International Engineering Journal For Research & Development

Vol.4
Issue 6

E-ISSN NO:-2349-0721

Impact factor : 6.03

# MOVING OBJECT DETECTION AND TRACKING IN OBJECT BASED VIDEO SCENES

**Kishor Dhake**
Department of Electronics and Telecommunication Engineering,
Prof Ram Meghe College of Engineering and Management,
Badnera-Amravati, India
kishorgdhake@gmail.com

**Dr. Surendra S. Dalu**
Head, Department of Electronics and Telecommunication Engineering,
Government Polytechnic,
Yavatmal, India
surendradalu@yahoo.co.in

**Mandar Dhake**
3rd year B.E. Department of Computer Science and Engineering,
Prof Ram Meghe College of Engineering and Management,
Badnera-Amravati, India
**mandar2203dhake@gmail.com**

---

***Abstract-***

***In recent years, the automatic object detection and tracking in a video scene is an interesting and most challenging area of computer vision with key applications in video surveillance for public safety against crime, defense technology, vehicle navigation etc. Video surveillance is a technology which works in dynamic environment in various events such as sports, public safety and management of traffic. Given an input video sequence, the system will automatically extract moving objects and will compute their bounding box, their centre of mass, their direction and their size.***

***Keywords—Object detection, object tracking, computer vision, video surveillance, bounding box***

## *INTRODUCTION*

The task of moving object detection in video is the initial step for the process of analyzing videos in computer vision. This is done either in every frame or when the object first time appears in the video. It can be achieved by the elimination of background stationary objects from the moving object of interest. During the detection step, a clustering method is introduced to detect the objects. There is a broad range of object detection system applications which are aided by these, like assisted driving [1, 2], industrial applications [3, 4] or video surveillance [5, 6, 7].

In object classification, the objects are classified as birds, humans, vehicles and other moving objects. The basic principle is to extract image features like HOG, SIFT, SURF [8, 9, 10] which are useful for a generic description of instances of specific object classes. The great challenge is to find the common characteristics of these features to separate multiple object classes.

Object tracking plays a vital role in the field of computer vision [11]. While object tracking involves the selection of Region of Interest (ROI) and keep tracking the motion and its positions from the video frames [12]. Object tracking algorithms have acquired priority due to the availability of highly sophisticated computers, good quality and inexpensive cameras. The basic steps for object detection classification and tracking [13] are shown in figure 1. In the object tracking, the video analysis involves, moving object detection, object classification, frame to frame object tracking.

To identify the moving object difference between two consecutive images is calculated, which is simple and easy to implementation. For various dynamic environments, this method has a strong adaptability, yet the final result of non-stationary object detection is not accurate as it is tough to obtain complete outline of that moving object.

Background modeling is the first step of background subtraction. It is used to obtain the reference model. This reference model is compared with each video sequence to determine possible variation. The existence of moving object is determined by the difference between the current video frame and the pervious frame in terms of pixels [14]. Many methods have been proposed for real-time moving object detection. At present, background subtraction, frame difference and optical flow are commonly used.

Moving object classification has a strong cue of periodic property shown by a non-rigid articulated motion of the object. Optical flow is also helpful for the object classification. Rigidity analysis and periodicity of moving entities are done by residual flow. The non-rigid dynamic objects such as human being can have greater average residual flow and display a periodic component, while the rigid objects are expected to have little residual flow.

The gradient orientation that occurs in the localized portions of an image is counted in texture based technique. It is then computed using overlapping local contrast normalization on a dense grid of uniformly spaced cells for improving accuracy.

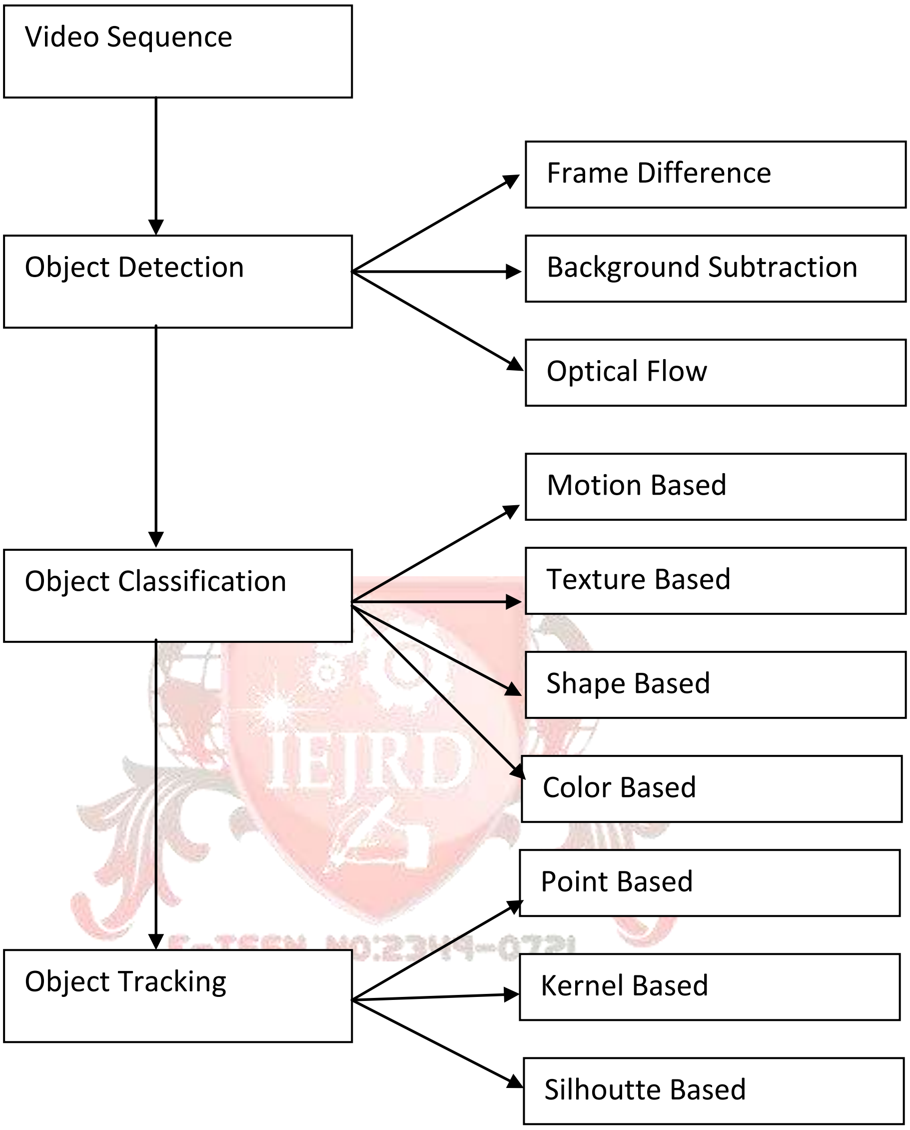


*Fig. 1: Basic steps for object detection and tracking*

Representation of points, box and blob are various motion regions available for classifying the moving objects based on the shape information. The mixture of image based and scene based parameters like image blob area, camera zoom and apparent aspect ratio of blob bounding box acts as the input features to the network. Each blob at every frame undergoes the classification and the histogram keeps the results. In this paper the proposed system will automatically extract moving objects and will compute their bounding box, their centre of mass, their direction and their size.

## PREVIOUS WORK

Detecting and estimating the viewpoint of 3D objects using Part-Based Models have shown great importance. Deformable Part Models (DPMs) as introduced by [15] have shown good results in object pose estimation and detection. DPMs are more generic and with the addition of 3D CAD model, viewpoints in 3D position can be inferred for 3D objects. Though the performance of DPMs are good but they are dependent on manual supervision of human.

Many authors investigated the object detection pose estimation in the context of regression problems uses the regression functions for the representation of local features and spatial arrangements. Similarly in uses cascaded pose regression in which each regress or is dependent on the previous repressors that compute simple image measurement from human annotated training images. Some authors have used statistical manifold modeling for pose and viewpoint estimation.

These approaches focus on the continuous pose estimation and belief that the location of object is given; but getting the clean foreground mask is difficult. Due to aforementioned problem, the performance of pose estimation using regression models is not good. However by using a unified model that jointly estimates and detects can overcome this issue as studied by defining a discriminative model of appearance that is learned from the pose space using structural support vector machines (SVMs).

High definition and wide view sensors are resulting in large volume of videos that then need to be processed and analyzed [16]. Two problems arise from these emerging trends: (1) the available wireless network bandwidth is not always adequate to send data from the platform to ground stations; (2) high workload is imposed on analysts to process the data. One solution to first problem is to perform video compression encoding before transmission to reduce the amount of data to be streamed, and thus meet the requirement of actual channel bandwidth in the wireless network environment. Another solution to second problem is to automate video recognition objects of interest to enable better and timely situational awareness, and thus reduce the video-user workload. However, accurate situational awareness is virtually impossible with the guarantee of object detection is not the scope of the current H.264 standard. In order to maintain acceptable recognition performance, it is necessary to select the appropriate coding parameters to accommodate available channel bandwidth.

The primary aim and contribution of this paper is to present a quantitative relationship between coding parameters and video object recognition performance. These results are suitable for wireless applications, for example, on unmanned intelligence, surveillance and reconnaissance platforms. It is worth noting that the results presented here are discussed for the first time to the best of our knowledge.

In object detection system presented in [17] the common combination of background subtraction and subsequent classification is used. The summary of the method is as follows: Firstly, the frames of an input video are segmented into foreground and background pixels. Using this information, potential objects are cropped from the frame and fed into a CNN for classification. The object class with the maximum activation on the output layer of the network is then assigned to the object while a minimal threshold ensures a high confidence level.

Video surveillances can be performed using three different activities; it can be manual, semi-autonomous or fully autonomous [18]. In Manual video surveillance the video contents are analyzed by human being and this type of Manual systems are mostly used in real time system. But Semi-automated video surveillance system consists of both human intervention and up to some extent of video processing. Most useful examples of this type of system are found in simple moving object detection in video sequence. A fully autonomous system is used for extra ordinary motion in captured video and then sent to the human expert for further analysis.

In case of fully-autonomous system, video sequence is the input which is taken from the source where performed the surveillance system. In such type of system, the system perform all kind of tasks, such as motion detection and tracking, and it also perform the high level of decision making activity such as unusual event detection. Video surveillance system must be able to supports automatically moving object detection and then tracking. Because analyzing of video sequence for a long time by human or by surveillance system is no longer practical and not feasible. But automatic detection of moving object generates batter human intervention.

The remote sensing images have some special characteristics, such as, wide range covering, great information including, short period in observation. Object detection is one of the hot research fields, it has great significance for resource survey, natural hazard monitoring and military object location. However, due to the complexity of remote sensing images, how to extract information effectively and detect objects accurately become the key points of application of remote sensing technology. A lot of methods about object detection and feature extraction have been proposed one after another (e.g., SIFT and HOG), and many of them have been proved very effective when dealing with object detection in natural images [19]. Recently, many of those methods have been introduced into the field of remote sensing. Such as, Bruzzone and Carlin [20] proposed a pixel-based object detection system, Sun [21] has proposed a detection framework based on spatial sparse coding bag of words and Zhang [22] has proposed an extending DTPBM model for object detection in high-resolution remote sensing images. All those methods could perform well for object detection under simple scenes, but due to the limitation of feature representation, the performance drops quickly when transfer it into complex scenes.

## PROPOSED METHODS

A small database is used with video files in mp4 format containing the data to be analyzed for this research work. In proposed methodology a method is discussed that automatically counts the number of moving objects in each frame and a bar plot represents the number of objects in each frame of the sequence as shown in figure 2 (x axis: frame number; y axis: number of objects).

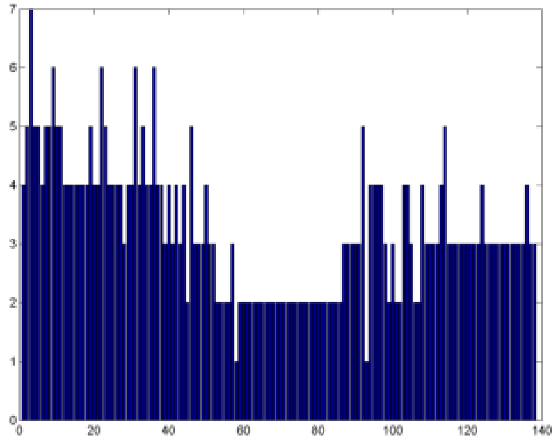

*Fig. 2: Number of objects in each frame*

The function BAR is used to achieve the bar plot. To avoid the interrupt of noise, the median filter is used to process the frames. A method is proposed that will compute for each frame the size, the centre of mass and the average color of each object, and that stores these values in a file. The operation mostly depends on the function region props. Size process uses the 'Area' to obtain the size of object in frames. Mass centre process uses 'Centroid' to obtain the mass centroid of the object. It should be noted that the means of row and volume. Because the first variable in region props is the volume number related to matrix and second variable is the row number of matrix. This figure illustrates the centroid and bounding box.

A frame F(t) with moving objects, a proposed method can find the position of the same objects in frame F(t+1) using the extracted features. The obtained correspondence information used to extract the object direction of motion (angle computed based on the horizontal and vertical displacement) and stored these values in a file. The resulting frames have colored bounding boxes superimposed on the original frames – the same used color for the same object in different frames. Figure 3 shows the objects tracked in various frames.

Frames are selected freely and then bounding the same object using the same color in the frame and the next frame. So actually the object is recognized in just two frames: F(t) and F(t+1) but not in the whole video.

Firstly, obtain the features to get ready to process. The features that can be used: size, position of mass, average color. And the fourth feature is calculated as: angle of motion by using the trigonometry function:

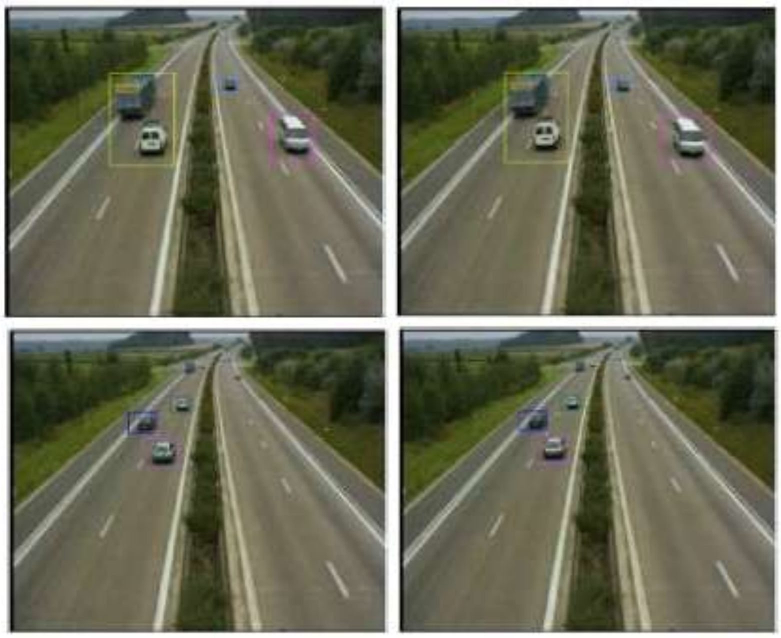

*Fig. 3: Objects tracked in different frames*

$$\text{angle} = \arctan^{((\text{Ynew}-\text{Yoriginal})/(\text{Xnew}-\text{Xoriginal}))} \quad (1)$$

Secondly, using the distance function for all the three features:

$$\text{Distance in space} = ((Y_{new} - Y_{original})^2/(X_{new} - X_{original})^2)^{1/2}; \quad (2)$$

$$\text{Distance in color} = ((R_{new} - R_{original})^2+(G_{new} - G_{original})^2+(B_{new} - B_{original})^2)^{1/2}; \quad (3)$$

$$\text{Distance in size}= ((size_{new} - size_{original})^2)^{1/2}; \quad (4)$$

Thirdly, the threshold is set for each value and it is identified whether the two object in F(t) frame and F(t+1) frame fit these threshold. Then if at least two of them fit the condition, consider it as the same object. If the environment is so badly, all three conditions can be used to limit the process. Another feature can also help us to recognize the object. The same object in two frames always in the same direction. So that it can identified through the angle calculated.

A method is proposed to extract a feature and to save them in a file. Extrema can find the 8 group values to provide the information of extrema positions. 'Extrema' — 8-by-2 matrix that specifies the extrema points in the region. Each row of the matrix contains the x- and y-coordinates of one of the points. The format of the vector is [top-left top-right right-top right-bottom bottom-right bottom-left left-bottom left-top]. This property is supported only for 2-D input label matrices. Proposed method illustrates the extrema of two different regions. In the region on the left, each extrema point is distinct; in the region on the right, certain extrema points (e.g., top-left and left-top) are identical.

## CONCLUSION

In this paper basic steps for object detection, classification and tracking have been discussed in detail. Various methods are presented for extracting moving objects and computing their bounding box, their centre of mass, their direction and their size.

## REFERENCES

H. Cho, Y.-W. Seo, B. Vijaya Kumar, and R. R. Rajkumar. a multi-sensor fusion system for moving object detection and tracking in urban driving environments. in robotics and automation (icra), 2014 ieee international conference on, pages 1836–1843. ieee, 2014.

S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel. Detection of traffic signs in real-world images: The german traffic sign detection benchmark. In Neural Networks (IJCNN), The 2013 International Joint Conference on, pages 1–8. IEEE, 2013.

A. Sharma, R. P. P. Singh, and P. Lehana. Evaluation of the accuracy of genetic algorithms for object detection in industrial environment. International Journal of Scientific and Technical Advancements, 1(3):105–111, 2015.

N. Somani, A. Perzylo, C. Cai, M. Rickert, and A. Knoll. Object detection using boundary representations of primitive shapes. In Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO), Zhuhai, China, 2015.

M. Evans, C. J. Osborne, and J. Ferryman. Multicamera object detection and tracking with object size estimation In Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on, pages 177–182. IEEE, 2013.

D. Varga, L. Havasi, and T. Sziranyi. Pedestrian detection in surveillance videos based on cs-lbp feature. In Models and Technologies for Intelligent Transportation Systems (MT-ITS), 2015 International Conference on, pages 413–417. IEEE, 2015.

X. Wang, M. Wang, and W. Li. Scene-specific pedestrian detection for static video surveillance. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 36(2):361–374, 2014.

N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 1, pages 886–893. IEEE, 2005.

D. G. Lowe. Distinctive image features from scaleinvariant keypoints. International journal of computer vision, 60(2):91–110, 2004.

H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In Computer vision–ECCV 2006, pages 404–417. Springer, 2006.

Alper Yihnaz, Ohio State University, Omar Javed, Object Video, Inc. , Mubarak Shah, University of Central Florida,"Object Tracking: A Survey", ACM Computing Surveys, Vol. 38, No. 4, Article 13, Publication date: December 2006.

Payal Panchal, Gaurav Prajapati, Savan Patel, Hinal Shah and Jitendra Nasriwal a, " A Review on Object Detection and Tracking methods", International Journal for Research in Emerging Science And Technology, Volume - 2, Issue-I , January 2015.

S. Karthikeyan, and S. R. Balaji, “A survey on moving object tracking using image processing,” in proc of IEEE 11 International Conference on Intelligent Systems and Control (ISCO), Feb 2017.

J.Joshan Athanesious, P.Suresh, "Systematic Survey on Object Tracking Methods in Video", Intelllational Journal of Advanced Research in Computer Engineering & Technology (UARCET) October 2012, 242-247.

Shoaib Azam, Aasim Rafique, and Moongu Jeon, “Vehicle pose detection using region based convolutional neural network,” in proc IEEE of International Conference on Control, Automation and Information Sciences (ICCAIS), Jan 2017.

Tao Liu, “More successful recognition: Seeking the relation of video object detection performance with video coding parameters,” in proc of 12th IEEE International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Jun 2016.

Erik Bochinski, Volker Eiselein, Tomas Sikora, “Training a convolutional neural network for multi-class object detection using solely virtual world data,” 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Nov 2016.

Kamal Kant Verma, Pradeep Kumar, Ankit Tomar, "Analysis of moving object detection and tracking in video surveillance system," 2nd International Conference on Computing for Sustainable Global Development (INDIACom), May 2015.

YuShe Cao, Xin Niu, and Yong Dou, "Region-based convolutional neural networks for object detection in very high resolution remote sensing images," 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Oct 2016.

Lorenzo Bruzzone, Lorenzo Carlin, "A multilevel context-based system for classification of very high spatial resolution images", Geoscience and Remote Sensing. IEEE Transactions on, vol. 44, no. 9, pp. 2587-2600, 2006: University Science,1989.

Hao Sun, Xian Sun, Hongqi Wang, Yu Li, Xiangjuan Li, "Automatic target detection in high-resolution remote sensing images using spatial sparse coding bag-of-words model", Geoscience and Remote Sensing Letters IEEE, vol. 9, no. 1, pp. 109-113, 2012.

Wanceng Zhang, Xian Sun, Kun Fu, Chenyuan Wang, Hongqi Wang, "Object detection in high-resolution remote sensing images using rotation invariant parts based model", Geoscience and Remote Sensing Letters IEEE, vol. 11, no. 1, pp. 74-78, 2014.