Computer Science Review 28 (2018) 157–177

Contents lists available at ScienceDirect

Computer Science Review

journal homepage: www.elsevier.com/locate/cosrev

Review article

New trends on moving object detection in video images captured by a moving camera: A survey

Mehran Yazdi [a,b,*], Thierry Bouwmans [b]

[a] Lab. of Signal and Image Proc., Faculty of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran
[b] Lab. MIA, University of La Rochelle, La Rochelle, France

ARTICLE INFO

Article history:
Received 26 October 2016
Received in revised form 29 January 2018
Accepted 2 March 2018

Keywords:
Moving object detection
Moving camera
Background subtraction
Motion compensation

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; modeling 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.

Contents

* Corresponding author at: Lab. of Signal and Image Proc., Faculty of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran.
E-mail addresses: yazdi@shirazu.ac.ir (M. Yazdi), thierry.bouwmans@univ-lr.fr (T. Bouwmans).

https://doi.org/10.1016/j.cosrev.2018.03.001
1574-0137/

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 modeling 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 modeling 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 modeling [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 [17–22] 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 [27–31], 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