J. Vis. Commun. Image R. 30 (2015) 164–180

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Moving object detection and tracking from video captured by moving camera ☆

Wu-Chih Hu [a], Chao-Ho Chen [b,*], Tsong-Yi Chen [b], Deng-Yuan Huang [c], Zong-Che Wu [d]

[a] Department of Computer Science and Information Engineering, National Penghu University of Science and Technology, Penghu, Taiwan, ROC
[b] Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan, ROC
[c] Department of Electrical Engineering, Dayeh University, Taiwan, ROC
[d] Whetron Electronics Co., Ltd, Taiwan, ROC

ARTICLE INFO

Article history:
Received 2 February 2015
Accepted 16 March 2015
Available online 1 April 2015

Keywords:
Object detection
Moving camera
Object tracking
Feature classification
Image difference
Object motion
Motion history
Ego-motion compensation

ABSTRACT

This paper presents an effective method for the detection and tracking of multiple moving objects from a video sequence captured by a moving camera without additional sensors. Moving object detection is relatively difficult for video captured by a moving camera, since camera motion and object motion are mixed. In the proposed method, the feature points in the frames are found and then classified as belonging to foreground or background features. Next, moving object regions are obtained using an integration scheme based on foreground feature points and foreground regions, which are obtained using an image difference scheme. Then, a compensation scheme based on the motion history of the continuous motion contours obtained from three consecutive frames is applied to increase the regions of moving objects. Moving objects are detected using a refinement scheme and a minimum bounding box. Finally, moving object tracking is achieved using a Kalman filter based on the center of gravity of a moving object region in the minimum bounding box. Experimental results show that the proposed method has good performance.

1. Introduction

Increasing demand for safety and security has resulted in more research on intelligent surveillance. Intelligent surveillance has a wide range of applications, such as moving object detection [1–4], object tracking [3–5], motion segmentation [6,7], object classification and identification [8], event detection [9], and behavior understanding and description [10]. The analysis and interpretation of video sequences captured by cameras is an active research field. Many applications in this field need to initially detect the moving objects in a scene. For example, video event detection requires detecting and tracking objects first, then recognizing what is happening around those tracked objects. Therefore, moving object detection and tracking play important roles in intelligent surveillance.

Moving objects can be detected based on video object segmentation, in which the foreground object (moving object) is extracted from each frame of a video sequence. Many methods have been proposed for video object segmentation. Generally, these methods can be roughly classified into two types [11,12]: background construction-based video object segmentation and foreground extraction-based video object segmentation. In background construction-based video object segmentation, the background information is first constructed, and then a video object in successive frames is obtained using background subtraction, that is, the difference between the background and the current frame. In foreground extraction-based video object segmentation, temporal information, spatial information, or temporal-spatial information is first used to obtain an initial video object, and then the video object in successive frames is obtained using motion information, change information, or other feature information.

Background construction-based video object segmentation can track fast-moving objects. Furthermore, its computational cost is low and its implementation is easy. Background modeling is necessary in background construction-based video object segmentation. It is mainly used in the background subtraction for foreground detection in various applications (such as video surveillance) to model the background and then detect the moving objects in a scene. The simplest method for background modeling is to obtain a background image without including any moving objects. Three

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.
E-mail addresses: wchu@npu.edu.tw (W.-C. Hu), thouho@cc.kuas.edu.tw (C.-H. Chen), chentso@cc.kuas.edu.tw (T.-Y. Chen), kevin@mail.dyu.edu.tw (D.-Y. Huang), zongche.wu@whetron.com.tw (Z.-C. Wu).

http://dx.doi.org/10.1016/j.jvcir.2015.03.003
1047-3203/

main conditions can ensure good results using background subtraction: the camera is fixed, the illumination is constant, and the background is static. However, the background can be affected under in some situations, such as under illumination changes. Therefore, the background representation model must be robust and adaptive.

Background modeling has been applied for the robust detection of moving objects in challenging environments. Effective methods robustly update models to deal with dynamic backgrounds and illumination changes. Traditional and recent approaches for background modeling used in background subtraction have been reviewed [13].

Generally, background subtraction involves building a scene representation, referred to as the background model, which is compared against incoming video frames to detect objects. However, the background, foreground, and camera all move in videos captured by a moving camera. Compared to fixed cameras, video captured by a moving camera is difficult to analyze since camera motion and object motion are mixed. When object segmentation is applied to such video, the shapes of moving objects fail to be effectively segmented and detected. Recently, Ferone and Maddalena [14] proposed neural background subtraction for moving object detection in video sequences captured by a pan-tilt-zoom (PTZ) camera. In this algorithm, the background model automatically adapts to the scene background variations, which can arise in a usual stationary camera setting or from PTZ camera movement. Furthermore, Hu et al. [15] proposed a method for the automatic detection of abandoned and removed objects for adaptive wide-field-of-view (FOV) surveillance based on an Internet Protocol camera on a rotational platform. In this approach, a frame-interval-based similarity measure and adaptive image-stitching-based wide-FOV image construction are used to obtain wide-FOV frames. Background modeling is built and used to detect abandoned and removed objects.

However, the center of a PTZ camera (or a camera on a rotational platform) is still fixed, unlike that of a moving camera. Therefore, background subtraction would fail for scenes captured by a moving camera. Video captured by a moving camera has a wide range of applications, such as mobile robots, self-driving cars, and vehicle video recorders. It is a challenging task to achieve accurate moving object detection and tracking for video captured by a moving camera.

Choi et al. [16] proposed a general framework for multiple-people tracking from a mobile vision platform. They used observation cues for given sensors (including depth) to obtain a three-dimensional (3D) coordinate system. Both the camera's ego-motion and people's paths are estimated to robustly determine all trajectories. Finally, the tracking is solved based on finding the maximum a posteri (MAP) solution of a posterior probability and the reversible jump Markov chain Monte Carlo (RJ-MCMC) particle filtering method. This method can robustly estimate a camera's motion from dynamic scenes and stably track people who are moving independently or interacting. However, a depth sensor is necessary, and thus its applications are limited.

Jodoin et al. [17] proposed a robust moving object detection method for both fixed- and moving-camera-captured video sequences that uses a Markov random field (MRF) to obtain label field fusion. Wang [18] used an MRF model to detect moving vehicles in various weather conditions, but this approach is limited because it is only applicable to gray-scale videos. In order to handle the spatial ambiguities of gray values, Ghosh et al. [19] proposed a region-matching-based motion estimation scheme to achieve moving object detection and tracking from video captured by a moving camera. In this approach, the fuzzy edge strength of each pixel location is incorporated in the MRF modeling, which preserves the object boundary for segmentation. The spatial segmentation problem is solved using the MAP estimation principle. Region-based estimation is used to find the moving objects in subsequent frames, where χ^2-test-based local histogram matching is used to detect the moving object in order to reduce search complexity. However, the computational cost of this approach is high, limiting it to specific offline applications. Furthermore, this approach does not yield good results when shadows or occlusion/disocclusion are present.

Hu et al. [20] proposed an effective method for the detection of moving objects for video captured by a moving camera. In this approach, feature points are first detected using a Harris corner detector, and then optical flow is used for feature matching for two consecutive frames. Next, these feature points are classified as belonging to foreground or background features with the assistance of multiple-view geometry. Then, the foreground regions are obtained based on the foreground feature points, and the image difference is calculated using affine transformation based on the background feature points. Moving object regions are obtained by merging the foreground regions and image difference. Finally, moving objects are detected based on the motion history of the continuous motion contours and refinement schemes. This approach is useful for real-time applications and requires no additional sensors. However, it works well only for the detection of slowly moving objects.

The present study proposes a method for moving object detection and tracking from video captured by a moving camera without additional sensors. This paper is the modified and extended version of our previous work [20] to greatly increase the performance. The proposed method can be useful for real-time applications and works well for the detection of fast moving objects. In the proposed method, the feature points in the frames are detected using the modified Harris corner detector proposed by Shi and Tomasi [21], and then further classified as belonging to foreground or background features with feature matching based on optical flow and the assistance of multiple-view geometry [20]. Next, the obtained foreground feature points are modified using a scheme for increasing the reliability of feature point classification. Foreground regions are obtained using an image difference scheme applied to the current frame and the previous frame transformed using a perspective transform with a homography matrix. Ego-motion compensation is proposed to obtain more reliable results from the image difference scheme. Then, regions of moving objects are obtained using an integration scheme applied to updated foreground feature points and obtained foreground regions. Furthermore, a compensation scheme based on the motion history of the continuous motion contours obtained from three consecutive frames is used to increase the regions of moving objects. Moving objects are detected using a refinement scheme and a minimum bounding box. Finally, a tracking scheme that uses a Kalman filter [22] based on the center of gravity of a moving object region in the minimum bounding box is proposed for tracking multiple moving objects. Experimental results show that the proposed method performs well for the detection and tracking of multiple moving objects in video sequences captured by a moving camera.

The rest of this paper is organized as follows. The feature point classification is described in Section 2. The moving object detection and tracking are proposed in Section 3. Section 4 presents experimental results and evaluations. Finally, the conclusions are given in Section 5.

2. Feature point classification

The proposed method has three main parts, namely the classification of feature points, the moving object detection, and the moving object tracking. A flowchart of the algorithm is shown in