A Review on Multiple Object Detection and Tracking in Smart City Video Analytics

Fancy Joy, V. Vijaya kumar

Abstract: Moving object detection and tracking are the two important challenging tasks in the smart city video analytics system. It is challenging due to occlusion, presence of shadows, cluttering, dynamic background, noise etc. Detection of moving objects, tracking, object matching across multi-camera, and re-identification are the basic steps of multi camera video analytics system. Multiple object detection and tracking in smart city video analytics can be developed according to appropriateness of society such as intelligent surveillance, smart parking, traffic monitoring, vehicle navigation, smart healthcare etc. The goal of this paper is to analyze and review various approaches towards multiple object tracking.

Keywords: Object detection, object tracking, video analytics, multi camera surveillance.

I. INTRODUCTION

Moving object detection and tracking are the two most significant steps in various video based applications such as video analytics, traffic monitoring, suspicious event detection etc. Video analytics is the emerging technology where computer vision and pattern recognition techniques are used to filter and manage real time surveillance video for intelligent monitoring (Ghasemi & Ravi Kumar, 2015). Smart city video analytics detects and determines temporal and spatial events to automatically analyze video. The key steps of smart city video analytics consist of object detection, tracking the object from frame to frame and camera to other camera, object matching across camera. Detection and tracking become challenging task due to occlusions, illumination variation, presence of shadows, complexity of background etc.

The organization of this paper is arranged as follows, Section 2 describes the steps for moving multi-object detection and tracking, Section 3 discuss the applications and challenges of multi-object detection and tracking and section 4 presents the conclusion.

II. MOVING OBJECT DETECTION AND TRACKING

Video analytics analyzes live or recorded video and generates data for pre-emptive action or datamining purposes (Vijayakumar & Mini T.V, 2017). Video data includes collection of video frames with temporal features. Object detection from a frame and keep tracking that objects throughout the multi camera surveillance is a challenging task. In order to track multiple objects across multi-camera surveillance needs to extract features or information of objects from video data.

Basic steps for moving object detection and tracking in smart city video analytics depicted in Figure 1.

Figure 1: Basic steps for multiple moving object detection and tracking.

Video data contains sequence of video frames. It may contain irrelevant data such as noise. Hence video preprocessing is required to convert collected data to a quality one. This process is also known as key frame extraction. Features of detected objects are tracked and stored in the corresponding repository for further processing such as multiple object detection and tracking. Multiple object tracking trace each detected object in video from frame to frame and camera to camera.

A. Video Preprocessing

Video analysis become complex task due to occlusion, blur in images, illumination variation, noise etc. To overcome these challenges an efficient key frame extraction technique is needed. Key frame extraction has a vital role in video preprocessing. Zhenget.al (2015) used parallel key frame extraction method using motion information and local maxima extraction for smart city surveillance. Makandarand Mulimani (2016) proved that key frame extraction using maximum value of video frame difference achieved better accuracy in key frame extraction. Video data contain different kinds of noises such as median noise, mean noise, impulse noise, guassian noise etc. Noises reduce the quality of images. Domanskiet.al (2009) M.F Hashmi et.al (2016) used morphological operations (dilation and erosion) for fast removal of noise from video frame. Samidhaand Amit (2013) suggested condensation approach for efficient video denoising.
Proposed method used non local means particle filter algorithm. But the approach is computationally expensive. Clement Godard et al. (2016) used deep burst denoising technique for reducing noise. The authors used deep neural architecture to process burst of images. Still needs to improvement for real time videos. Hou and Shen (2018) suggested image denoising with morphology and size adaptive block matching method achieved better results in less computational complexity.

B. Object Detection

It is used to find out the presence of object in a video frame and to detect those objects. It is classified into four categories such as template matching, knowledge extraction based, Object Based Image Analysis (OBIA) based, and machine learning based. Template based method uses a rigid template for detecting moving objects in small scale variation and depends on scale and variation (Feng et al., 2008). Knowledge-based techniques are based on rules that codify the human knowledge about the objects of interest and its characteristics. Knowledge based, OBIA based method comprises two steps; image segmentation and classification for moving object detection (Aydin et al., 2017). Machine learning based uses object detection machine learning technique mainly for feature extraction, feature fusion, classification etc. (Cheng & Han, 2016).

C. Multiple Object Detection

Multiple object detection is quite different from single object detection. It can be used for many applications such as sports, traffic monitoring, crowded scenes, security monitoring etc. Author Del-Bianco et. al (2012) developed an efficient frame work for multiple object detection and tracking. They used foreground detection with ellipse fitting method, which is the combination of Expectation Maximization algorithm and Levenberg Marquardt algorithm. The method used novel Bayesian tracking approach that can handle multimodal distribution, false detection and missing detections. Morimitsu et al. (2014) used spatio-temporal approach using graphs and probability maps for multiple object detection. Low level features, temporal consistency and structural properties were used to obtain better results. Improvement is needed to detect multiple objects independent of its color. Ashna and Nishidha proposed multiple object detection and tracking method based on local sparse representation. For multiple object detection the authors employ super pixel image segmentation followed by detection step with support vector machine (SVM) algorithm. They used sparse representation method along with incremental sub space learning for template update, solve the problem of occlusion. Tiancheng Li et al. (2017) proposed clustering for filtering (C4F) method multiple object detection. Prior knowledge about object is not needed for detection and it can also handle time varying uncertainties such as noise, clutter, misdetection etc. But it is only suitable for specific applications. Ankush Mahoraet al. (2017) used frame difference method with gaussian mixture model and background model. The method can overcome varying illumination condition, dynamic background, shadows, camouflage, and bootstrapping. Supreeth and Chandrashekar (2018) proposed mixture of Guassian based background subtraction method for multiple moving object detection. It can handle varying illumination condition, shadows, camera jitter, and occlusion. Mahalingam and Subramoniam (2018) used mixture of adaptive Gaussian model for multiple object detection. The method improves accuracy. But enhancement is needed to detect multiple objects in very crowded scene, extreme illumination variation condition and occlusion. Multiple object detection in smart city video analytics still become challenging one. A flexible algorithm is required for multiple object detection.

D. Feature Extraction and Matching

Extracted appearance based features can be color, shape, edge, position color histogram, histogram oriented gradients etc. can form a feature space (yazdi and Bowmans, 2018.). GLOH, SURF, and SIFT descriptors are used to represent moving object. Shift Invariant Feature Transform (SIFT) is more efficient because it generate features which are robust to noise, scale variation, and illumination but it is relatively slow to compute and match (Mikolajczyz and Schmid, 2005). This can be overcome by speed Up Robust Features (SURF) (Bay et al., 2006;1WBeng et al., 2013). SURF is used 64 vector for storage and in case of feature matching features are matched only if they have same contrast (Bay et al., 2008; Karami et al., 2017). Bhosale et al. (2014) also used SURF algorithm for feature extraction for multiple object detection. Binary Robust Independent Elementary Features (BRIEF) used binary strings as efficient feature descriptor (Calonder, 2010). BRIEF is very fast for both building and matching. Rublee et al. used Oriented FAST and Rotated BRIEF (ORB), very fast binary descriptor based on BRIEF and resistant to noise (Rublee et al., 2011). ORB act as an alternative to SIFT and SURF. Histogram oriented gradient (HOG) feature extraction method calculate the gradients, generate a histogram and normalize blocks. Local histogram used to represent moving object (Yuqian Li & Guangda Su, 2015). It is susceptible to illumination changes.

E. Object Tracking

Tracking can be defined as the problem of estimating the path of an object in the image plane as it moves around a scene (Ghasemi & Ravi Kumar, 2015). Multiple cameras required to cover wide area. Tracking in multi-camera surveillance is quite different from single camera surveillance. Tracking is the term is used to detect the moving objects position and tracking them through video sequences across different camera. Object tracking approaches categorized in to three; point tracking, kernel tracking, silhouette tracking. Point tracking method detect object in successive video frames and represent it as points, then associate these point to based previous state (object position and motion) of object (Yilmaz et al., 2006). External technique is needed to detect object in each frames. Kernel tracking track the object by computing kernel (object shape and appearance) motion in consecutive frames (Yilmaz et al., 2006).
Motion indicate the translation, rotation and affine. Silhouette tracking is based on estimating the object region in the each frame(Yilmaz et.al, 2006). Object region consists of information such as appearance, shape etc. It uses either shape matching algorithm or contour evaluation algorithm for tracking.

F. Multiple Object tracking

Multiple object tracking is more challenging than single object tracking due to it needs one to one correspondence between the observed object and target frame, tracking similar object, etc. Authors Xin Li et.al(2010) ,Uke and Futane (2016) ,BimaSahbani, et.al(2016) used kalman filter algorithm for multiple object tracking. Multiple target tracking by particle filter algorithm was discussed by the authors Jung et.al(2013),Shinji Fukui et.al(2016). Xin Li et.al(2010) used feature based kalman filter method for object motion detection for multiple object tracking. The proposed method is able to ensure an efficient and robust tracking with merge and split of multiple objects. Jung et.al(2013) presented dual layer particle filter algorithm for multiple object tracking. They used parent and child particle as first and second layer respectively. Intersection Kernel support vector machine was used by the parent particle layer for detection whereas child particle layer for tracking .The proposed system can be applied for video surveillance application, human computer interaction and intelligent transport system. Shinji Fukui et.al(2016) discussed particle filter based tracking with image based localization. Authors used 3D information of feature points and the camera position for image localization. The method can track robustly with objects with similar appearance and rotating camera environment. BimaSahbani.et.al(2016) proposed kalman filter with iterative Hungarian algorithm with euclidian distance for fast multi object tracking. The authors focused on multiple players tracking football. The proposed multiple object tracking computation could be implemented with less computational resources. Uke and Futane(2016) presented parallel kalman filter algorithm for motion estimation. They proposed hybrid method of object detection using motion estimation and tracking by parallel kalman filter. Shape and corner features were used for extract tracking of objects. The method achieves better precision and recall values. Chandrajit M et.al(2016) used color and hue moments for multiple object tracking. Nearest neighbor classifier and chi-square dissimilarity is applied to track objects in successive frames. Still needs improvement in order to track multiple objects under shadow and occlusion. AljosaOsep et.al,(2016) proposed general tracking approach for tracking multiple objects simultaneously without explicitly learning about classifiers. The authors used competitive semantic segmentation algorithm for tracking. The method achieves competitive results in 3D tracking in close- range objects. Suprecht and Chandrashekar(2018) used k-means clustering for efficient multiple moving object tracking with initialization and learning phase. The proposed method handled occlusion, camera jitter and shadows. Multiple object tracking is a difficult task compared to single object tracking. A better algorithm is needed for real time multi tracking.

III. APPLICATIONS AND CHALLENGES

The main applications of moving object detection and tracking are the follows (Fancy and Vijayakumar,2018):

A. Motion based recognition:
Identification of object or its motion based on motion in series of images(Yilmaz et.al, 2006).Zhang et. al (2013) used color based meanshift segmentation algorithm for detection and Bayesian kalman filter with improved mean shift for tracking in multi camera network.It requires improvement to increase performance.Coskun and Unal (2016),used camshift technique for tracking, but technique fail to perform video with full occlusion. Oiwa et al. (2016) constructedprobabilistic background model for tracking the objects from video sequences. The speed and accuracy of the method is high compared with the existing one.

B. Automatic surveillance:
Observing a scene to identify dubious activities or improbable actions (Yilmaz et.al, 2006).Sarkar et.al(2012) used neural network based detection and tracking method in low resolution surveillance video, but the method can’t detect faces from gray scale images and also unable to detect very small face images that is away from camera. SamyS . A. ghonemy (2015) uses multi objects visual color tracking algorithm in real time surveillance. But it is difficult to track the objects in dynamic environment.Prevaltali et al. (2017) suggested P-tracking for distributed multi camera multiple object tracking based on distributed multi – clustered particle filtering for surveillance video. Performance can be improved with more cameras.

C. Video Indexing:
Automatic annotation and retrieving the videos in multimedia databases (Wilma et.al, 2006).Video indexing is video concept detection. With the advancement of storage devices large kind of video data are now publicly accessible. In order to retrieve the required data from bulk data it needs an automatic mechanism video indexing.Zhaet.al(2011) used new active learning approachie, local learning regularized least square model to utilize local structure of video for interactive video indexing. Shruthi and priyanvadhat(2017) constructed different feature extraction model with dominant frame generation for each video frame. But it is not applicable in real time video.Chamasemani.et.al(2017) used abstraction forindexing and retrieval of surveillance video. Method reduces the computation complexity compared to various process used for video retrieval and proved significant reduction in processing time also.

D. Traffic monitoring:
Directing traffic flow by gathering real time statistics uses GPU based Gabor filter as a directional filter to filter the vehicle movement towards the camera. Accuracy of the method needs to be improved. Hung presented a traffic monitoring system for a mixed traffic flow via road estimation and analysis to estimate the accurate traffic flow. It is based on the road analysis and detection of vehicle queue using surveillance camera.The method also used for automatic timing control of traffic signal.
E. Vehicle navigation
Path planning based on videos and capability of avoiding obstacles. Shinte and chorage used intelligent car parking system using localization approach of robots. The system optimizes the performance and resource utilization. Sensors are used for avoiding obstacles. Roman S. Kulikov and Alexandr I. Perov developed method for vehicle navigation in urban canyon using global navigation satellite system (GNSS). Accuracy can be improved by integrating proposed method with other available method.

The challenges which affect the object tracking are follows:

F. Illumination variation:
Light variation on surface due to day time variation, weather, obstacles to light source etc. Authors used object tracking method under varying illumination condition. Still some improvements needed.

G. Noise in images:
Noise is the factor which mainly affects the quality of video frame. A low quality video can cause moving object detection and tracking a complex one. Halkarnikaret proposed object detection and tracking method gave good result in noisy environment.

H. Partial and full object occlusion:
In real time the subject can be hidden by other object fully or partially. Coskun presented object tracking using camshift method. But the method is sensitive to full occlusion.

I. Presence of shadow:
Shadows occur due to block light from source. Cucchiara proved that moving objects can be detected even in presence of shadow.

J. Complex object motion:
Tracking of very slow and very fast moving object is a challenging task. Zavery developed efficient object detection and tracking approach even the objects in very slow and very fast moving.

K. Clutter:
It indicates confusing background which causes detection and tracking difficult. Kim proposed moving object tracking under cluttered environment. Method was sensitive to illumination variation.

IV. RESULTS
In this paper, we found that multiple object tracking has become a major application area in computer vision. Surveillance generally demands that objects are tracked over a long periods of time, and in varying conditions. Tracking become challenging because of illumination variation, occlusion, motion of the camera, complex object shapes, objects such as waving trees, etc.

V. CONCLUSION
In this paper strongly recommended to many tracking algorithm have been proposed to handle these issues, but the problem is still exist. The surveillance area can be extended with multiple cameras and the vision information can overcome occlusion. This paper discusses various steps for multi-object detection and tracking in smart city video analytics. A better algorithm is required to detect and track multi objects in smart city video analytics.

REFERENCES
5. awardanaAdiprawita, Kalman Filter and Iterative-Hungarian Algorithm Implementation for Low Complexity Point Tracking as Part of fastmultiple Object Tracking System, in IEEE 6th International Conference on System Engineering and Technology (ICSET), 2016.
11. Clement Godard, Kevin Matzen and Matt Uyttendaele, Deep Burst Demossing,arXiv:1712.05790 vol.120.

Published By: Blue Eyes Intelligence Engineering & Sciences Publication


27. Lijun He, Xiaoya Qiao, Shuai Wen and Fan Li, Robust Object Tracking Based on Motion Consistency, Sensors 2018, 18, 572, 2018.


44. Yuqian Li and Guangda Su, Simplified histogram oriented feature extraction algorithm for hardware implementation, in International conference on computers, communications, and system, 2015.


AUTHORS PROFILE

1. FANCY JOY, Department of Computer Science, Sri Ramakrishna College of Arts and Science, Coimbatore, India.

2. Dr. V. VIJAYAKUMAR, Department of Computer Science, Sri Ramakrishna College of Arts and Science, Coimbatore, India.