Optik 124 (2013) 6081–6088

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Moving object detection based on improved VIBE and graph cut optimization

Jianfang Dou*, Jianxun Li

Department of Automation, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China

ARTICLE INFO

Article history:
Received 30 November 2012
Accepted 28 April 2013

Keywords:
Background model
Moving object detection
VIBE
Mean shift
Feature clustering

ABSTRACT

Extracting foreground moving objects from video sequences is an important task and also a hot topic in computer vision and image processing. Segmentation results can be used in many object-based video applications such as object-based video coding, content-based video retrieval, intelligent video surveillance and video-based human–computer interaction. In this paper, we present a novel moving object detection method based on improved VIBE and graph cut method from monocular video sequences. Firstly, perform moving object detection for the current frame based on improved VIBE method to extract the background and foreground information; then obtain the clusters of foreground and background respectively using mean shift clustering on the background and foreground information; Third, initialize the S/T Network with corresponding image pixels as nodes (except S/T node); calculate the data and smoothness term of graph; finally, use max flow/minimum cut to segmentation S/T network to extract the motion objects. Experimental results on indoor and outdoor videos demonstrate the efficiency of our proposed method.

1. Introduction

Moving objects detection for a static camera has been extensively studied for many years [1,2]. Moving objects detection plays a very important role in many vision applications with the purpose of subtracting interesting target area and locating the moving objects from image sequences. It is widely used in vision systems such as traffic control, video surveillance of unattended outdoor environments, video surveillance of objects, activity recognition, object tracking and behavior understanding. Accurate moving object detection is essential for the robustness of intelligent video-surveillance systems.

Background subtraction and temporal differencing are two popular approaches to segment moving objects in an image sequence under a stationary camera. Background subtraction detects moving objects in an image by evaluating the difference of pixel features of the current scene image against the reference background image, such as the Gaussian mixture model (GMM) [3]. GMM is a widely used approach because of its self learning capacity and its robustness to variations in lighting. However, it still has some shortcomings. One problem is that it does not explicitly model the spatial dependencies of neighboring background pixels' colors. Therefore, some false positive pixels will be produced in highly dynamic scenes where dynamic texture does not repeat exactly. Temporal differencing such as W4 [4] use three parameters to model each pixel of background: the maximum of pixel gray, minimum, and the mean of two maximum differences of two successive frames in a period time. However, It is very effective to accommodate environmental changes and generally can only recover partial edge shapes of moving objects. For non-stationary cameras, optical flow method can be used, which assign to every pixel a 2-D velocity vector over a sequence of images. Moving objects are then detected based on the characteristics of the velocity vectors. Optical flow methods are computationally intensive, and can only detect partial edge shapes of moving objects. To improve the accuracy of the background model and to deal with highly dynamic scenes, the spatial information is exploited. Li [5] use spatial information at feature level, such as color and gradient. This can improve the accuracy of the background model and is most suitable for the stationary background. Olivier Barnich and Marc Van Droogenbroeck proposed Vibe algorithm [6,7]. The method adopt neighboring pixels to create the background model, by comparing the background model and the current input pixel values to detect targets, the method also gives three steps to update the field to adapt to changes in the environment. It uses randomly selected old samples to approximate the color distribution of the background. As described by the author, the advantage of the randomization is that it avoids replacement of the oldest samples. However, when the background have similar color with foreground

* Corresponding author.
E-mail address: specialdays_2010@163.com (J. Dou).

0030-4026/$ – see front matter © 2013 Elsevier GmbH. All rights reserved.
http://dx.doi.org/10.1016/j.ijleo.2013.04.106

(i.e. camouflage problem), it will result in missing the foreground point and thus cause lack of accurate contour for post motion analysis such as recognition. And it also but slowed to eliminate ghost region. When the foreground object through these areas, the detection accuracy will drop; at the same time, goes up the high rate of false detection algorithm because of the ghosting region.

Recently, the graph-cut algorithm has been used to detect video moving objects though the energy minimization technique under the framework of MAP-MRF (Maximum a posterior-Markov random field). Most graph-cuts [8,9] algorithms focus on the iterative process and the priori information of moving objects in order to improve the detection accuracy. Stereo-based segmentation [10] seems to achieve the most robust results by fusing color, contrast and stereo matching information. But it requires special stereo input. However, graph cut requires labeling of the source and sink seeds by a human operator. Through our improved VIBE method, this can get initial foreground region and then provide the seeds of foreground and background.

In this paper, in order to overcome the shortcoming of origin VIBE and provide accurate silhouette with good spatial and temporal consistency, we present a novel moving object detection method based on improved VIBE and graph cut method from monocular video sequences. Firstly, perform moving object detection for the current frame based on improved VIBE method to extract the background and foreground information; Then obtain the clusters of foreground and background respectively using mean shift clustering on the background and foreground information; Third, initialize the S/T Network with corresponding image pixels as nodes (except S/T node); Calculate the data and smoothness term of graph; Finally, use max flow/minimum cut to segmentation S/T network to extract the motion objects. Experimental results on indoor and outdoor videos demonstrate the efficiency of our proposed method.

The paper is organized as follows: Section 2 reviews the previous work related to layer extraction. Section 3 addresses how to extract layer descriptions from a short video clip. Section 4 deals with the use of the occlusion order constraint, three-state pixel graph, and a multiframe graph cuts algorithm for obtaining layer segmentation in the presence of occlusion. Finally, in Section 5, we demonstrate several results obtained by our method.

2. Review of VIBE algorithm

Visual Background Extractor (ViBe) as a universal background subtraction method is proposed by Olivier. The algorithm adopts neighboring pixels to establish the background model by comparing the background model with the current pixel value. The implementation of the algorithm is subdivided into three steps [7]:

The first step is to initialize the single frame image from each pixel in the background model. Since there is no temporal information in a single frame, it is supposed that neighboring pixels share a similar temporal distribution. This means that the value of a pixel and its neighbor pixel values in spatial domain has a similar distribution. The right size range of the neighborhood is chosen to make sure that the background model includes a sufficient number of different samples, while keeping in mind that as the neighborhood scope increases, the correlation between pixel values at different locations decreases. It is assumed that $T=0$ denotes the first frame; $BG^0(x,y)$, $N_G(x,y)$, $P^0(x,y)$ is the pixel background model value, spatial neighborhood value, pixel value, then:

$$BG^0(x,y) = \{P_m^0(x^n, y^n) | (x^n, y^n) \in N_G(x,y)\} \quad (1)$$

where (x^n,y^n) in $N_G(x,y)$ is selected such as probability. $m=1,2,\ldots,N$ is the number of samples.

Secondly there is a simple decision process to determine whether an input pixel belongs to the background or not, then to update the background pixel model. They denote by the pixel value $P(x,y)$ in a given Euclidean color space taken by the pixel located at (x,y) in the image, and each background model is modeled by a collection of 20 background samples. When $T=t$, the background model of (x,y) is $BG^{t-1}(x,y)$, the pixel value is $P^t(x,y)$, the formula for determining the input frame image for foreground objects segmentation is as follows:

$$P^t(x,y) = \begin{cases} \text{foreground} & |BG^{t-1}(x,y) - P^t(x,y)| > R \\ \text{background} & |BG^{t-1}(x,y) - P^t(x,y)| < R \end{cases} \quad (2)$$

where superscript r is randomly chosen; R is the fixed shold (which in color space is a spherical radius). When $P^t(x,y)$ is larger than or equal to a given background candidate #min (the minimal cardinality), we think it is the corresponding background pixel, otherwise is foreground. In ViBe algorithm, #min is set to 2. It means that it is sufficient to find at least two samples close enough to classify the pixel in the background.

Because the existence of a moving object in the first frame will introduce an artifact commonly called a ghost (regions of connected foreground points that do not correspond to any real object). Although using the subsequent frames can update the background to remove the ghost, the process is a little slow. Because of the existence of ghost, the accuracy of detection will be decreased. Otherwise, the detection result of ViBe algorithm is sensitive to illumination changes, dynamic scenes, cluttered background.

A good background subtraction method has to be adapted to changes of the background caused by different lighting conditions but also to those caused by the addition, removal, or displacement of some of its parts, camera jitter.

The third step is to randomly update the background model with each new frame. Because of the strong statistical correlation between a pixel and its neighbor pixel, when a pixel is detected as the background pixel, its neighbor pixel is highly possible to be considered as the background pixel with high possibility. Consequently, it allows using the background pixels to update the background model of neighboring pixels. Through the random updating policy we can merge the foreground objects which halt suddenly or stay long into the background model.

3. Our proposed algorithm

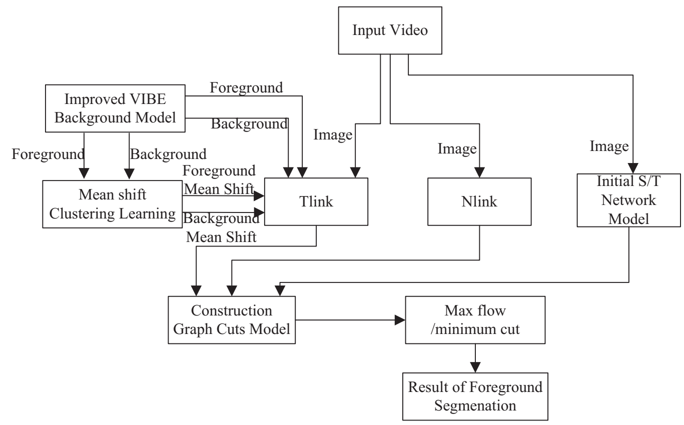

Fig. 1. Flowchart foreground segmentation algorithm GC_STVIVE.

The block diagram of our algorithm is demonstrated by Fig. 1.

The steps of our proposed approach – foreground segmentation for moving object in this paper can be briefly summarized as follows: