods, Pattern Recognition, 45(4), (2012), 1684-1695.

[68] N. Al-Najdawi, H. E. Bez, J. Singhai, and E. A. Edirisinghe, A survey of cast shadow detection algorithms, Pattern Recognition Letters, 33(6), (2012), 752-764.

[69] A. Tiwari, P.K. Singh, and S. Amin, A survey on shadow detection and removal in images and video sequences, 6th International Conference in Cloud System and Big Data Engineering (Confluence), 2016 6th International Conference, (Jan. 2016), 518-523.

[70] H. Xia, S. Shuxiang, and H. Liping, A modified Gaussian mixture background model via spatiotemporal distribution with shadow detection, Signal, Image and Video Processing, 10(2), (2016), 343-350.

[71] R. Song, M. Liu, M. Wu, J. Wang, C. and Liu, A Shadow Elimination Algorithm Based on HSV Spatial Feature and Texture Feature, In International Conference on Emerging Internetworking, Data & Web Technologies, (2017), 585-591.

[72] Ling, H.: BLUT dataset. http://www.dabi.temple.edu/~hbling/code_data.htm#L1_Tracker (2011)

[73] A. Treptow and A. Zell, Real-time object tracking for soccer-robots without color information, Robotics and Autonomous Systems, 48(1), (2004), 41-48.

[74] C. Hua, Q. Chen, H. Wu, and T. Wada, A noise-insensitive object tracking algorithm, Asian Conference on Computer Vision, (Nov. 2007), 565-575.

[75] M. Unger, M. Asdsch, and P. Hosten, Enhanced background subtraction using global motion compensation and mosaicing, International Conference on Image Processing (ICIP), (Oct. 2008), 2708-2711.

[76] Y. Li, H. Ai, T. Yamashita, S. Lao, and M. Kawada, Tracking in low frame rate video: A cascade particle filter with discriminative observers of different life spans (PAMI), 30(10), (2008), 1728–1740.

[77] C. Spampinato, Y. H. Chen-Burger, G. Nadarajan, and R. B. Fisher, Detecting, tracking and counting fish in low quality unconstrained underwater videos (VISAPP), 2, (2008), 514-519.

[78] Y. Wu, H. Ling, J. Yu, F. Li, X. Mei, and E. Cheng, Blurred target tracking by blur-driven tracker, International Conference on Computer Vision, (Nov. 2011), 1100-1107.

[79] P. Dollár, R. Appel, S. Blondie, and P. Perona, Fast feature pyramids for object detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(8), (2014), 1532-1545.

[80] X. Zhang, W. Hu, N. Xie, H. Bao, and S. Maybank, A robust tracking system for low frame rate video. International Journal of Computer Vision, 115(3), (2015), 279-304.

[81] H. K. Chen, X. G. Zhao, S. Y. Sun, and M. Tan, PLS-CCA heterogeneous features fusion-based low-resolution human detection method for outdoor video surveillance, International Journal of Automation and Computing, 14(2), (2017), 136-146.

[82] S. Rowe and A. Blake, Statistical mosaics for tracking, Image and Vision Computing, 14(8), (1996), 549-564.

[83] Y. Ren, C. S. Chua, and Y. K. Ho, Statistical background modelling for non-stationary camera, Pattern Recognition Letters, 24(1), (2003), 183-196.

[84] D. Avola, L. Cinque, G. L. Foresti, C. Massaroni, and D. Pannone, A keypoint-based method for background modeling and foreground detection using a PTZ camera, Pattern Recognition Letters, 96, (2017), 96-105.

[85] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2003.

[86] http://jacarini.dinf.usherbrooke.ca/dataset2014/

[87] M. Irani, B. Rousso, and S. Peleg, Recovery of ego-motion using region alignment, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(3), (1997), 268-272.

[88] M. Irani and P. Anandan, A unified approach to moving object detection in 2d and 3d scenes, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(6), (1998), 577-589.

[89] H.S. Sawhney, Y. Guo, and R. Kumar, Independent motion detection in 3d scenes, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(10), (2000), 1191-1199.

[90] D. Zhou, V. Frémont, B. Quost, Y. Dai, and H. Li, Moving object detection and segmentation in urban environments from a moving platform, Image and Vision Computing, 68, (2017), 76-87.

[91] http://www.cvpapers.com/datasets.html

[92] J.Y.A. Wang and E.H. Adelson, Representing moving images with layers, IEEE Transactions on Image Processing, 3(5), (1994), 625-638.

[93] J. Xiao and M. Shah. Motion layer extraction in the presence of occlusion using graph cut, IEEE transactions on pattern analysis and machine intelligence, 27(10), (2005), 1644-1659.

[94] T. Chen and S. Lu, Object-level motion detection from moving cameras, IEEE Transactions on Circuits and Systems for Video Technology, 27(11), (2017), 2333-2343.

[95] E. Hayman and J. O. Eklundh, Staristical background subtraction for a mobile observer, IEEE International Conference on Computer Vision (ICCV), (2003), 67-74.

[96] Y. Shen, P. Guturu, T. Damarla, B. Buckles, and K. Namuduri, Video stabilization using principal component analysis and scale invariant feature transform in particle filter framework, IEEE Transactions on Consumer Electronics, 55(3), (2009), 1714-1721.

[97] T. Li-Fen,P. Qi, and Z. Si-Dong, A Moving Object Detection Method Adapted to Camera Jittering, Journal of Electronics and Information Technology, 35(8), (2014), 1914-1920.

[98] J. M. Wang, H. P. Chou, S. W. Chen, and C. S. Fuh, Video stabilization for a hand-held camera based on 3D motion model, 16th IEEE International Conference on Image Processing (ICIP), (Nov. 2009), 3477-3480.

[99] I. Koh, S. Ro, J. Kim, K. Min, and J. Chong, A novel digital image stabilization for mobile applications, IEEE International Conference on Consumer Electronics (ICCE), (2011), 209-210.

[100] B. Daubney, D. Gibson, and N. Campbell, Estimating pose of articulated objects using low-level motion, Computer Vision and Image Understanding, 116(3), (2012), 330-346.

[101] T. Schmidt, R. A. Newcombe, and D. Fox, DART: dense articulated real-time tracking with consumer depth cameras, Autonomous Robots, 39(3), (2015), 239-258.

[102] C. Lin, C. M. Pun, and G. Huang, Highly non-rigid video object tracking using segment-based object candidates. Multimedia Tools and Applications, 76(7), (2017), 9565-9586.

[103] M. Hoynck, M. Unger, and J. R. Ohm, Robust object region detection in natural video using motion estimation and region-based diffusion, Institute of Communication Engineering (IENT) RWTH Aachen University D-52056 Aachen, Germany, (2004).

[104] D. H. Parks and S. Fels, Evaluation of background subtraction algorithms with postprocessing, In Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance, (2008), 192–199.

[105] I. Kartika and S. S. Mohamed, Frame differencing with post-processing techniques for moving object detection in outdoor environment, IEEE 7th International Colloquium on Signal Processing and its Applications (CSPA), (2011), 172–176.

[106] Y. Dorai, F. Chausse, S. Gazzah, and N. E. B. Amara, Multi target tracking by linking tracklets with a convolutional neural network (VISIGRAPP), 6, (2017), 492-498.

[107] K. Fragkiadaki, P. Arbelaez, P. Felsen, and J. Malik, Learning to segment moving objects in videos, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2015), 4083-4090.

[108] Q. Zheng and M. Yang, A video stabilization method based on inter-frame image matching score, Global Journal of Computer Science and Technology, 17(1), (2017).

[109] D. Comaniciu, V. Ramesh, and P. Meer, Real time tracking of non-rigid objects using mean shift, IEEE Conference on Computer Vision and Pattern Recognition, 2, (June 2000), 142-149.

[110] P. K. T. Ponga and R. Bowden, A real time adaptive visual surveillance system for tracking low-resolution color targets in dynamically changing scenes, Image and Vision Computing, 21(10), (2003), 913-929.

[111] T. Yang, Q. Pan, J. Li, and S. Z. Li, Real-time Multiple Objects Tracking with Occlusion Handling in Dynamic Scenes, IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR), 1 (2005), 970-975.

[112] P. Heinemann, M. Plagge, A. Treptow, and A. Zell, Tracking Dynamic Objects in a RoboCup Environment-The Attempto Tübingen Robot Soccer Team, RoboCup-2003: Robot Soccer World Cup VII, Lecture Notes in Computer Science (CD-Supplement). Springer Verlag. (2003).

[113] H. Grabner, M. Grabner, and H. Bischof, Real-Time Tracking via On-line Boosting, British Machine Vision Conference (BMVC), 1(5), (2006), p. 6.

[114] J. U. Cho, S. H. Jin, J. E. Byun, H. Kang, X. D. Pham, and J. W. Jeon, A Real Time Object Tracking System Using a Particle Filter, IEEE/RSJ International Conference on Intelligent Robots and Systems, (Oct. 2006), 2822-2827.

[115] M. Shah, O. Javed, and K. Shafique, Automated Visual Surveillance in Realistic Scenarios, IEEE MultiMedia, 14(1), (2007), 30-39.

[116] C. Bibby and I. Reid, Robust Real-Time Visual Tracking using Pixel-Wise Posteriors, Proceeding of 10th European Conference on Computer Vision, (Oct. 2008), 831-844.

[117] K. Huanga, L. Wanga, T. Tana, and S. Maybankb, A real-time object detecting and tracking system for outdoor night surveillance, Journal of Pattern Recognition, 41(1), (2008), 432-444.

[118] S. X. Li, H. X. Chang, and C. F. Zhu, Adaptive pyramid mean shift for global real-time visual tracking, Journal of Image and Vision Computing, 28(3), (Mar. 2010), 424-437.

[119] M. Danelljan, F. Shahbaz Khan,M. Felsberg, and J. Van de Weijer, Adaptive Color Attributes for Real-Time Visual Tracking, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2014), 1090-1097.

[120] A. Agarwal and S. Suryavanshi, Real-Time* Multiple Object Tracking (MOT) for Autonomous Navigation, Technical report, (2017) (cs231n.stanford.edu/reports/2017/pdfs/630.pdf)

[121] S. Minaeian, J. Liu, and Y. J. Son, Effective and Efficient Detection of Moving Targets From a UAV's Camera, IEEE Transactions on Intelligent Transportation Systems, 19(2), (2018), 497-506.

[122] H. K. Galoogahi, A. Fagg, C. Huang, D. Ramanan, and S. Lucey, Need for speed: a benchmark for higher frame rate object tracking, arXiv preprint arXiv:1703.05884, (2017).

[123] Z. Zivkovic and F. Van der Heijden, Efficient adaptive density estimation per image pixel for the task of background subtraction, Pattern Recognition Letters, 27(7), (2006), 773–780.

[124] K. M. Yi, K. Yun, S. W. Kim, H. J. Chang, H. Jeong, and J. Y. Choi, Detection of moving objects with non-stationary cameras in 5.8ms: bring motion detection to your mobile device, IEEE Conference on Computer Vision and Pattern Recognition Workshop, (2013), 27-34.

[125] C. Cuevas, R. Mohedano, and N. Garcia, Statistical moving object detection for mobile devices with camera, IEEE International Conference on Consumer Electronics (ICCE), (Jan. 2015), 15-16.

[126] F. A. Setyawan, J. K. Tan, H. Kim, and S. Ishikawa, Detection of moving objects in a video captured by a moving camera using error reduction, SICE Annual Conference, Sapporo, Japan, (Sept. 2014), 347-352.

[127] Y. Jin, L. Tao, H. Di, N. I Rao, and G. Xu, Background modelling from a free-moving camera by multi-layer homography algorithm, IEEE International Conference on Image Processing (ICIP), (2008), 1572-1575.

[128] P. Lenz, J. Ziegler, A. Geiger, and M. Roser, Sparse scene flow segmentation for moving object detection in urban environment, Intelligent Vehicles Symposium (IV), (2011), 926-932.

[129] T. Minematsu, H. Uchiyama, A. Shimada, H. Nagahara, and R. Taniguchi, Evaluation of foreground detection methodology for a moving camera, Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV), (Jan. 2015), 1-4.

[130] A. Viswanath, R. K. Behera, V. Senthamilarasu, and K. Kutty, background Modelling from a moving camera, International Symposium on Computer Vision and the Internet (VisionNet), (2015), 289-296.

[131] L Maddalena and A Petrosino, A self-organizing approach to background subtraction for visual surveillance applications, IEEE Transactions on Image Processing, 17(7), (2008), 1168-1177.

[132] Y. Zhu and A. Elgammal, A Multilayer-Based Framework for Online Background Subtraction with Freely Moving Cameras, arXiv preprint arXiv:1709.01140, (2017).

[133] Y. Zhou and S. Maskell, Moving Object Detection Using Background Subtraction for a Moving Camera with Pronounced Parallax, Sensor Data Fusion: Trends, Solutions, Applications Conference (SDF), (2017).

[134] Y. Wu, X. He, and T. Q. Nguyen, Moving object detection with a freely moving camera via background motion subtraction, IEEE Transactions on Circuits and Systems for Video Technology, 27(2), (2017), 236-248.

[135] L. Gong, M. Yu, and T. Gordon, Online codebook modelling based background subtraction with a moving camera, 3rd International Conference on Frontiers of Signal Processing (ICFSP), (2017), 136-140.

[136] N. Goyette, P.M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, Changedetection. net: A new change detection benchmark dataset, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, (June 2012), 1-8 (http://wordpress-jodoin.dmi.usherb.ca/dataset2012/).

[137] M. Babaee, D. T. Dinh, and G. Rigoll, A deep convolutional neural network for background subtraction, arXiv preprint arXiv:1702.01731, (2017).

[138] Y. Sheikh, O. Javed, and T. Kanade, Background subtraction for freely moving cameras, IEEE International Conference on Computer Vision (ICCV), (Sept. 2009), 1219-1225.

[139] S. Singh, C. Arora, and C. V. Jawahar, Trajectory aligned features for first person action recognition, Pattern Recognition, 62, (2017), 45-55.

[140] S. Zhang, J. B. Huang, J. Lim, Y. Gong, J. Wang, N. Ahuja, and M. H. Yang, Tracking Persons-of-Interest via Unsupervised Representation Adaptation, arXiv preprint arXiv:1710.02139, (2017).

[141] E. J. Candès, X. Li, T. Ma, and J. Wright, Robust principal component analysis?, Journal of the ACM (JACM), 58(3), (2011), p. 11.

[142] C. Chen, S. Li, H. Qin, and A. Hao, Robust salient motion detection in non-stationary videos via novel integrated strategies of spatio-temporal coherency clues and low-rank analysis, Pattern recognition, 52, (2016), 410-432.

[143] G. Chau and O. Rodrıguez, Panning and Jitter Invariant Incremental Principal Component Pursuit for Video Background Modelling, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2017), 1844-1852.

[144] C. Gao, B. E. Moore, and R. R. Nadakuditi, Augmented robust PCA for foreground-background separation on noisy, moving camera video, arXiv preprint arXiv:1709.09328, (2017).

[145] R. R. Nadakuditi, Optshrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage, IEEE Transactions on Information Theory, 60(5), (2014), 3002-3018.

[146] L. Thomaz, E. Jardim, A. da Silva, E. da Silva, S. Netto, and H. Krim, Anomaly Detection in Moving-Camera Video Sequences using Principal Subspace Analysis, IEEE Transactions on Circuits and Systems I: Regular Papers, 65(3), (2018), 1003-1015.

[147] S. Hare, S. Golodetz, A. Saffari, V. Vineet, M. M. Cheng, S. L. Hicks, and P. H. Torr, Struck: Structured output tracking with kernels, IEEE transactions on pattern analysis and machine intelligence, 38(10), (2016), 2096-2109.

[148] K. Zhang, L. Zhang, and M. H. and Yang, Real-time compressive tracking, In European conference on computer vision, (2012), 864-877.

[149] M. Danelljan, F. Shahbaz Khan, M. Felsberg, and J. Van de Weijer, Adaptive color attributes for real-time visual tracking, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2014), 1090-1097.

[150] D. Du, H. Qi, L. Wen, Q. Tian, Q. Huang, S. and Lyu, Geometric hypergraph learning for visual tracking, IEEE transactions on cybernetics, (2017), 4182-4195.

[151] T. Bouwmans, C. Silva, C. Marghes, M. Zitouni, H. Bhaskar, and C. Frelicot, On the Role and the Importance of Features for Background Modeling and Foreground Detection, Computer Science Review, (28), (2018), 26-91.

[152] Q. Zhao, Z. Yang, H. and Tao, Differential earth mover's distance with its applications to visual tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(2), (3010), 274-287.

[153] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), (2005), 886-893.

[154] K. Mikolajczyk and C. Schmid, A performance evaluation of local descriptors, IEEE transactions on pattern analysis and machine intelligence, 27(10), (2005), 1615-1630.

[155] H. Bay, T. Tuytelaars, and L. Van Gool, SURF: Speeded up robust features, Computer vision (ECCV), (2006), 404-417.

[156] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International journal of computer vision, 60(2), (2004), 91-110.

[157] S. Maji, Comparison of Local Feature Descriptors, University of California, Berkeley, (2006).

[158] N. Roshanbin and J. Miller, A comparative study of the performance of local feature-based pattern recognition algorithms, Pattern Analysis and Applications, 20(4), (2017), 1145-1156.

[159] S. W. Ha and Y. H. Moon, Multiple object tracking using SIFT features and location matching, International Journal of Smart Home, 5(4), (2011), 17-26.

[160] C. Kim, F. Li, A. Ciptadi, and J. M. Rehg, Multiple hypothesis tracking revisited, In Proceedings of the IEEE International Conference on Computer Vision, (2015), 4696-4704.

[161] L. Leal-Taixé, C. Canton-Ferrer, and K. Schindler, Learning by tracking: Siamese CNN for robust target association, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, (2016), 33-40.

[162] C. Ma, J. B. Huang, X. Yang, and M. H. Yang, Hierarchical convolutional features for visual tracking, In Proceedings of the IEEE International Conference on Computer Vision, (2015), 3074-3082.

[163] C. Stauffer and W. E. L. Grimson, Adaptive background mixture models for real-time tracking, IEEE Conference on Computer Vision and Pattern Recognition, 2, (1999), 246-252.

[164] A. K. Chauhan and P. Krishan, Moving object tracking using Gaussian mixture model and optical flow, International Journal of Advanced Research in Computer Science and Software Engineering, 3(4), (2013), 243-246.

[165] M.A. Bagherzadeh and M. Yazdi, Fast object tracking with long-term occlusions handling in dynamic scenes, Second RSI/ISM International Conference on Robotics and Mechatronics (ICRoM), (Oct. 2014), 823-827.

[166] J. Ning, L. Zhang, D. Zhang, and C. Wu, Robust object tracking using joint color-texture histogram, International Journal of Pattern Recognition and Artificial Intelligence, 23(7), (2009), 1245-1263.

[167] J. Pan, B. Hu, and J. Q. Zhang, Robust and accurate object tracking under various types of occlusions, IEEE Transactions on Circuits and Systems for Video Technology, 18(2), (2008), 223-236.

[168] Y. Zhong, A. K. Jain, and M. P. Dubuisson-Jolly, Object Tracking Using Deformable Templates, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22(5), (2000), 544-549.

[169] X. Liu, L. Lin, S. Yan, H. Jin, and W. Jiang, Adaptive Object Tracking by Learning Hybrid Template Online, IEEE Transactions On Circuits and Systems For Video Technology, 21(11), (2011), 1588-1599.

[170] D. Comaniciu, V. Ramesh, and P. Meer, Kernel-based object tracking, IEEE Transactions on pattern analysis and machine, 25(5), (2003), 564-575.

[171] B. Babenko, M. H. Yang, and S. Belongie. Robust object tracking with online multiple instance learning, IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(8), (2011), 1619-1632.

[172] M. A. Bagherzadeh and M. Yazdi, Regularized least-square object tracking based on $\ell_{2,1}$ minimization, 3rd RSI International Conference on Robotics and Mechatronics (ICROM), (Oct. 2015), 535-539.

[173] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool, Robust tracking-by-detection using a detector confidence particle filter, IEEE 12th International Conference on Computer Vision, (2009), 1515-1522.

[174] A. Milan, K. Schindler, and S. Roth, Multi-target tracking by discrete-continuous energy minimization, IEEE transactions on pattern analysis and machine intelligence, 38(10), (2016), 2054-2068.

[175] N. Le, A. Heili, and J. M. Odobez, Long-term time-sensitive costs for CRF-based tracking by detection, In European Conference on Computer Vision, (2016), 43-51.

[176] J. Chen, H. Sheng, Y. Zhang, and Z. Xiong, Enhancing detection model for multiple hypothesis tracking, In Conf. on Computer Vision and Pattern Recognition Workshops, (2017), 2143-2152.

[177] A. Sadeghian, A. Alahi, and S. Savarese, Tracking the untrackable: Learning to track multiple cues with long-term dependencies, *arXiv preprint arXiv:1701.01909*, *4*(5), (2017), p.6.

[178] J. Kuen, K. M. Lim, and C. P. Lee, Self-taught learning of a deep invariant representation for visual tracking via temporal slowness principle, Pattern Recognition, 48(10), (2015), 2964–2982.

[179] C. Ma, J. B. Huang, X. Yang, and M. H. Yang, Hierarchical convolutional features for visual tracking, In Proceedings of the IEEE International Conference on Computer Vision, (2015), 3074-3082.

[180] H. Li, Y. Li, and F. Porikli, Deeptrack: Learning discriminative feature representations online for robust visual tracking, IEEE Transactions on Image Processing, 25(4), (2016), 1834-1848.

[181] H. Nam and B. Han, Learning multi-domain convolutional neural networks for visual tracking, IEEE Conference on Computer Vision and Pattern Recognition, (2016), 4293-4302.

[182] N. Wang and D. Y. Yeung, Learning a deep compact image representation for visual tracking, In Advances in neural information processing systems, (2013), 809-817.

[183] N, Wang, S. Li, A. Gupta, and D.Y. Yeung, Transferring rich feature hierarchies for robust visual tracking, arXiv preprint arXiv:1501.04587, (2015).

[184] L. Wang, T. Liu, G. Wang, K. L. Chan, and Q. Yang, Video tracking using learned hierarchical features, IEEE Transactions on Image Processing, 24(4), (2015), 1424-1435.

[185] L. Wang, W. Ouyang, X. Wang, and H. Lu, Visual tracking with fully convolutional networks, In Proceedings of the IEEE International Conference on Computer Vision, (2015), 3119-3127.

[186] M. Zhai, M. J. Roshtkhari, and G. Mori, Deep learning of appearance models for online object tracking, arXiv preprint arXiv:1607.02568, (2016).

[187] P. Li, D. Wang, L. Wang, and H. Lu, Deep visual tracking: Review and experimental comparison, Pattern Recognition, 76, (2018), 323-338.

[188] AVSS2007 (IEEE International Conference on Advanced Video and Signal based Surveillance): http://www.eecs.qmul.ac.uk/~andrea/avss2007_d.html/ (2016).

[189] PET2006 (PETS workshop in Conjunction with IEEE Conference on Computer Vision and Pattern Recognition) http://www.cvg.rdg.ac.uk/PETS2006/data.html/ (2016).

[190] PET2007 (PETS workshop *in Conjunction with 11th IEEE International Conference on Computer Vision*)http://www.cvg.reading.ac.uk/PETS2007/data.html/ (2016).

[191] CAVIAR project http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/ (2014).

[192] R. Tron and R. Vidal, A benchmark for the comparison of 3-D motion segmentation algorithms, IEEE Conference on Computer Vision and Pattern Recognition, (June 2007), 1-8.

[193] A. Elqursh and A. M. Elgammal, Online moving camera background subtraction, European Conference on Computer Vision, (Oct. 2012), 228-241.

[194] T. Lim, B. Han, and J. H. Han, Modelling and segmentation of floating foreground and background in videos, Pattern Recognition, 45(4), (2012), 1696-1706.

[195] S. Kwak, T. Lim, W. Nam, B. Han, and J. H. Han, Generalized background subtraction based on hybrid inference by belief propagation and Bayesian filtering, IEEE Conference on Computer Vision, (Nov. 2011), 2174-2181.

[196] P. Sand and S. Teller, Particle video: long-range motion estimation using point trajectories, IEEE Conference on Computer Vision and Pattern Recognition, (2006), 2195-2202.

[197] X. Cui, J. Huang, S. Zhang, and D. N. Metaxas, Background subtraction using low rank and group sparsity constraints, European Conference on Computer Vision, (Oct. 2012), 612-625.

[198] D. Zamalieva and A. Yilmaz, Background subtraction for the moving camera: A geometric approach, Computer Vision and Image Understanding, 127, (2014), 73-85.

[199] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, CDnet 2014: An Expanded Change Detection Benchmark Dataset, IEEE Workshop on Change Detection (CDW), (2014), 387-394 (http://wordpress-jodoin.dmi.usherb.ca/dataset2014/).

[200] http://homepages.inf.ed.ac.uk/rbf/CVonline/Imagedbase.htm/ (2016).

[201] https://motchallenge.net/ (2016).

[202] http://www.vision.ee.ethz.ch/datasets/ (2015).

[203] S. Dubuisson and C. Gonzales, A survey of datasets for visual tracking", Machine Vision and Applications, 27(1), (2016), 23-52.

[204] G. Zhang, J. Jia, W. Hua, and H. Bao, Robust bilayer segmentation and motion/depth estimation with a handheld camera, IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(3), (2011), 603-617.

[205] F. Liu and M. Gleicher, Learning color and locality cues for moving object detection and segmentation, IEEE Conference on Computer Vision and Pattern Recognition, (June 2009), 320-327.

[206] M. Everingham, S. M. Ali Eslami, L. Van Gool, C. K. L. Williams, J. Winn, and A. Zisserman, The PASCAL visual object classes challenge: A retrospective (IJCV), 111(1), 2015, 98-136.

[207] B. Ristic, B. N. Vo, and D. Clark, Performance evaluation of multi-target tracking using the OSPA metric, 13th Conference on Information Fusion (FUSION), (July 2010), 1-7.

[208] B. Ristic, B. N. Vo,D. Clark, and B. T. Vo, A metric for performance evaluation of multi-target tracking algorithms, IEEE Transactions on Signal Processing, 59(7), (2011), 3452-3457.

[209] J. C. Nascimento and J. S. Marques, Performance evaluation of object detection algorithms for video surveillance, IEEE Transactions on Multimedia, 8(4), (2006), 761-774.

[210] C. Jaynes, S. Webb, R. Steele, and Q. Xiong, An open development environment for evaluation of video surveillance systems, Third IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS), (June 2002), 32-39.

[211] D. Doermann and D. Mihalcik, Tools and techniques for video performance evaluation, 15th International Conference on Pattern Recognition, 4, (2000), 167-170.

[212] J. Black, T. Ellis, and P. Rosin, A novel method for video tracking performance evaluation, IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS), (Oct. 2003), 125-132.

[213] T. Akilan, Q. J. Wu, and Y. Yang, Fusion-based foreground enhancement for background subtraction using multivariate multi-model Gaussian distribution. Information Sciences, 430, (2018), 414-431.

[214] D. Zamalieva, A. Yilmaz, and J. W. Davis, A multi-transformational model for background subtraction with moving cameras, Computer Vision (ECCV), (2014), 817-819.

[215] A. R. Pathak, M. Pandey, S. Rautaray, and K. Pawar, Assessment of Object Detection Using Deep Convolutional Neural Networks, In Intelligent Computing and Information and Communication, (2018), 457-466.

[216] P. Scovanner, S. Ali, and M. Shah, A 3-dimensional sift descriptor and its application to action recognition, In Proceedings of the 15th ACM international conference on Multimedia, (2017), 357-360.

[217] A. Klaser, M. Marszałek, and C. Schmid, A spatio-temporal descriptor based on 3d-gradients, In BMVC 2008-19th British Machine Vision Conference, (2008), 275-276.

[218] G. Willems, T. Tuytelaars, and L. Van Gool, An efficient dense and scale-invariant spatio-temporal interest point detector, In European conference on computer vision, (2008), 650-663.