



# International Journal for Innovative Engineering and Management Research

A Peer Reviewed Open Access International Journal

www.ijiemr.org

## COPY RIGHT



# ELSEVIER

# SSRN

2020 IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 4th Sept 2020. Link

[:http://www.ijiemr.org/downloads.php?vol=Volume-09&issue=ISSUE-09](http://www.ijiemr.org/downloads.php?vol=Volume-09&issue=ISSUE-09)

Title: A STUDY ON OBJECT TRACKING AND IDENTIFYING NAME OF OBJECT USING OPEN CV

Volume 09, Issue 09, Pages: 41-51

Paper Authors

**DR.N. RAGHAVENDRA SAI, DR.PATHAN HUSSAIN BASHA , G.SAI CHAITANYA KUMAR**



USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per **UGC Guidelines** We Are Providing A Electronic Bar Code

## A STUDY ON OBJECT TRACKING AND IDENTIFYING NAME OF OBJECT USING OPEN CV

DR.N. RAGHAVENDRA SAI<sup>1</sup>, DR.PATHAN HUSSAIN BASHA<sup>2</sup>, G.SAI CHAITANYA KUMAR<sup>3</sup>

Assoc. Professor, Department of Computer Science and Engineering, KoneruLakshmaiah Education Foundation,Vaddeswaram, AP, India.

<sup>2</sup>Assoc.Professor and HOD, Department of Computer Science and Engineering, MalineniLakshmaiah Engineering College ,Singrayakonda,Prakasam (Dt),AP,India

### Abstract:

A couple of years back, the formation of the product and equipment picture preparing frameworks was basically constrained to the improvement of the UI, which a large portion of the software engineers of each firm were occupied with. The circumstance has been essentially changed with the appearance of the Windows working framework when most of the engineers changed to taking care of the issues of picture preparing itself. Be that as it may, this has not yet prompted the cardinal advancement in explaining normal errands of perceiving faces, vehicle numbers, street signs, investigating remote and clinical pictures, and so on. Each of these "everlasting" issues is explained by experimentation by the endeavors of various gatherings of the designers and researchers. As present day specialized arrangements are end up being unnecessarily costly, the errand of computerizing the production of the product devices for tackling scholarly issues is detailed and seriously illuminated abroad. In the field of picture preparing, the necessary toolbox ought to be supporting the examination and acknowledgment of pictures of already obscure substance and guarantee the viable advancement of utilizations by common developers. Similarly as the Windows toolbox bolsters the production of interfaces for taking care of different applied issues. Basically, finding an item in progressive casings of a video is called following. The definition sounds straight forward however in PC vision and AI, following is an expansive term that incorporates theoretically comparable yet in fact various thoughts. For instance, all the accompanying distinctive however related thoughts are commonly concentrated under item following. In this paper we are following articles in video and distinguishing name of item utilizing open cv.to actualize this idea here we are utilizing thick optical stream, meager optical stream, kalman sifting, mean shift and cam move calculations.

**Keywords :** Object Tracking ,OpenCV, Python

Article acknowledgment is to portray an assortment of related PC vision errands that include exercises like distinguishing objects in advanced photos. Picture grouping includes exercises, for example, anticipating the class of one item in a

picture. Item limitation is alludes to distinguishing the area of at least one articles in a picture and drawing a flourishing box around their degree. Article discovery accomplishes crafted by consolidates these two errands and

confines and orders at least one items in a picture. At the point when a client or expert alludes to the expression "object acknowledgment", they frequently signify "object recognition". It might be trying for amateurs to recognize distinctive related computer vision tasks.

Thing ID [9] and zone in modernized pictures has gotten one of the most noteworthy applications for endeavors to ease customer, save time and to achieve parallelism. This is definitely not another methodology yet improvement in object area is up 'til now required in order to achieve the concentrated on track even more profitably and unequivocally. The guideline purpose of inspecting and exploring PC vision is to reenact the direct and method of natural eyes authentically by using a PC and later on develop a structure that diminishes human undertakings. PC vision is such kind of examination field which endeavors to see and addresses the 3D information for world articles. Its basic purpose behind existing is reproducing the visual pieces of 3D addresses ensuing to examining the 2D information expelled. Certified 3D objects are addressed by 2D pictures.

The method of thing acknowledgment examination is to choose the number, zone, size, position of the articles in the data picture. Thing area is the essential thought for following and affirmation of articles, which impacts the efficiency and exactness of article affirmation. The essential thing area procedure is the concealing based technique, perceiving objects reliant on their concealing regards [4]. The methodology is used by virtue of its strong flexibility and quality, in any case, the distinguishing proof speed ought to be

improved, in light of the fact that it requires testing each and every possible window by extensive chase and has high computational capriciousness.

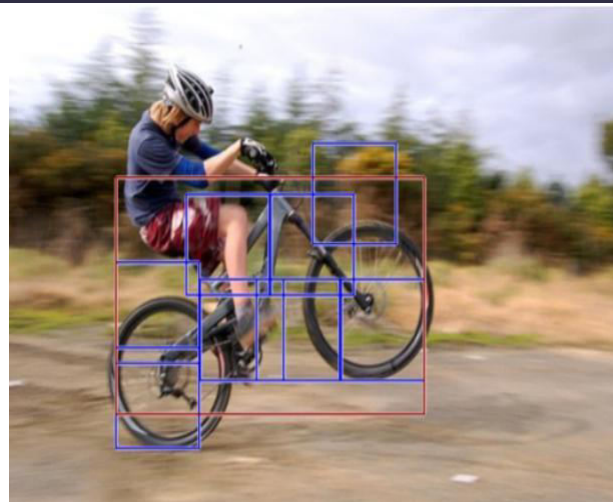
Thing revelation from a confusing establishment is a troublesome application in picture getting ready. The goal of this endeavor is to perceive objects set over a surface from a complicated establishment picture using various frameworks. The area of the things can be extended using robotization and mechanical self-sufficiency for winnowing of the articles like apples, bananas from the contrasting tree using the image taking care of systems and it will be easier, faster and favorable to separate the apples and bananas instead of the manual winnowing.

The standard execution measure that is consistently used for the thing order division issue is brought Intersection-over-Union (IOU) [3]. Given an image, the comparable qualities between the foreseen region and the ground-truth district for an article present in the image can be found with the help of IOU evaluates and can be described as the size of the intersection point apportioned by the relationship of the different territories. For example, if a particular estimation predicts each and every pixel of an image to be its experience, the IOU measure can effectively rebuff for that, as the combination between the envisioned and ground-truth zones would be zero, it will convey an IOU count of zero. OpenCV library executed in python2.7 close by the help of Numpy is used and the universe of thing recognizable proof is researched.

## **2. Background:**

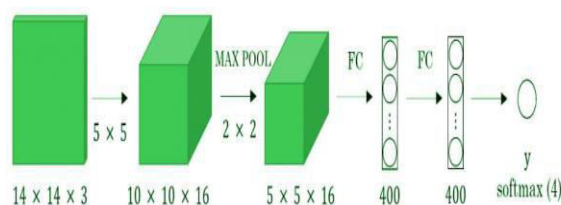
The point of article location is to distinguish all occasions of items from a

referred to class, for example, individuals, vehicles or faces in a picture. By and large, just few cases of the item are available in the picture, yet there is an exceptionally huge number of potential areas and scales at which they can happen and that need to be investigated. Every location of the picture is accounted for with some type of posture data. This is as basic as the area of the article, an area and scale, or the degree of the item characterized as far as a jumping box. In some different circumstances, the posture data is increasingly point by point and contains the boundaries of a direct or non-straight change. For instance for face identification in a face locator may register the areas of the eyes, nose and mouth, notwithstanding the jumping box of the face. A case of a bike identification in a picture that determines the areas of specific parts is appeared in Figure 1. The posture can likewise be characterized by a three-dimensional change determining the area of the article comparative with the camera. Item recognition frameworks consistently develop a model for an article class from a lot of preparing models. On account of a fixed inflexible item in a picture, just a single model might be required, yet more for the most part different preparing models are important to catch certain parts of class inconstancy



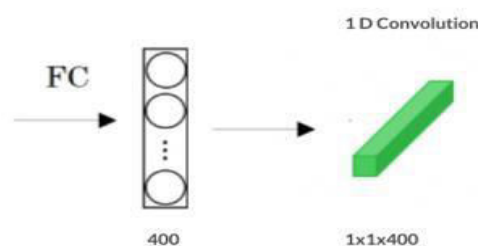
**Figure1**

Convolutional implementation of the sliding windows Before we discuss the implementation of the sliding window using convnets, let us analyze how we can convert the fully connected layers of the network into convolutional layers. Fig. 2 shows a simple convolutional network with two fully connected layer each of shape.



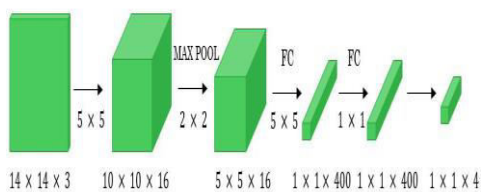
**Figure 2**

A fully connected layer can be converted to a convolutional layer with the help of a 1D convolutional layer. The width and height of this layer is equal to one and the number of filters are equal to the shape of the fully connected layer. An example of this is shown in Fig3.



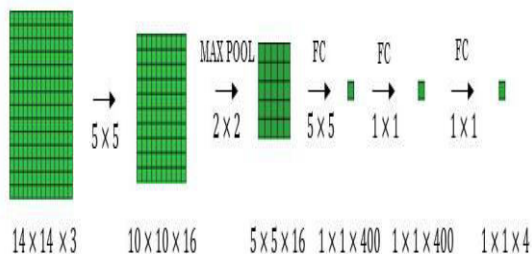
**Figure 3**

We can apply the concept of conversion of a fully connected layer into a convolutional layer to the model by replacing the fully connected layer with a 1-D convolutional layer. The number of filters of the 1D convolutional layer is equal to the shape of the fully connected layer. This representation is shown in Fig4. Also, the outputs of the max layer is also a convolutional layer of shape (1,1,4), where 4 is the number of classes to predict.



**Figure 4**

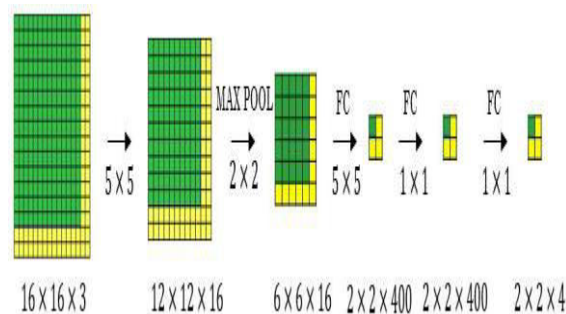
Now, let's extend the above approach to implement a convolutional version of the sliding window. First, let us consider the ConvNet that we have trained to be in the following representation (no fully connected layers).



**Figure 5**

Let's assume the size of the input image to be 16x16x3. If we are using the sliding window approach, then we would have passed this image to the above

ConvNet four times, where each time the sliding window crops the part of the input image matrix of size 14x14x3 and pass it through the ConvNet. But instead of this, we feed the full image (with shape 16x16x3) directly into the trained ConvNet (see Fig. 6). This results will give an output matrix of shape 2x2x4. Each cell in the output matrix represents the result of the possible crop and the classified value of the cropped image. For example, the left cell of the output matrix (the green one) in Fig. 6 represents the result of the first sliding window. The other cells in the matrix represent the results of the remaining sliding window operations.



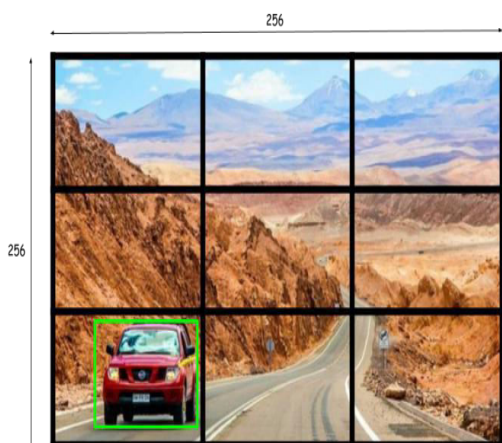
**Figure 6**

The stride of the sliding window is decided by the number of filters used in the Max Pool layer. In the example above, the Max Pool layer has two filters, and for the result, the sliding window moves with a stride of two resulting in four possible outputs to the given input. The main advantage of using this technique is that the sliding window runs and computes all values simultaneously. Consequently, this technique is really fast. The weakness of this technique is that the position of the bounding boxes is not very accurate.

A better algorithm that tackles the issue of predicting accurate bounding boxes while using the convolutional sliding window technique is the YOLO algorithm. YOLO

stands for **you only look once** which was developed in 2015 by Joseph Redmon, Santosh Divvala, Ross Girshick and Ali, Farhadi. It is popular because it achieves high accuracy while running in real-time. This algorithm requires only one forward propagation pass through the network to make the predictions.

This algorithm divides the image in to grids and then runs the image classification and localization algorithm (discussed under object localization) on each of the grid cells. For example, we can give an input image of size 256x256. We place a 3x3 grid on the image (see Fig. 7).



**Figure 7**

Next, we shall apply the image classification and localization algorithm on each grid cell. In the image each grid cell, the target variable is defined as  $Y_{i,j} = [p, c, b, x, y, b, h, w, c, 1, c, 2, c, 3, c, 4]$

Do everything once with the convolution sliding window. Since the shape of the target variable for each grid cell in the image is 1x9 and there are 9(3x3) grid cells, the final output of the model will be:

$$Final\ Output = 3 \times 3 \times 9$$

└───┬───┘
└───┬───┘  
 Number of grid cells      Output label for each grid cell

The advantages of the YOLO algorithm is that it is very fast and predicts much more accurate bounding boxes. Also, in practice to get the more accurate predictions, we use a much finer grid, say 19x19, in which case the target output is of the shape 19x19x9.

### 3. Literature Survey:

In different fields, there is a need to recognize the objective article and furthermore track them successfully while dealing with impediments and other included complexities. Numerous specialists (Almeida and Guting 2004, Hsiao-Ping Tsai 2011, Nicolas Papadakis and Aure lie Bugeau 2010 ) endeavored for different methodologies in object following. The idea of the strategies generally relies upon the application space .Some of there search works which made the development to proposed work in the field of article following are delineated as follows.

#### 3.1 Object Detection

Item recognition is a significant assignment, yet testing vision task. It is a basic piece of numerous applications, for example, picture search, picture auto-comment and scene understanding, object following. Moving article following of video picture successions was one of the most significant subjects in PC vision. It had just been applied in numerous PC vision fields, for example, savvy video reconnaissance (Arun Hampapur 2005), man-made brainpower, military direction, wellbeing discovery and robot route,

clinical and natural application. As of late, various fruitful single-object following framework showed up, however within the sight of a few items, object recognition gets troublesome and when articles are completely or mostly impeded, they are obtruded from the human vision which further expands the issue of identification. Diminishing brightening and obtaining edge. The proposed MLP based article following framework is made hearty by an ideal determination of unique feature sandal so by implementing the Ad a boost strong classification method.

### **3.1.1 Background Subtraction**

The foundation deduction strategy by Horprasert et al (1999), had the option to adapt to neighborhood brightening changes, for example, shadows and features, even globe enlightenment changes. In this technique, the foundation model was factually demonstrated on every pixel. Computational shading mode, incorporate the splendor contortion and the chromaticity mutilation which was utilized to recognize concealing foundation from the customary foundation or moving closer view objects. The foundation and closer view deduction technique utilized the accompanying methodology. A pixel was displayed by a 4-tuple  $[E_i, s_i, a_i, b_i]$ , where  $E_i$  - a vector with expected shading esteem,  $s_i$  - a vector with the standard deviation of shading esteem,  $a_i$  - the variety of the brilliance contortion and  $b_i$  was the variety of the chromaticity twisting of the  $i$ th pixel. In the following stage, the difference between the background picture and the current picture was assessed. Every pixel was at long last grouped into four classifications: unique foundation, concealed foundation or shadow, featured

foundation and moving closer view object. Liyuan Li et al (2003), contributed a technique for recognizing frontal area questions in non-fixed complex conditions containing moving foundation objects. A Bayes choice standard was utilized for order of foundation and closer view changes dependent on between outline shading co-event insights. A methodology to store and quick recover colour co-occurrence insights was likewise settled. In this technique, closer view objects were recognized in two stages. To begin with, both the forefront and the foundation changes are separated utilizing foundation deduction and transient differencing. The incessant foundation changes were then perceived utilizing the Bayes choice guideline dependent on the educated shading co-event insights.

### **3.1.2 Template Matching**

Template Matching is the technique of finding small parts of an image which match a template image. It slides the layout from the upper left to the base right of the picture and contrasts for the best match and the format. The layout measurement ought to be equivalent to the reference picture or littler than the reference picture. It perceives the portion with the most elevated relationship as the objective. Given a picture  $S$  and a picture  $T$ , where the element of  $S$  was both bigger than  $T$ , yield whether  $S$  contains a subset picture  $I$  where  $I$  and  $T$  are reasonably comparative in design and if such  $I$  exists, yield the area of  $I$  in  $S$  as in Hager and Bellhumeur (1998). Schweitzer et al (2011), determined a calculation which utilized both upper and brings bound down to distinguish 'k' best matches. Euclidean separation and Walsh change portions are

utilized to ascertain coordinate measure. The positive things incorporated the use of need line improved nature of choice with respect to which bound-improved and when great matches exist inborn cost was predominant and it improved execution. However, there were requirements like the nonattendance of good matches that lead to line cost and the number-crunching activity cost was higher.

The proposed strategies dint use line along these lines dodging the line cost rather utilized layout coordinating. Visual following techniques can be generally sorted in two different ways to be specific, the element based and area based strategy as proposed by Ken Ito and Shigeyuki Sakane (2001). The element based approach assesses the 3D posture of an objective article to fit the picture includes the edges, given a 3Dgeometrical model of an item. This strategy requires a lot of computational expense. Area based can be characterized into two classes to be specific, parametric strategy and view-based technique. The parametric strategy accept a parametric model of the pictures in the objective picture and ascertains ideal fitting of the model to pixel information in a locale. The view-based strategy was utilized to locate the best match of a district in a hunt region given the reference layout. This has the preferred position that it doesn't require a lot of computational intricacy as in the element based methodology.

## **4. Methodology**

### **4.1 Open CV**

OpenCV (Open Source Computer Vision) is an open source PC vision and AI programming library [10]. OpenCV was at first worked to give a typical framework to

applications identified with PC vision and to build the utilization of machine recognition in the business items. As it is a BSD-authorized item so it turns out to be simple for organizations to use and alter the current code in OpenCV.

Around 3000 calculations are as of now implanted inside OpenCV library, every one of these calculations being productively streamlined. It underpins constant vision applications. These calculations are ordered under great calculations, condition of craftsmanship PC vision calculations and AI calculations. These calculations are effectively executed in Java, MATLAB, Python, C, C++ and so on and are all around upheld by working framework like Window, Mac OS, Linux and Android.

A full-highlighted CUDA and OpenCL interfaces are as a rule effectively created for the improvement of innovation. There are in excess of 500 unique calculations and considerably increasingly such capacities that form or bolster those calculations. OpenCV is composed locally in C++ and has a templated interface that works flawlessly with STL holders.

For OpenCV to work effectively with python 2.7 we have to introduce NumPy bundle first.

### **4.2 OpenCV will utilize following calculations to follow object in recordings**

**4.2.1 Dense Optical stream:** These calculations help gauge the movement vector of each pixel in a video outline.

**4.2.2 Sparse optical stream:** These calculations, similar to the Kanade-Lucas-Tomashi (KLT) highlight tracker, track the area of a couple of highlight focuses in a picture.



**4.2.3 Kalman Filtering:** An extremely famous sign handling calculation used to anticipate the area of a moving article dependent on earlier movement data. One of the early uses of this calculation was rocket direction! Additionally as referenced here, "the on-board PC that guided the drop of the Apollo 11 lunar module to the moon had a Kalman channel".

**4.2.4 Meanshift and Camshift:** These are calculations for finding the maxima of a thickness work. They are likewise utilized for following.

**4.2.5 Single article trackers:** In this class of trackers, the main casing is checked utilizing a square shape to demonstrate the area of the item we need to follow. The article is then followed in resulting outlines utilizing the following calculation. In most genuine applications, these trackers are utilized related to an article identifier.

**4.2.6 Multiple article track discovering calculations:** In situations when we have a quick item finder, it bodes well to recognize different items in each edge and afterward run a track discovering calculation that distinguishes which square shape in one casing relates to a square shape in the following edge.

**4.2.7 Browse System Videos:** Using this module application permit client to transfer any video from his framework and application will associate with that video and begin playing it, while playing on the off chance that application identify any article, at that point it will stamp that object with jumping boxes, while playing video on the off chance that client needs to quit following, at that point he have to squeeze 'q' key from console to stop video

playing.

**4.2.8 Start Webcam Video Tracking:** Using this module application interface itself with inbuilt framework webcam and start video gushing, while at the same time spilling in the event that application identify any item, at that point it will encompass that object with bouncing boxes, while playing press 'q' to stop web cam spilling.

### **4.3 Object Classification In Moving Object Detection**

Item order approach depends on shape, movement, shading and surface. The grouping should be possible under different classes, for example, trees, creatures, people, objects and so on. Following items and examining their highlights is a key idea of article order.

#### **4.3.1 Shape Based**

A blend of picture based and scene-based item boundaries, for example, picture mass (twofold enormous article) zone, the viewpoint proportion of mass jumping box and camera zoom is given as contribution to this identification framework. Grouping is performed based on the mass at every single edge. The outcomes are kept in the histogram.

#### **4.3.2 Motion Based**

At the point when a straightforward picture is given as a contribution without any articles moving, this order isn't required. When all is said in done, non-inflexible enunciated human movement shows an intermittent property, thus this has been utilized as a solid hint for characterization of moving articles. In light of this valuable piece of information.

#### **4.3.3 Color Based**

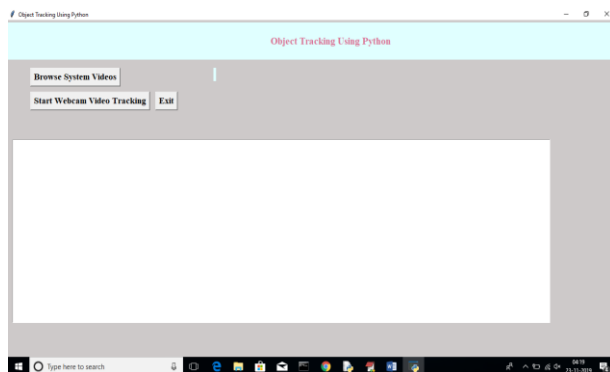
Though color is not an appropriate measure alone for detecting and tracking

objects, but the low computational cost of the color based algorithms makes the color a very good feature to be exploited. For example, the color-histogram-based technique is used for detection of vehicles in real-time. Color histogram describes the color distribution in a given region, which is robust against partial occlusions.

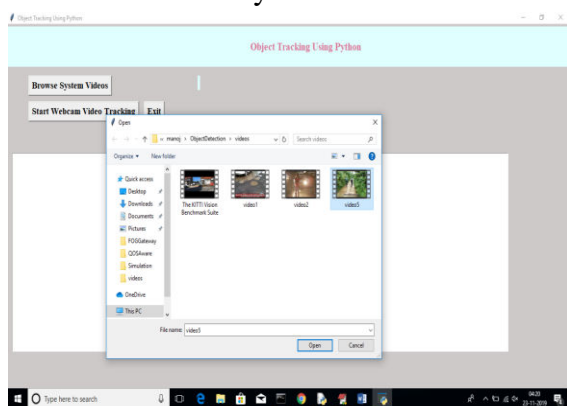
### 4.3.4 Texture-Based

The texture-based approaches with the help of texture pattern recognition work similar to motion-based approaches. It provides better accuracy, by using overlapping local contrast normalization but may require more time, which can be improved using some fast techniques.

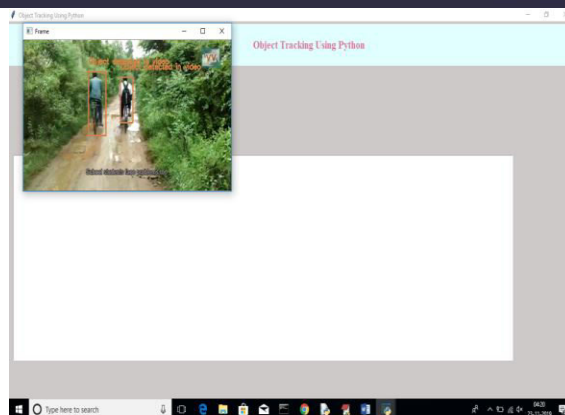
## 5. Results and Discussions



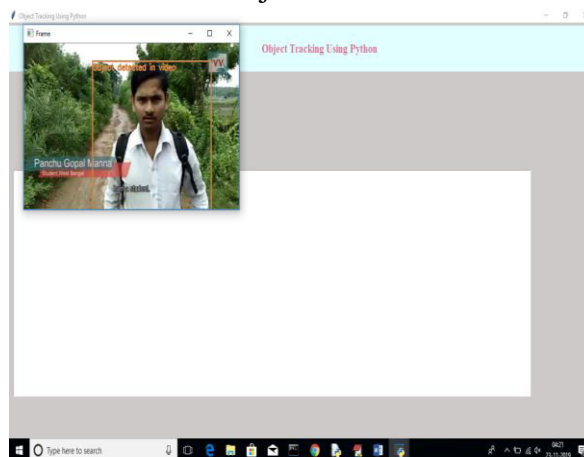
**Fig 5.1** Now click on 'Browse System Videos' button to upload videos from system



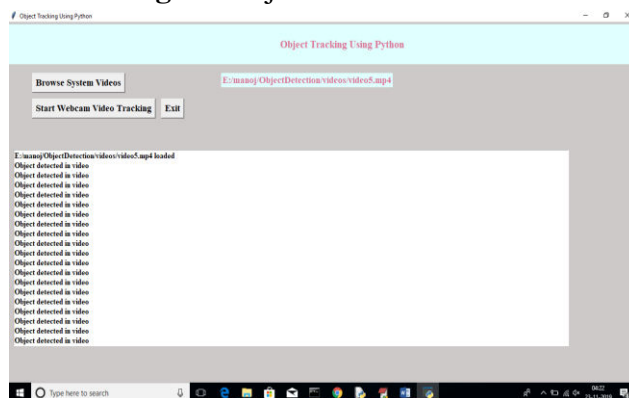
**Fig 5.2** In above screen I am uploading one video, after upload will get below screen



**Fig 5.3** In above video we can see application start tracking objects from video and mark them with bounding boxes. Similarly we can upload any video and track objects from video



**Fig 5.4** Object Detection



**Fig 5.5** In above screen now click on another button called 'Start Webcam Video Tracking' to connect application to web cam and start streaming. After connecting to webcam will get below screen

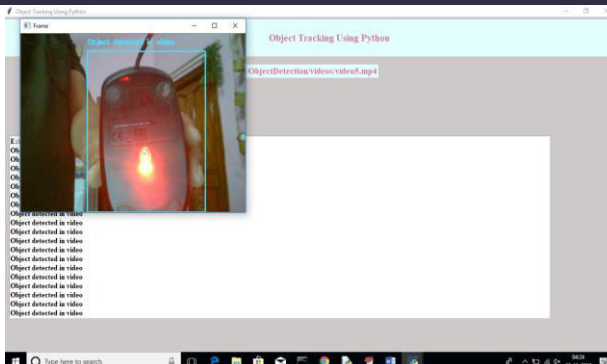


Fig 5.6 In above screen we can see objects is getting tracked from webcam also. In above screen it track computer mouse from web cam video

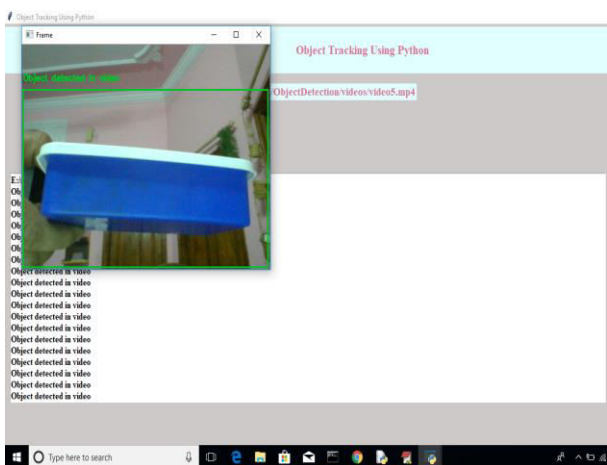


Fig 5.7 Object Detection In Video

## 6. Conclusion

The potential outcomes of utilizing PC vision to take care of genuine issues are huge. The fundamentals of article discovery alongside different methods for accomplishing it and its extension has been talked about. Python has been favored over MATLAB for coordinating with OpenCV in light of the fact that when a Matlab program is run on a PC, it gets caught up with attempting to decipher all that Matlab code as Matlab code is based on Java. OpenCV is fundamentally a library of capacities written in C\C++. Furthermore, OpenCV is simpler to use for somebody with small programming

foundation. Along these lines, it is smarter to begin examining on any idea of article recognition utilizing OpenCV-Python. Highlight understanding and coordinating are the significant strides in object identification and ought to be performed well and with high precision.

## Reference

- [1] KhushbooKhurana and ReetuAwasthi, "Techniques for Object Recognition in Images and Multi-Object Detection", (IJARCET), ISSN:2278-1323,4th, April 2013.
- [2] Latharani T.R., M.Z. Kurian, Chidananda Murthy M.V, "Various Object Recognition Techniques for Computer Vision", Journal of Analysis and Computation, ISSN: 0973-2861.
- [3] MdAtiqurRahman and Yang Wang, "Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation," in Object detection, Department of Computer Science, University of Manitoba, Canada, 2015.
- [4] Nidhi, "Image Processing and Object Detection", Dept. of Computer Applications, NIT, Kurukshetra, Haryana, 1(9): 396-399, 2015.
- [5] R. Hussin, M. RizonJuhari, Ng Wei Kang, R.C.Ismail, A.Kamarudin, "Digital Image Processing Techniques for Object Detection from Complex Background Image,"Perlis, Malaysia: School of Microelectronic Engineering, University Malaysia Perlis, 2012.
- [6] S.Bindu, S.Prudhvi, G.Hemalatha, Mr.N.RajaSekhar, Mr. V.Nanchariah, "Object Detection from Complex Background Image using Circular Hough Transform", IJERA, ISSN: 2248-9622, Vol. 4, Issue 4(Version 1), April 2014, pp.23-28.



[7] Shaikh, S.H; Saeed, K, and Chaki.N,“Moving Object Detection Using Background Subtraction” Springer, ISBN:978-3-319-07385-9.

[8] Shijian Tang and Ye Yuan,“Object Detection based on Conventional Neural

Network”.

[9](2017,January 17).Object Detection

[Online]. Available:

[http://en.m.wikipedia.org/wiki/Object\\_detection](http://en.m.wikipedia.org/wiki/Object_detection)