

# Survey on Various Techniques on Moving Vehicle Detection, Tracking and Counting

Rajendra Manandhar <sup>1+</sup>, and Yongmei Liu <sup>1</sup>

<sup>1</sup> College of Computer Science and Technology, Harbin Engineering University, Harbin, China

**Abstract.** This survey describes different methods used for detection, object representation, tracking and counting of the moving vehicles on the road. With the change of time and advances in system, there have been a lot of improvements in the application, and many authors have provided different techniques for achieving certain goal. In this survey, we included many techniques that are available and currently being studied.

**Keywords:** Background Subtraction, Object Detection, Blob Tracking, Vehicle Counting

## 1. Introduction

Traffic Problem is an important concern in many cities in the world. This causes imbalance between the number of vehicles and the roads, causing the road traffic overcrowding, particularly in large cities. Lack of transportation system also creates the same problem. Intelligent transportation systems (ITS) are innovative applications whose purpose is to provide advance services related to different approaches of transport and traffic management, for safe and smarter use of the transport system.

Video camera has been used to count the number of objects for some time now. Before that, and still now, sensor equipment have been used to count the number of objects. While those equipment provide benefits, there are also some drawbacks, for example the sensor is only used for real time, which means the user cannot input other types of inputs. The equipment are also expensive and not multitasking.

Conventional vehicle detection methods include infrared detector, inductive loop detector, radar detector, etc. Normally the traffic management system requires many sensors. Video cameras can be used for multitasking – not only for the security, but also used instead of some of the sensor equipment. Video camera is inexpensive and likely to collect a huge amount of information, such as number of vehicles, speed, vehicle class, vehicle tracks, etc., which can also deduce information such as incidents, speeding, origin-destination of vehicles, macroscopic traffic statics, etc.[1]

How to get the best out of the camera? The simple answer is computer vision (CV). In simple words, CV is a study and understanding of images and videos taken by a digital camera. CV has grown reputation in recent years and been used in many fields and has also been applied for answering transportation problems. CV is the methods which is based on the software module that does not require special hardware. Using a usual image or video recording device and the simple pattern matching theory, one can find the number of object in the image or video scene from the known size of the object.

There are three key steps in video analysis: detection of interesting moving object, which is a moving vehicle in current case. Another key step is tracking that moving object, i.e. vehicle from frame to frame. And analysis of the object tracks to recognize their behaviour, here counting them.

<sup>+</sup> Corresponding author. Tel.: +86-15645032063  
E-mail address: rajendra@hrbeu.edu.cn

The rest of the paper is organized as follows. In section 2, vehicle detection techniques are discussed. In section 3, object representation is surveyed and vehicle tracking is mentioned in section 4 and finally in section 5 the vehicle counting is discussed.

## **2. Vehicle Detection**

Object detection is the process of finding occurrences of real-life objects in images or videos. Object detection algorithms normally use extracted features and learning algorithms to identify models of an object group. It is generally used in applications like image recovery, security, surveillance, etc. Different applications have different requirements, and thus have to treat differently, vehicle detection, tracking and counting has to be treated likewise.

Object detection is the basic step for every computer vision system since it offers the motivation and simplifies the processing on following analysis steps. Due to dynamic changes in natural scenes. Mostly in outdoor scenarios, like sudden brightness and weather changes, monotonous motions that cause disorder, moving object detection is a difficult problem to process consistently.

Mostly used methods for moving object detection in CV applications are:

### **2.1. Point detectors**

Point detectors are used to find interest points in image which have an expressive texture in their respective localities. A desirable quality of an interest point is its invariance to change in illumination and camera viewpoint.

### **2.2. Segmentation**

The aim of image segmentation is to partition the image into perceptual similar regions. Every segmentation algorithm address two problems, the criteria for a good partition and the method for achieving efficient partitioning [2]. Some of the common segmentation techniques are:

- Mean-Shift Clustering,
- Image Segmentation using Graph-cuts,
- Active Contours

### **2.3. Temporal differencing**

Temporal differencing attempts to detect moving objects by making use of the pixel-by-pixel difference of consecutive frames in the video sequence. This method is highly adaptive to dynamic scene changes, however, it generally fails in detecting whole relevant pixel of some types of moving objects. This method failed to detect all the moving object.

### **2.4. Optical flow analysis**

Optical flow methods use the flow vectors of moving things over time to identify moving areas in an image. They can spot signal in video sequences even from a moving camera, nonetheless, most of the optical flow approaches are computationally difficult and cannot be used in real-time without specific hardware.

### **2.5. Statistical methods**

More innovative methods that make use of the statistical features of different pixels have been developed to overcome the limitations of basic background subtraction approaches. These statistical approaches are mainly inspired by the background subtraction methods with regards of keeping and dynamically updating statistics of the pixels that belong to the background image procedure. Foreground pixels are recognized by matching each pixel's statistics with that of the background model. This method is becoming more widespread due to its consistency in scenes that comprehend noise, lighting changes and shadow [3].

### **2.6. Background subtraction**

Object recognition can be done by constructing an illustration of the scene called the background model and then finding divergence from the model for each new frame. Any major change in an image section from

the background model suggests a moving object. The pixels creating the sections experiencing change are selected for additional processing.

Background subtraction is a commonly used method for motion separation in motionless scenes. It tries to identify moving areas by subtracting the present 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 change is above the threshold are categorized as foreground. After generating a foreground pixel map, certain morphological post-processing procedures such as erosion, enlargement and closing are made to shrink the effects of noise and improve the identified areas. The reference background is updated with new images over time to familiarize to dynamic scene changes.

### 3. Object Representation

An object can be defined as something that is of concern for additional study. For example, automobiles on the road, ships in the ocean, airplanes in the sky etc. are set of objects that may be significant to track in particular domain. The object of interest can be characterized by their shapes and appearances.[4]

The common object shape representation used for tracking are:

#### 3.1. Points

An object is represented by a point, i.e. the centroid. It is suitable when the object is very small in the image.

#### 3.2. Silhouette and contour

Contour defines the border of the object and the area inside the contour is known as silhouette of the object. This type of representation is appropriate for tracking complex non-rigid shapes.

#### 3.3. Geometric shapes

The object shapes are represented by a rectangle, ellipse or other basic geometric shapes. Motion of the object is typically represented by translation, transformation of affine. They are appropriate for simple rigid objects, although they are also used for tracking non-rigid objects.

#### 3.4. Articulated shape models

The objects are composed of parts that are held together with joints. The relationship between the parts are governed by kinematic motion models. Object can be represented by modelling the constituent parts using cylinders or ellipses, shown as in Fig. 1(e).

#### 3.5. Skeletal models

Object skeleton can be taken out by applying medial axis transform to the object silhouette. The model is usually used as shape representation for identifying objects. Skeleton representation can be used to model both articulated and rigid objects.

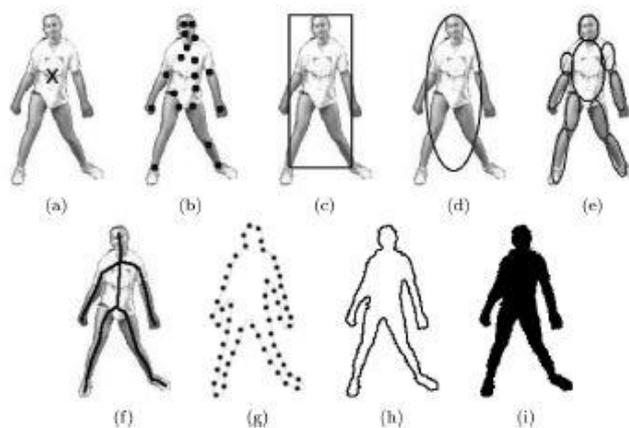


Fig. 1: Common object shape representation: (a) Centroid, (b) Multiple-Points, (c) Rectangular Patch, (d) Elliptical Patch, (e) Part-Based Multiple Patches, (f) Object Skeleton, (g) Complete Object Contour, (h) Contour Points on Object Contour and (i) Object Silhouette

There are number of ways to characterize the appearance feature of objects. Some common representation in context of object tracking are:

### **3.6. Probability densities for object appearance**

The probability density estimates of the object appearance can either be parametric, such as Gaussian and mixture of Gaussian, or non-parametric, such as Kernel Density Estimation and histograms. The probability densities of object appearance features can be calculated from the image regions identified by the shape models.

### **3.7. Templates**

Templates are made using simple geometric shapes or silhouettes. An advantage of a template is that it conveys both spatial and appearance data. They are appropriate for tracking objects whose positions do not fluctuate significantly during the sequence of tracking.

### **3.8. Active appearance models**

Active appearance models are created by instantaneously modelling the object shape and appearance. In broad-spectrum, the object shape is defined by a set of standards. The standards can exist in the object border or they can exist inside the object area. For each standards, an appearance vector is kept which is in the form of texture, color, or gradient magnitude.

### **3.9. Multi-view appearance model**

These models convert different views of an object. One method is to create a subspace from the given views. For appearance representation, Principle Component Analysis, Subspace approaches and Independent Component Analysis have been used.

There is a rigid association between the object representation and object tracking. Object representation are usually selected according to the application domain.

## **4. Vehicle Tracking**

Tracking can be well-defined as the problem of approximating the trail of an object in the image as it passages around the scene. Particularly, a tracker allocates consistent tags to the trailed objects in different frames of a video. Furthermore, depending on the tracking area, a tracker can also deliver object-centric statistics, such as area, orientation, shape and other information of object being tracked. Tracking objects can be difficult due to

- Complicated motion of the object,
- Low quality of image resulting noise,
- Information lost due to prediction of the 3D world on a 2D image,
- Flexible nature objects in the view,
- Partial obstructions in the view,
- Changes in lighting in the image sequence,
- Complex and un-uniform shapes of the object,
- Real-time processing requirements

Practically, there is a solid association between the object representation and the tracking. Object representation are generally selected according to the application domain. For tracking objects, point representation is remarkably suitable. Other than that, there are additional representations, like

- Object silhouette and contour,
- Skeletal models
- Primitive geometric shapes
- Articulated shape models

Choosing the accurate features plays a vital part in tracking. In general, the most anticipated properties of a graphical feature is its individuality so that the objects can be certainly illustrious in the feature space. Feature selection is narrowly associated to object representation. For instance, edges of the object are used as

feature for contour-based representation whereas color is used as the feature for histogram-based appearance model[4]. Some familiar features are

#### 4.1. Edge

Object borderline commonly make strong alterations in image intensities. Edge detection is used to recognize these alterations. A significant property of edge is that they are not very much delicate to lighting changes compared to that of color feature.

#### 4.2. Color

The obvious color of an object is subjective mainly by two physical factors, 1) the spectral power scattering of the light and 2) the surface reflectance properties of the object. In image processing, the RGB (red, green, blue) color space is commonly used to characterize color. Nevertheless, the RGB color space is not a perceptually constant color space.

#### 4.3. Optical flow

Optical flow is a compact field of displacement vectors which defines the conversion of each pixel in a region. It is calculated using the intensity limit, which take up intensity constancy of corresponding pixels in successive frames.

#### 4.4. Texture

Texture is a degree of the concentration variation of a surface which measures properties such as sleekness and symmetry.

The tracking category can be categorized as shown in Fig.2, the categories can be summarized as follows:

#### 4.5. Point tracking

Objects detected in successive frames are characterized by points, and the link between the points are based on the preceding object state which can contain object position and motion. This method requires an external tool to distinguish the object in each frame.

- *Deterministic Methods*: Deterministic methods for point correspondence describes a cost of correlating every object in frame  $t-1$  to particular object in the frame  $t$  using a set of motion restrictions. The correspondence cost is usually defined by using a mixture of proximity, common motion, maximum velocity, minor velocity change, firmness and proximal consistency.
- *Statistical Methods*: Measurement acquired from video sensors always contain noise. Furthermore, the object motion can experience arbitrary agitations, for instance, reversing vehicles. Statistical counterpart approaches resolve these doubts into account during object state assessment. The statistical counterpart approaches use the state space method to model the object properties such as velocity, position, and acceleration.

#### 4.6. Kernel tracking

Kernel denotes the object form and figure. For instance, the kernel can be a quadrilateral pattern or an oval shape with a related histogram. Objects are tracked by calculating the motion of the kernel in successive frames. This motion is typically in the form of a parametric change such as rotation, translation and affine.

- *Tracking using template*: Templates and density-based appearance models have been extensively used since they are relatively simple and costs low computationally. It can track single object or multiple objects at a time.
- *Tracking using Multi-view Appearance Model*: In earlier methods, the appearance models, that is templates, histograms, etc. are typically created online. The interested object may look different if viewed from different direction, and it may vividly change during the course of tracking, in such situation the appearance model may no longer be usable, and the trail of the object being tracked might be lost. To tackle such difficulty, different views of the object can be learned offline and used for tracking.

#### 4.7. Silhouette tracking

Silhouette tracking is done by approximating the object region in each frame. In this method, the information inside the object region are used, they can be in the form of appearance density and shape models that are usually in the form of edge maps. Given the object models, silhouettes are tracked by either shape matching or contour evolution.

- *Shape Matching*: Shape matching can be done like the template matching tracking, where an object silhouette and related model are searched in current frame. By computing the likeness of the object with the assumed object silhouette model created from the preceding frame, the search is done. The object model is reinitialized to handle appearance variations in each frame after the object is placed.
- *Contour Tracking*: In contour tracking method, an initial contour is elaborated iteratively in the preceding frame to its new position in the current frame. The elaboration needs that some parts of the object in the current frame overlay with the object region in the preceding frame. Tracking by elaborating a contour can be done using two different methods. The first method uses state space models to model the contour shape and motion. The second method directly elaborates the contour by reducing the contour energy using direct reduction methods such as gradient descent.

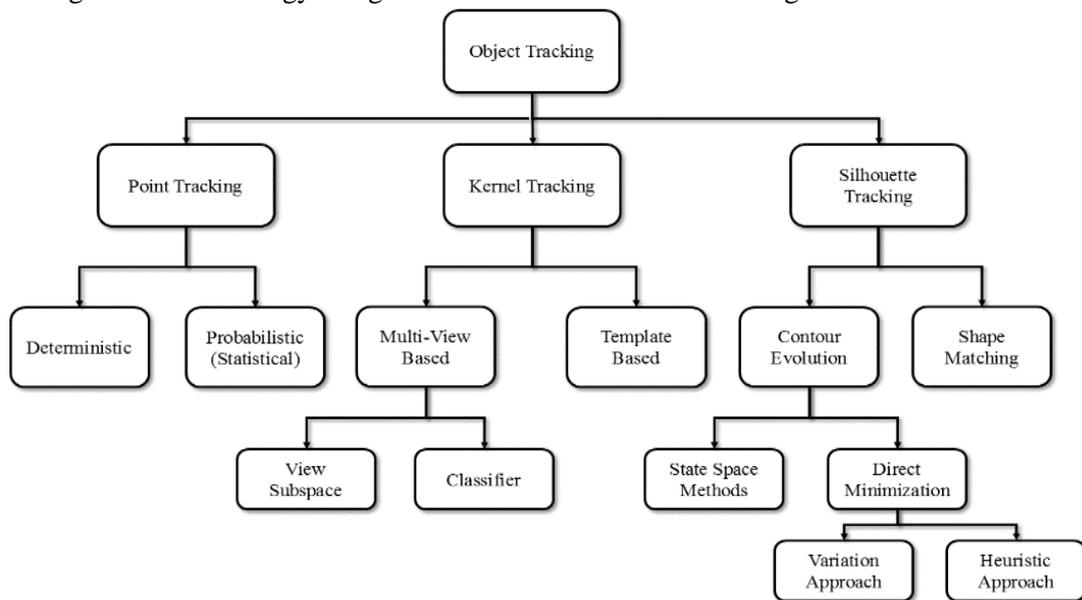


Fig. 2: Taxonomy of Tracking Methods

## 5. Vehicle Counting

Counting is expressed as the act of finding the number of components of a pre-determined set of objects in a defined space area. Conventionally, counting is done by counter increment by a unit for every component of the set, in some order, whereas marking each component, to avoid counting same component again, until no unmarked components are left in the space.

Region of interest (ROI) is a selected region in the image or the video where the object of interest is expected to be analysed. Literally, ROI is a polygonal section from a 2D map as shown in Fig. 3. For vehicle counting, a section of the road can be considered as ROI, only vehicles that enters the ROI can be counted.

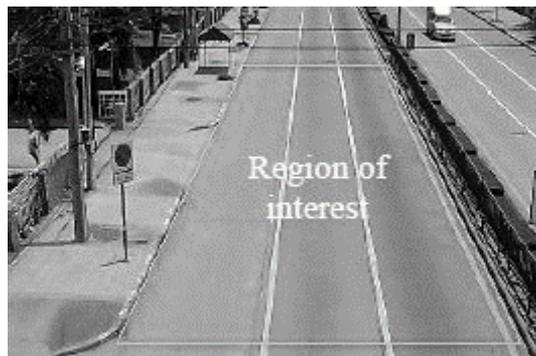


Fig. 3: Region of Interest (ROI) in a vehicle counting system.

There are advantages as well as some disadvantages of using ROI for vehicle detection and counting. The advantages are that it will reduce the computation area and it is highly effective and quick. While the small area in the video frame or image may skip some of the object of interest, in case of vehicle counting it will be difficult to define the ROI in different situation, such as snowy road.

Once the object of interest in the video frame is detected and tracked, now is the time for application. Object counting in the video can be done in many different ways. The common techniques that are being followed for counting vehicles in the video sequences are

### **5.1. Spatio-temporal contour technique**

After detecting foreground pixels in each frame, that is the object of interest, a contour of object is constructed in the spatio-temporal domain. The contour is then analysed to determine the number of objects.

### **5.2. Virtual detection line**

A virtual detection zone is defined within the region of interest. When the object of interest passes through the virtual detection line, the object will be counted.

### **5.3. Virtual loop**

The detected and tracked objects are counted at user defined virtual loops that can be defined at any position of the frame. Object counting is based on the intersection of convex region formed by the convex hull of the particles associated to the tracked vehicle with the polygonal regions defining the virtual loops.

## **6. Conclusion**

In this paper, common methods for object detection, tracking and counting has been discussed giving the interest of object to be tracked has been considered as a vehicle on the road. Starting from the scratch, what one need to follow to do the object analysis in the video sequence from detection of the object of interest then representing the object before tracking and then vehicle counting can be done as the analysis part of the study has been discussed in this paper.

## **7. References**

- [1]. Salvi, G. *An Automated Nighttime Vehicle Counting and Detection System for Traffic Surveillance*. in *2014 International Conference on Computational Science and Computational Intelligence*. 2014.
- [2]. Shi, J. and J. Malik, *Normalized cuts and image segmentation*. *IEEE Transactions on pattern analysis and machine intelligence*, 2000. 22(8): p. 888-905.
- [3]. Wang, L., W. Hu, and T. Tan, *Recent developments in human motion analysis*. *Pattern recognition*, 2003. 36(3): p. 585-601.
- [4]. Yilmaz, A., O. Javed, and M. Shah, *Object tracking: A survey*. *ACM Comput. Surv.*, 2006. 38(4): p. 13.