Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2016, 3(6):426-432**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Moving Object Detection and Tracking in Video Surveillance System

Sabri M. A. A. Ahmed[1], Othman O Khalifa[2]*

[1]College of Computer Science & Information Technology, Sudan University of Science and technology, Khartoum, Sudan
[2]Electrical and Computer Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia

Abstract Detecting and tracking objects in crowded areas is a challenging issue in the field of Video Surveillance System. Nowadays the increase of digital video cameras, and the availability of video storage and high performance video processing hardware, opens up conceivable outcomes for tackling many video understanding problems. Developing a real-time video understanding technique which can process the large amounts of data becomes very important. The object detection first step used in surveillance applications aims to separation of foreground objects from the background. Many algorithms proposed to solve the problem of object detection, however, it still lack of tracking multiple objects in real time. Object tracking used to find a moving object detected in motion detection stage from one frame to another in an image sequence. This paper focuses on review of various techniques used in object detection and object tracking.

Keywords Object detection, Frame difference, Background subtraction, Segmentation, Object Tacking

Introduction

Visual traffic surveillance has attracted significant interest in computer vision, because of its tremendous application prospect. The main problems of many surveillance and video understanding systems are the themes of target detection, classification and tracking because effective segmentation of foreground (region of interest in image) and background plays a key role in subsequent detection, classification and tracking results [1]. For example, traffic surveillance system can be used to detect vehicles, generate trajectories from video data and extract important information like gap, accelerating vehicles and stopped vehicle detection; this information can be used to manage traffic on special roads, specify the speed and is also useful for Intelligent Transportation Systems. A generic video processing framework for smart algorithms [2] is shown in Figure 1. This framework provides a good structure for Robust Vision-based Moving Target Detection and Tracking System because it can be considered as a surveillance application.

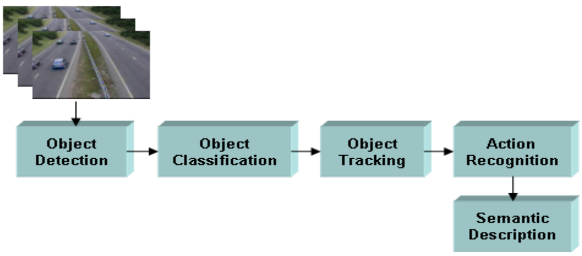

Figure 1: A generic framework for smart video processing algorithms [2]

The paper is organized as follows: Section II covers techniques, related to moving object detection, such as background subtraction, optical flow, statistical methods and temporal differencing, while section III covers major approaches used for object classification, section IV object tracking techniques and related work are reviewed. Finally, conclusions are given in section V.

Moving Object Detection

Each application needs different requirements to use smart video processing in an efficient way. However, the common first step between all applications is the separation of foreground objects from background [3]; this step is used for detecting regions that correspond to moving objects such as humans and cars in video.

Motion detection is a difficult problem to process reliably because of the dynamic changes in natural scenes such as the motion of trees and rain; so there are many techniques used for moving object detection such as background subtraction, optical flow, statistical methods and temporal differencing [2, 4-5]. The main goal of these techniques is: segmentation of an image, or video stream, into object vs. non-object regions. This is based on matching regions of interest to reasonably detailed target models. Another requirement of these systems is to have a large number of pixels on the target [1, 5]. Every tracking method requires an object detection mechanism either in every frame or when the object first appears in the video. From the preliminary literature review, object detection methods can be divided into three categories namely point-based detectors, segmentation-based detectors, and background modeling detectors. The architecture of the Human Detection System [6] is shown in Figure 1. Also, tracking an object aims at locating the object and finding the region it encompasses in the image at every time instant. Object tracking methods can be divided into 3 categories to correspond to the categories used for object detections. These categories are tracking by Point Correspondence, Tracking by Matching Primitives, and Tracking by Contour Evolution.

Background Model and Subtraction

Background subtraction is a common technique used for motion segmentation in static scenes. To detect moving regions this technology subtracts the current image pixel-by-pixel from a reference background image that is created by averaging images over time in an initialization period. The pixels where the difference is above a threshold are classified as foreground. After creating a foreground pixel map, some processing operations are performed to reduce the effects of noise and enhance the detected regions. With new images over time the reference background is updated to adapt to dynamic scene changes [2, 4].

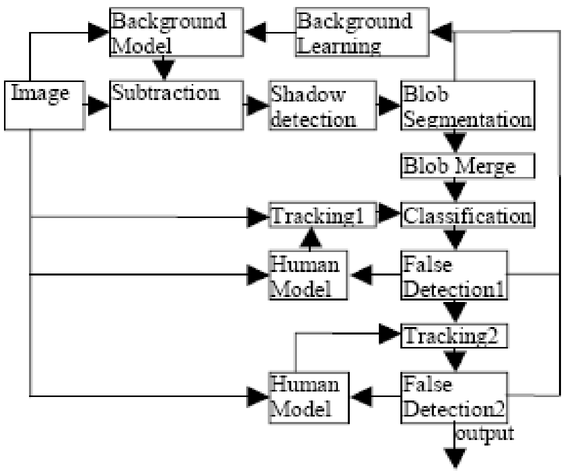

Figure 2" The Architecture of Human Detection System

The simple version of this scheme where a pixel at location (x, y) in the current image is marked as foreground [2][6] as follows:

$$|I_t(x,y) - B_t(x,y)| > T \qquad (1)$$

Is satisfied where T is a predefined threshold. The background image Bt is updated by the use of an Infinite Impulse Response (IIR) filter as follows:

$$\mathrm{B}_{t+1} = \alpha \mathrm{I}_t + (1-\alpha)\mathrm{B}_t \qquad (2)$$

Adaptation coefficient is donated by α; The foreground pixel map creation is followed by morphological closing and the elimination of small-sized regions. About eight background subtraction techniques were reviewed as follows: Gaussian average, Temporal median filter, Mixture of Gaussians, Kernel density estimation (KDE), Sequential KD approximation, Co-occurrence of image variations and Eigen backgrounds technique. Those techniques were ranging from simple approaches, used for maximizing speed and the size of the memory required, to complicated approaches, used for accomplishing the highest possible accuracy under any potential circumstance [2-3]. All approaches were prepared for real-time performance. Although background subtraction techniques perform well at extracting most of the relevant pixels of moving regions even they stop, they are usually sensitive to dynamic changes when, for instance, stationary objects uncover the background (e.g. a parked car moves out of the parking lot) or sudden illumination changes occur [2].

Statistical Methods

To overcome the fault of basic background subtraction methods many methods that make use of the statistical characteristics of individual pixels have been developed.

These statistical methods are mainly inspired by the background subtraction methods in terms of keeping and dynamically updating statistics of the pixels that belong to the background image process. Foreground pixels are identified by comparing each pixel's statistic with that of the background model [2-4].

Recently, it has become popular to model the pixel wise color distribution of the background through statistical methods. Parametric methods for example assume a parametric model for each pixel in the image, and then try to classify it as foreground or background using this model [6-7]. Gaussian probability is the simplest approach for assuming distribution for each pixel. Then model update each pixel values from the new frames (images) in the video. After the model collects enough information that was used to decide pixel (x,y) is background pixel by satisfies this equation [7-8]:

$$\mathrm{I}\,(\mathrm{x},\mathrm{y}) - \mathrm{Mean}(\mathrm{x},\mathrm{y}) < (C \times Std(x,y)) \quad (3)$$

Where: I (x,y) is pixel intensity, C is a constant, Mean(x,y) is the mean, Std(x,y) is the standard deviation, If it does not satisfy, it is marked as foreground pixel [8].

However the probability distribution of such background cannot be captured using a single Gaussian. Good foreground object detection results were reported by applying Mixture of Gaussians (MoG) to outdoor scenes. Further investigations showed that MoG with more than two Gaussians can degrade the performance in foreground object detection; but there was also a problem of determining number of Gaussians to be used to get optimal results [9].

The W4 system uses a statistical background model where each pixel is represented with its minimum (M) and maximum (N) intensity values and maximum intensity difference (D) between any consecutive frames observed during initial training period where the scene contains no moving objects. A pixel in the current image It is classified as foreground if it satisfies [10]:

$$|\mathrm{M}(\mathrm{x},\mathrm{y}) - \mathrm{It}(\mathrm{x},\mathrm{y})| > D(x,y) \quad (4)$$

Or

$$|\mathrm{N}(\mathrm{x},\mathrm{y}) - \mathrm{It}(\mathrm{x},\mathrm{y})| > D(x,y) \quad (5)$$

After thresholding, a single iteration of morphological erosion is applied to the detected foreground pixels to remove one-pixel thick noise. In order to grow the eroded regions to their original sizes, a sequence of erosion and dilation is performed on the foreground pixel map. Also, small-sized regions are eliminated after applying connected component labeling to find the regions. The statistics of the background pixels that belong to the non-moving regions of current image are updated with new image data [10].

Another way to perform foreground object detection is use Kalman filters. Each pixel is modeled using a Kalman filter. A robust Kalman filter is used in tracking explicit curves. A robust Kalman filter framework is for the recovery of moving objects. However, the framework does not model dynamic, textured backgrounds. Kalman filter also works on dynamic videos [9].

Temporal Differencing (DT)

There are many variants on the DT method, but the simplest is to take consecutive video frames and determine the absolute difference [1,3]. A threshold function is then used to determine change. If In the intensity of the nth frame, then the pixel wise difference function Δn is [1,3]:

$$\Delta_n = |I_n - I_{n-1}| \tag{6}$$

and a motion image M_n can be extracted by thresholding

$$M_n(u,v) = \begin{cases} I_n(u,v) & , \quad \Delta_n(u,v) \geq T \\ 0 & , \quad \Delta_n(u,v) \geq T \end{cases} \tag{7}$$

The threshold T has been determined empirically to be ≈15% of the digitizer's brightness range. For a digitizer providing 255 grey levels, a value of T≈40 should be used [3]. Temporal differencing is powerful to dynamic changes of the environment [9], however this technique generally gives poor results in extracting all the relevant pixels and leaves holes inside moving objects.

Object Classification

Real world video usually contains different objects such as humans, vehicles, clutter, etc; so it is very important to classify the type of detecting object then track it reliably id the correct way. In this section, we review two major approaches for moving object classification which is a shape-based and motion-based method [1-2,13-14].

Shape-based Classification

To detect object regions shape-based classification schemes are used common features such as the bounding rectangle, area, gradient, and silhouette. It depends on the objects' silhouette contour length and area information; the objects can be classified into three groups: human, vehicle and other. The method in [1] depends on the assumption that humans are, in general, smaller than vehicles and have complex shapes. Dispersedness is used as the classification metric and it is defined in terms of object's area and contour length (perimeter) as follows:

$$\text{Dispersedness} = \frac{\text{Perimeter}^{\;2}}{\text{Area}} \tag{8}$$

Classification is performed at each frame and tracking results are used to improve temporal classification consistency [1].

Motion-based Classification

There are many methods used temporal motion features of objects to recognize their classes; As generally those methods used to classify rigid objects (e.g. vehicles) from non-rigid objects (e.g. human). The method in [13] is to classify objects depending on temporal self-similarity of a moving object. As an object that exhibits periodic motion evolves, its self-similarity measure also shows a periodic motion. The method exploits this clue to categorize moving objects using periodicity. Optical flow analysis is also useful to distinguish rigid and non-rigid objects. The method in [14] proposed makes use of the local optical flow analysis of the detected object regions. It is expected that non-rigid objects such as humans will present high average residual flow whereas rigid objects such as vehicles will present little residual flow. Also, the residual flow generated by human motion will have a periodicity. By using this cue, human motion, thus humans, can be distinguished from other objects such as vehicles.

Object Tracking

The main goal of object tracking is to find a moving object detected in motion detection stage from one frame to another in an image sequence [15-16, 19]; so it's an important task in most of the surveillance systems because it provides cohesive temporal data about moving objects which are used both to enhance lower level processing such as motion segmentation and to enable higher level data extraction such as activity analysis and behavior recognition [2]. The common object tracking problems of erroneous segmentation are long shadows, partial and full occlusion of objects with each other and with stationary items in the scene [2, 15]. In this section, we review the object tracking by classifying it [15, 16] as point tracking, kernel tracking, and contour tracking according to the object representation method.

Point Tracking

The point tracking method represents the target being tracked by points which are detected in consecutive frames with the tracking procedure. Point representation of a target object has robustness to the changes of rotation, scale, and affine transform. Point tracking can be classified into the deterministic and the statistical methods depending on the matching method used for finding point correspondences [15-16].

The deterministic method uses proximity, maximum velocity, common motion, and rigidity constraints to match point correspondence. The statistical method represents an object by the state-space of object parameters such as position, velocity, and size. When tracking is performed with state-space representation, the state is estimated using the dynamic model of state transition, and updated by the correction stage using the measurement from the image [15]. Representative methods for estimating the dynamic model in statistical point tracking include the Kalman filter and the particle filter. The particle filter calculates state probability using the sequential importance sampling method and corrects state probability using the measurement. It can handle non-Gaussian state and non-Gaussian noise. Thus, a particle filter can track a point in a general environment. However, if the state and noise distribution follow the Gaussian distribution, then the Kalman filter provides a better optimal solution [15].

Kernel Tracking

The kernel tracking represents a target object by a primitive object region such as a rectangular, ellipse, or circle, and tracking is performed by computing object motion from one frame to the next. Usually, the motion of the object is assumed in the form of a parametric model such as translation, conformal, and affine transform [15]. Kernel tracking is a popular method, because it is robust to uncertain spatial deformations and its broad range of convergence. Kernel tracking can be classified into template model and appearance model. The template model matches the target using a similarity measure between the template and a candidate image. Rectangular and ellipse templates have been widely used to characterize the object and histogram of the template and their intensity values are used to calculate the similarity score [15]. The template model approach has been widely used because of its computational simplicity. In [17] the Visual Surveillance and Monitoring (VSAM) system uses cross-correlation method to track the object detected by a motion detection module. In W4[18] cross-correlation function is also used to track human body parts. Recently, there have been some attempts to make the computation speed of similarity function faster, or to reduce the search area of similarity measurement to shorten the computation time. In VSAM the object tracker uses the sub-sampling method with motion information to reduce the computational cost in the template matching process.

Contour Tracking

Contour tracking method iteratively evolves an initial contour which represents the target object using an outline contour from the previous image to the next. Contour’s representation ability provides an efficient tracking method to a target object with a complex shape and various changes of shape over time [5, 15].

Thus, recent studies have applied contour tracking method to the non-rigid object such as human tracking. Contour tracking method was present based on the particle filter, as known as the Condensation algorithm. Condensation algorithm is the first application to use particle filtering for object tracking in the computer vision community. It can handle the non-Gaussian distribution of the state and the noise to overcome the limitations of the Kalman filter in a complex cluttered image, showing non-Gaussian distribution [15].

Extensive research has been conducted for object detection and tracking, the [17] was present algorithm for real-time detection and tracking of moving targets in terrestrial scenes using a mobile camera. The algorithm consists of two modes: detection and tracking. In the detection mode, background motion is estimated and compensated using an affine transformation. The resultant motion rectified image is used for detection of the target location using split and merge algorithm. When the target is identified, algorithm switches to the tracking mode. Modified Moravec operator is applied to the target to identify feature points. The feature points are matched with points in the region of interest in the current frame. The corresponding points are further refined using disparity vectors. The tracking system is capable of target shape recovery and therefore it can successfully track targets with varying distance. However this algorithm used gray images to recognize the object because the color-based recognition is complicated; hence, the system can be affected by the state of the surrounding environment.

A new method was presented for vehicle detection, tracking and classification based colour. The detection stage has used the background subtraction technique, where the background image is subtracted from the current frame and the Pixel values greater than the set threshold are considered as part of the foreground. In the tracking stage Kalma filter was employed to track the detected vehicle's position from frame to frame. After that tracked vehicle segmented at the point when its colour is more visible (at the point closer to the camera), than extract small patch from the segmented vehicle and used it as a colour sample. The last stage of the method the extracted patch is compared against the gallery to determine the colour class (red, green, white or black) of the vehicle. However this method can detect and track only one object (car) that near from the camera and ignore the rest objects [18].

A method was proposed that records only the part of the video that contains important information and ignore the rest of the video, as example the part that contains motion in the scene. The proposed method solves two problems: the times consumed to record and view the video and excessive storage space required to store the video. This goal was achieved with a digital video camera and Digital signal processing algorithm that detects motion; the surveillance system that was developed based on TI DSP 'C54' and their algorithm called block-based MR-SAD (Mean Reduced – Sum Average Difference). However this method need special hardware and does not work with exist surveillance systems camera [19].

Joshi *et al* introduced architecture of an automatic real-time video surveillance system, capable of autonomously detecting anomalous behavioral events. The proposed system automatically adapts to different scenarios without human intervention, and applies self-learning techniques to automatically learn typical behavior of targets in each specific environment. Anomalous behaviors are detected whenever the observed trajectories deviate from the typical learned prototypes. However the proposed system does not accomplish the true requirements of a surveillance system, where typically the observed area comprises different levels of security. Another negative aspect of this approach is the use of simple spatial information from the center-of-mass position of the object, without considering other important features (e.g. color) [20].

Another detection technique applied in video surveillance and monitoring systems is an improved motion detection algorithm [18-19, 21] based on an integrated algorithm consisting of the temporal frame differencing, optical flow, double background filtering, and morphological processing methods. In this algorithm a temporal frame differencing of a single frame is implemented to obtain the region of change. A single-frame differencing is computed an average weight coefficient is multiplied on the differencing function. The adaptive threshold is represented by three times the mean of the temporal differencing function. After an optical flow detection is calculated to detect the real object movement. The next steps are using a double background filtering and morphological image processing based on the prior processed information. Despite the good detection results that this algorithm gets, it is described as a very complicated and high time processing algorithm.

Conclusion

Moving target detection and tracking are an important research field of video processing for its great potential in Military and Civil applications. In this paper, we review various techniques used in object detection and object tracking used in Video-Based Surveillance System. We focus on summarizing the different technologies used in Object Detection and tracking. In addition, many related works have been reviewed.

References

[1]. J. Lipton, H. Fujiyoshi, and R.S. Patil. Moving target classification and tracking from real-time video. In Proc. of Workshop Applications of Computer Vision, pages 129–136, 1998.

[2]. Y. Dedeoglu, Moving object detection, tracking and classification for smart video surveillance, Master's Thesis, Dept. of Computer Eng. Bilkent University, Ankara, 2004.

[3]. N. Paragios, and R. Deriche.. Geodesic active contours and level sets for the detection and tracking of moving objects. IEEE Trans. Patt. Analy. Mach. Intell. 22, 3, 266–280, 2000.

[4]. S. Y. Elhabian, K. M. El-Sayed, “Moving object detection in spatial domain using background removal techniques- state of the art”, Recent patents on computer science, Vol 1, pp 32-54, Apr, 2008.

[5]. Elgammal, A. Duraiswami, R.,Hairwood, D., Anddavis, L. 2002. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. Proceedings of IEEE 90, 7, 1151–1163.

[6]. J. Zhou and J. Hoang. “Real Time Robust Human Detection and Tracking System”, IEEE Intl. Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum, June 20, 2005

[7]. N. Paragios, and R. Deriche.. Geodesic active contours and level sets for the detection and tracking of moving objects. IEEE Trans. Patt. Analy. Mach. Intell. 22, 3, 266–280, 2000.

[8]. Ambardekar, M. Nicolescu, G. Bebis, “Efficient Vehicle Tracking and Classification for an Automated Traffic Surveillance System,” Signal and Image Processing, August 2008.

[9]. Raad Ahmed Hadi, Ghazali Sulong, and Loay Edwar George, "Vehi-cle Detection and Tracking Techniques: a Concise Review," In the In-ternational Journal of Signal & Image Processing (SIPIJ), Volume 5, Number 1, pp. 1-12, February 2014.

[10]. Y. chalapathi, I S. prabha, Moving Vehicle Detection Using Kalman Method, International Journal of Advanced and Innovative Research (IJAIR), Volume 1, Issue 2, July 2012.

[11]. Haritaoglu, D. Harwood, and L.S. Davis. W4: A real time system for detecting and tracking people. In Computer Vision and Pattern Recognition, pages 962–967, 1998.

[12]. L. Wang, W. Hu, and T. Tan. Recent developments in human motion analysis. Pattern Recognition, 36(3):585–601, March 2003.

[13]. W.M. Hu, T.N. Tan, L. Wang, and S.J. Maybank “A Survey on Visual Surveillance of Object Motion and Behaviors,” IEEE Trans. on Systems, Man, and Cybernetics—Part C: Applications and Reviews,vol. 34, no. 3, pp. 334-352, August 2004.

[14]. R. Cutler and L.S. Davis. Robust real-time periodic motion detection, analysis and applications. In IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 8, pages 781–796, 2000

[15]. J. Lipton. Local application of optic flow to analyse rigid versus non-rigid motion. Technical Report CMU-RI-TR-99-13, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, December 1999.

[16]. In Su Kim, Hong Seok Choi, Kwang Moo Yi, Jin Young Choi, and Seong G. Kong. Intelligent Visual Surveillance - A Survey. International Journal of Control, Automation, and Systems (2010) 8(5):926-939.

[17]. Yilmaz, A., Javed, O., and Shah, M. 2006. Object tracking: A survey. ACM Comput. Surv. 38, 4, Article 13,December 2006.

[18]. Behrad, A. Shahrokni, S.A. Motamedi, A robust vision-based moving target detection and tracking system. In the Proceeding of Image and Vision Computing Conference, University of Otago, Dunedin, New Zealand, November 2001.

[19]. Nan Lu, Jihong Wang, Q.H. Wu and Li Yang, “An Improved Motion Detection Method for Real-Time Surveillance”, IAENG International Journal of Computer Science, 35:1, IJCS_35_1_16, 2008.

[20]. Kinjal A Joshi, Darshak G. Thakore, A Survey on Moving Object Detection and Tracking in Video Surveillance System, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, July 2012.

[21]. Payal Panchal1 , Gaurav Prajapati2 , Savan Patel3 , Hinal Shah4 and Jitendra Nasriwala, A Review on Object Detection and Tracking Methods, International Journal For Research In Emerging Science and Technology, Volume-2, Issue-1, January-2015