A REVIEW ON
OPTICAL FLOW-BASED OBJECT DETECTION AND TRACKING USING MATLAB

By :-

Anshuman Agarwal[1] **Shivam Gupta**[2]

[1]Sophomore, Computer Science & Engineering,
HBTI, Kanpur
[2]Sophomore, Electronics & Communication Engineering,
HBTI, Kanpur

Under the guidance of
Er. Dushyant Kumar Singh,
Assistant Professor,
Computer Science & Engineering Department,
Motilal Nehru National Institute of Technology, Allahabad, UP (India)

ABSTRACT

Object detection in a series of frames plays a vital role in the field of computer vision. The aim of this research paper is to detect objects that are in motion with respect to the background that is still. The detection of an object in a series of video frames is determined by the generated optical flow vectors instead of the concept that uses subtraction of the image background. The method put forward below is based on optical flow vectors (*Lucas Kanade)* is facile and as demonstrated below, would be implemented without any complications. The program is written in a way that it is easily modifiable to be used in specific application areas. It is tested on datasets available online, real time videos and also on videos recorded manually. There are various applications of this candid approach of detecting objects using the above mentioned technique. Some of the applications of this technique are to use it for interaction between machines and humans, security and surveillance, supplemented authenticity, monitoring traffic in roads, medicinal image processing and other video processing. For removal of noises in the areas other than the moving objects, median filter is used and the unwanted objects are removed by applying the algorithms based on thresholding in morphological operations. Also the blob analysis is required for setting the object type restrictions. The results show that the proposed system is successfully detecting and tracking moving objects in different types of videos.

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Background

Image is nothing but a combination of pixels which are spread around on the window in a regular pattern and that each point in a pixel has an intensity value that comprises an image as a whole. This program does not only try to make the computer see the image but also detect specific details in the image. Living beings can observe the image by many characteristics of it and use our previous experiences in detecting the image. But for a machine, an image is nothing but a two dimensional array of pixel intensities. So we have to devise techniques and methods to achieve this goal.

There are many articles and reports on the topic of object detection throughout. This field has also a promising future as it would help to automate various processes with least human involvement. In the many articles and researches that have already taken place on this topic, this problem of making the computer learn this as humans do and track and detect objects and also segment them have been made possible. We would restrict ourselves to the concept of detecting object that are in motion with respect to the background from a stationary camera.

There were many algorithms proposed for the above tasks which are listed below:

- Frame differencing
- Viola Jones detection
- Skin colour modelling

In an image a distinct boundary that separates two homogenous regions is taken as an edge and Frame differencing and edge detection algorithm will subtract the two consecutive frames based on these edges. If the difference comes out to be non-zero values, it is considered to be moving otherwise not. But it has some limitations that during capturing the video due to the movement in air or any other source might cause the disturbance in the position of the camera resulting into the false detection of the non-moving objects.

The Viola-Jones algorithm uses Haar-like features, that is, a scalar product between the image and some Haar-like templates. Although it can be trained to detect a variety of object classes, it was motivated primarily by the problem of face detection. But it has some limitations like the detector is most effective only on frontal images of faces and it is sensitive to lighting conditions.

The preliminary steps in skin detection are the representation of image pixels in colour spaces, suitable distribution of skin and non skin pixels, and after that skin colour modelling. According to skin colour's distribution characteristics on colour space, skin colour pixels can be detected quickly with skin colour model. But it has obvious disadvantage like skin colour also varies from person to person belonging to different ethnic groups and from persons across different regions and many others.

1.2 Problem Statement

Since long time, it has been a very tedious job to detect the objects in a series of frames. As mentioned above there have been various methods for object detection but they all have many shortcomings. Major objective here is to design a program that would segment moving objects in the frame with that of the background motion thus successfully generating optical flow vectors, a binary image with moving objects clearly segmented and finally enclosing the moving objects in a block.

1.3 Motivation

Optical flow vectors can be calculated through various types of algorithms, but the most popular among them is Lucas-Kanade (LK) method which was introduced in 1981 (Lucas and Kanade, 1981). In this method both Lucas and Kanade concluded that the motion can be indicated by optical flow and its flow and thereby producing accurate results theoretically. As mentioned above, this method has various applications in almost every area, yet it is not widely implemented as optical flow requires special optimized and designed system. Still, as it would be explained further, the accuracy of optical flow and the very idea behind it make it an authenticated perfect solution for motion detection. Research on optical is being conducted since a long time and various improvements have been made, but there has not been a very clear and elucidative presentation to this approach. Therefore it is required to re-organize its practicability and performance, which is one of the motives in this research project.

MATLAB is becoming very demanding nowadays in computer vision field especially for its library that are mainly developed for Image Processing. Another important feature is that all the libraries which are in-built are platform-independent, which are compatible with all kinds of Operating systems like Windows, Linux and MacOS. Such high portability minimizes burdens and complexity for the programmers to install various versions of the image processing toolbox. . Nowadays, MATAB has humongous collection of popular computer vision computation algorithm, among which optical flow is the best.

1.4 Assumptions

- The first assumption here is that brightness will remain constant from time to time, or only change smoothly and slowly from time to time. Drastic increase on brightness can have adverse effects on the optical flow designed to run and compute, thereby resulting in improper detection and tracking on moving objects. But optical flow still has the capability to track movement if the changes of brightness are smooth and slow from time to time.
- The nearby points in the image plane move in a similar manner (velocity-smoothness constraint).

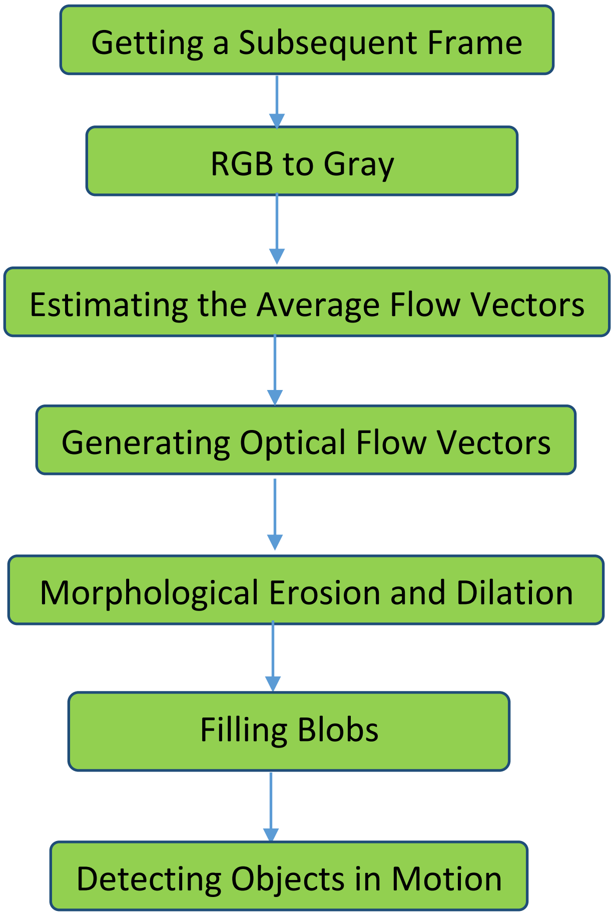


Fig. 1. *Block diagram of the proposed algorithm*

2. LITERATURE REVIEW

2.1 Optical flow (i)

Now the Optical Flow method which is used to detect the moving objects uses the concept of vectors to detect motion in two subsequent frames of a video. It is known that vectors both have a magnitude and direction so the moving parts of the two frames are segmented from the still background and are depicted by vectors directing towards the direction and motion. The original surroundings are in a three-dimensional coordinate system with time as an extra variable but for simplicity it is converted into a two-dimensional coordinate system with time as the third variable.

So finally a function in a two dimensional environment with variables as I (x,y,t) is obtained. It is a function of dynamic brightness of time and location. As already mentioned in the assumptions before, the pixels of the moving object are changed due to which the motion vectors are generated, so the brightness intensity near those pixels should remain constant with respect to the frames. So the following expression can be used,

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \qquad (1)$$

Using Taylor series for the right hand part of (1)
Obtain:

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x,y,z) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t + H.O.T \qquad (2)$$

From (1) and (2), with neglecting higher order terms (H.O.T.) and after modifications,

$$I_x.v_x + I_y.v_y = -I_t \qquad (3)$$

Or in formal vector representation

$$\nabla I^T.\vec{v} = -I_t \qquad (4)$$

Where ∇I is so-called the spatial gradient of brightness intensity and $\vec{v}$ is the optical flow vector of the image pixel and I_t is the time derivative of the brightness intensity.

Equation (4) is most important for optical flow calculation and is called 2-D Motion Constraint equation or gradient Constraint. It can be seen that it is a single equation with two unknown variables and is known as the famous aperture problem.

It is important to note that the optical flow method is computationally very expensive and is very slow. It needs faster processors to run optimally. Optical flow vectors are estimated by a number of techniques and the two most popular ones are mentioned below. All these techniques come from the first equation and the assumption that the brightness intensity of the neighbourhood pixels should remain constant stays. Main aim here would be to obtain optical vectors using the differential equations only.
The two methods discussed above are given below and the first method which is used in the project is described further in detail:-
•Lucas-Kanade
•Horn-Schunck

2.2 Lucas-Kanade Method

$$\rho_{LK} = \sum_{x,y\in\Omega}[W^2(x,y)[\nabla I(x,y,t).\vec{v} + I_t(x,y,t)]^2] \qquad (5)$$

Where Ω is the neighbourhood of the pixel; W (x,y) are the weights allocated to individual pixels in Ω (typically 2-D Gaussian coefficients).To determine the minimum error it is mandatory to calculate the derivation of the error term ρ_{LK} by individual components of velocity and equating the result equal to zero. Finally, a matrix form of the expression for the optical flow is generated as follows:

$$\vec{v} = [A^T W^2 A]^{-1} A^T W^2 \vec{b} \quad (6)$$

For N pixels (N = n2, for nxn of Ω neighbourhood) and (x_i,y_i) $\in\Omega$ at time t:

$$A = [\nabla I(x_1,y_1),\ldots\ldots.,\nabla I(x_N, y_N)] \quad (7)$$

$$W = \text{diag}\,[W(x_1,y_1),\ldots\ldots.,W(x_N, y_N)] \quad (8)$$

$$\vec{b} = -\,[I_t(x_1,y_1),\ldots\ldots., I_t(x_N, y_N)] \quad (9)$$

So the expression for the resultant velocity for one pixel will be obtained as the solution of the system (6). The convolution by averages of Gaussian or difference temporal gradient filter is preferred over the calculation of the sums to reduce the algorithm complexity.

3. NOISE FILTERING

Any video that is captured with a still camera, it is associated with a variety of noises that may not be visible in the original video but are clearly seen in the binary frame of the video. This is an obstruction in detecting moving objects as it leads to sudden changes in the intensity levels of the pixels in the neighbourhood the desired pixels of the objects that are in motion. So either a very high or a very low intensity levels are observed.

So there is an important need to eliminate these noises out of the binary frame for image segmentation. To remove the noises, the sources of these should be known which are:

1. Slight motion of the camera due to which the still background seems to be moving.
2. Motion in the background due to environmental conditions like wind, rain, etc.

With these noises, moving objects will never be clearly detected so a technique known as Filtering is used to remove these noises. The major objective of these algorithms is to identify areas of homogeneity and eliminate them, smooth the edges of the objects which are helpful in further processing and to remove the sudden intensity changes in the background due to noises.

Signal filter belong to the class which are helpful to detect and suppress these noises. It is wide range and one of its sub-type known as the median filters is used in this report. As shown in the block diagram, the image is converted into a gray scale image, so what the median filter does is that it replaces the value of the processing pixel by the median value of the block of pixels that is declared previously. Median filters are an excellent choice as the effectively remove the unnecessary noises and smoothen the binary image which can be processed easily further.

4. IMAGE SEGMENTATION

An important prospect in the next step is to connect the optical flow vectors carefully and to segment them into different object which is determined by some thresholding values. The final aim is to separate objects into blobs and to fill the holes in the blobs

Thresholding function is applied for the conversion of the gray image to binary and consequently the moving objects are separated distinctly. So after applying filtering and the thresholding function, the moving parts of the background, detected as objects are removed.

5. MORPHOLOGICAL OPERATIONS

In the binary image of the moving object obtained there are some improper boundaries, so Morphological operations are performed to extract significant features from those images that are useful in the representation and description of the shapes in those moving region. It is widely used in image segmentation and pattern recognition. Morphological closing and eroding is done respectively to remove portions of the objects which are unnecessarily detected as moving.

In morphological eroding object, the structuring element is taken as a square. Let E be a Euclidean space or an integer grid, and A a binary image in E. The erosion of the binary image A by the structuring element B is defined by:

$$A \ominus B = \{ z \in E \mid B_Z \subseteq A \}$$

Where:

B =

0	0	0
0	0	0
0	0	0

The morphological closing operation and the applied structural element C is described as follows:

$$A \bullet C = (P \oplus C) \oplus C$$

Where:

C =

1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1

The matrix P which includes information of moving object is obtained through threshold segmentation of the image.

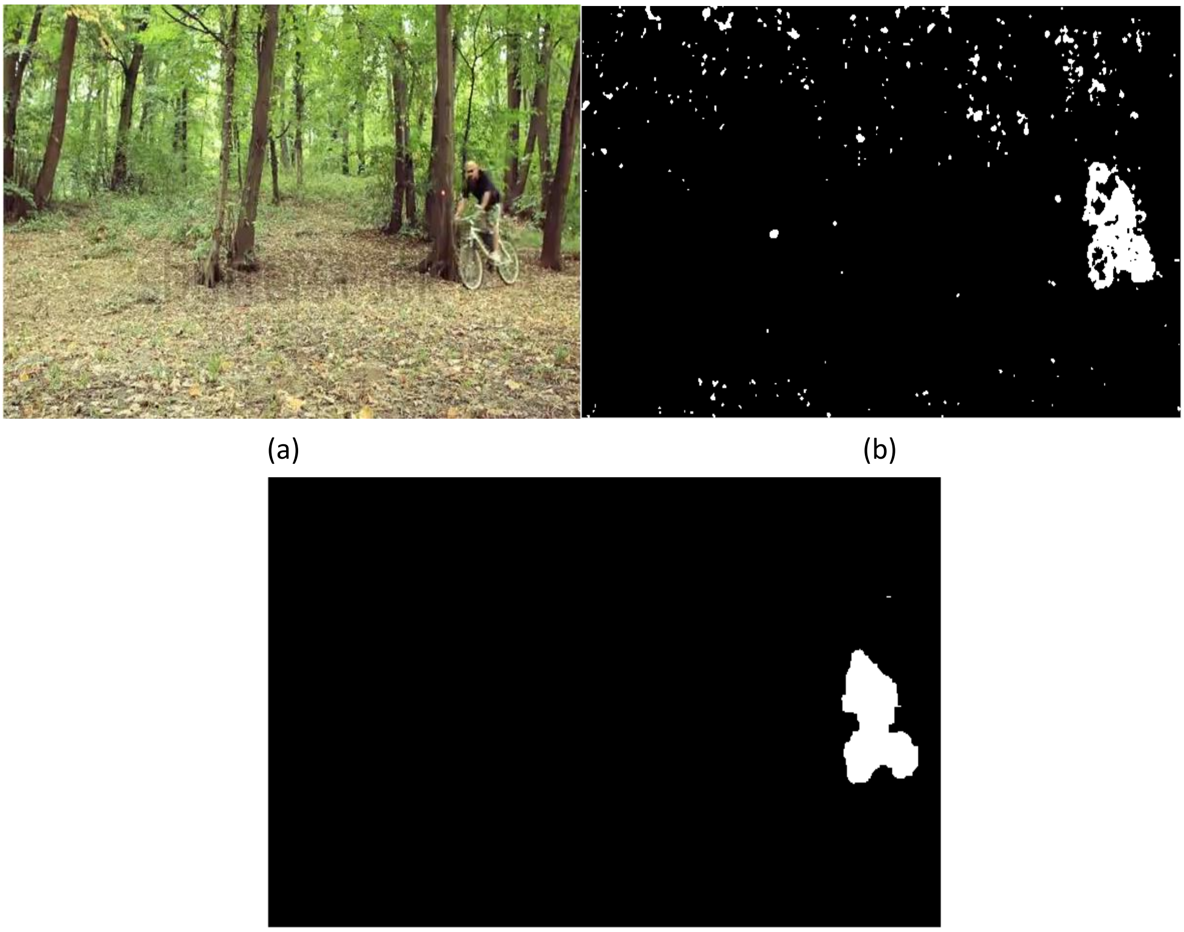
6. BLOB ANALYSIS

Blob analysis is useful in detecting any type of 2- dimensional shapes in an image. The detection uses the concept of spatial characteristics through particular criteria. In many applications where the computation is time taking, there are some unwanted objects in the series of frames that are detected, so blob analysis is used to eliminate those blobs based on certain spatial characteristics and consider only the relevant blobs for required analysis. Only those blobs which satisfy our system as the moving objects are considered. Rest of the unnecessary blobs are eradicated by setting limitations on the relative features in the algorithm and holes are filled in the required blobs.

7. EXPERIMENTAL RESULTS

This part sums up all the various steps that were discussed above in form of results of an experiment using a standard dataset of woods in which a man is riding a cycle. It is also tested on other datasets from traffic closed-circuit TV (CCTV), real-time videos and other manually recorded videos to analyze the working of the program.

As already mentioned above in the block diagram showing various operations, each frame that is extracted from the video is converted from the RGB colorspace to Gray colorspace (Fig. 3 (a)); as applying operations like filtering, thresholding and other operations work well and give better results in this colorspace. So after the conversion of colorspace, optical flow is implemented using the Lucas-Kanade method to determine the optical flow vectors. The image is then filtered to smoothen it and remove blurriness. After this, a binary image is obtained with moving object as white while rest of the background as black. On looking at and comparing the Fig. 2 (a). And Fig. 2 (b), in addition to the moving object, other white spots can also be seen and are detected as objects like the noises of moving leaves in the background. Now to remove these, we apply morphological operations, namely Dilation and Erosion to both enhance the object boundary and remove the noise spots using a square as a structuring element (Fig. 2 (c)).

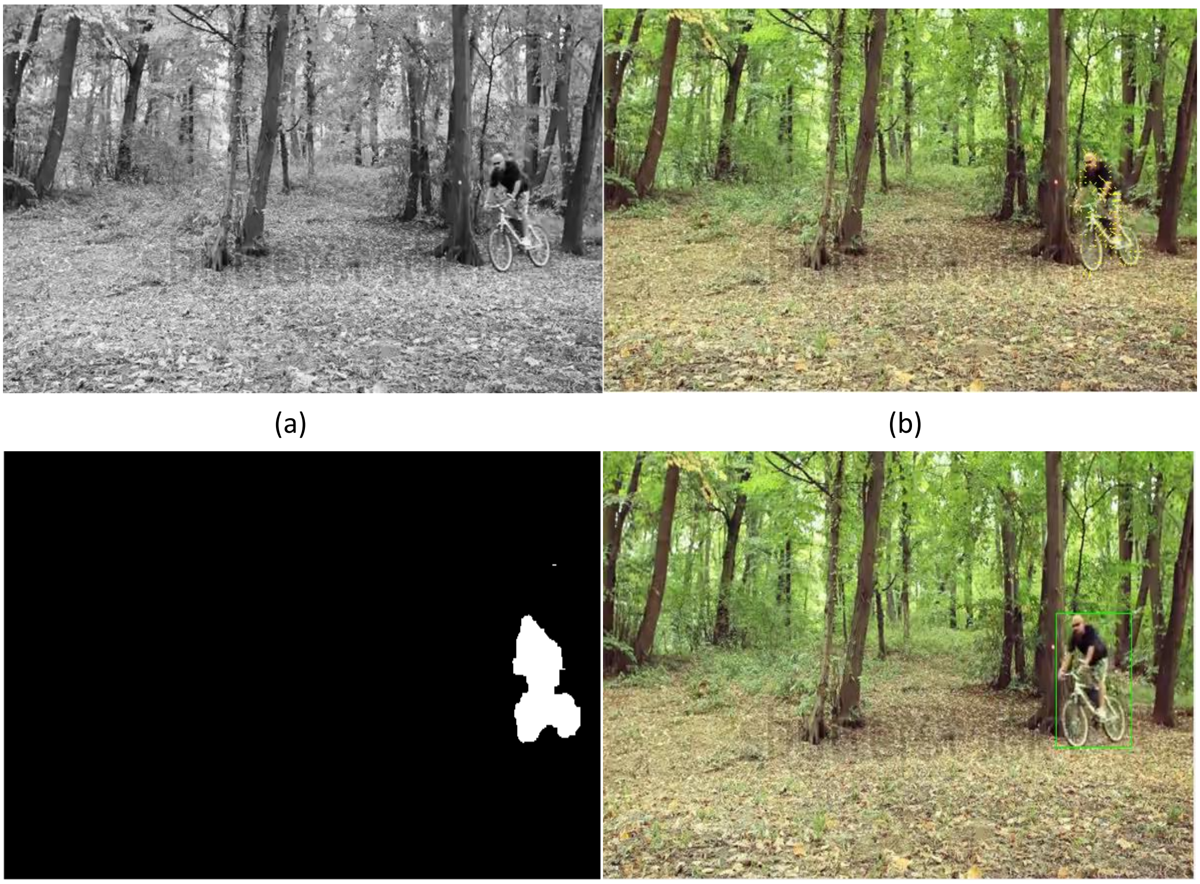


(a) (b)

(c)

Fig. 2. *An extracted frame (a) Original Video (b) Binary Image Without Morphological Close (c) Binary Image after applying Morphological Dilation and Erosion.*

The algorithm used continues with blob analysis to detect particular objects, and remove unnecessary blobs depending upon the sizes like rectangle, square in the background and features like motion (Fig. 3 (c)). The blobs are created according to our need that what kind of objects are to be detected based on sizes, etc. and the constraints applied on them. One of the most important blob constraints is the density of the blob which is the amount of pixels that constitutes one blob. Another blob constraint is connectivity coefficient, which describes the number of pixels which are inter-linked to each other. So the accuracy is directly proportional to both density and connectivity coefficient and the computation is also time consuming. The boxes of fixed sizes are generated around the developed blob and are ready for the next step.



(a) (b) (c) (d)

Fig. 3. *A forest video frame (a) Gray Scale Image (b) Optical Flow Motion Vectors (c) Binary Image with Blobs (d) Tracking Moving Object with Boundary Boxes*

The last stage of the algorithm is the detection and tracking of moving objects that is shown above in the image (Fig. 3 (d)). Prior to detection of the moving object i.e., the man on cycle, make sure whether all of the generated bounding boxes are containing the required objects or not (in our case the above mentioned man on cycle). In this algorithm, when the area of the bounding box is below 1200, it is classified as the object.

8. PRECISION AND RECALL

Precision actually depends on the identifications that the program has made and then to estimate how many of them are correct.

Recall on the other hand depends on the identification made by the program with respect to the original video and then to estimate how many of them actually are correctly identified.

Mentioned below are the mathematical formulae to estimate the precision and recall for any object identification algorithm.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

- True Positive is the number of objects that are correctly identified.
- False Positive are the objects that are detected by the algorithm but they are incorrect identifications.
- False Negative are those number of objects which should have been identified by the program but they are not detected.

All the datasets that have been tested do not detect objects that are not correct, (here stationary objects are not detected due to the constraints that have been imposed).

Video name	Number of Frames	True positive (%)	False positive (%)	False negative (%)	True Negative (%)	Precision (%)	TPR (%)	FPR (%)
Record.mp4	75	88.89	6.11	11.11	93.89	93.47	88.89	6.11
Car.mp4	100	78.75	9.58	21.25	90.42	89.15	78.75	9.58
Vissiontraffic.mp4	200	86.73	7.50	13.27	92.50	92.04	86.73	7.50

Table 1. *Representation of Precision, TPR, FPR in different videos.*

*Where **TPR** stands for True positive rate also known as Recall and **FPR** stands false positive rate also known as Fall-out.*

9. CONCLUSION

In the end, to conclude, it would be important to state that the algorithm mentioned above will be able to detect and track the moving object in the sequence of video frame taken from the static camera in any kind of background and terrain. In every subsequent frame initially the average flow vectors are estimated and then the generation of optical flow vectors takes place. For the better accuracy of the detection morphological erosion and dilation is performed. Lucas-Kanade has been chosen for the estimation of optical flow because of its high accuracy and its basic principle that uses the change of intensity between two consecutive video frames for motion detection. Now the filtering is done to smooth out the boundaries of the moving object using median filters. Morphological closing is applied to the thresholding image obtained to thin out the parts that are detected as objects. As it is already known that the optical flow method is computationally expensive, so finally filling of holes is done using blob analysis as it is faster. Ultimately as shown in the figures, the algorithm will detect only those moving objects that will satisfy the restrictions applied on the blob areas rest will remain as undetected.

10. REFERENCES

i. https://en.wikipedia.org/wiki/Precision_and_recall

ii. Revathi, R., and M. Hemalatha. "Certain Approach of Object Tracking using Optical Flow Techniques." International Journal of Computer Applications 53, no. 8 (2012).

iii. Klappstein, J., Vaudrey, T., Rabe, C., Wedel, A. and Klette, R., 2009, January. Moving object segmentation using optical flow and depth information. In Pacific-Rim Symposium on Image and Video Technology (pp. 611-623). Springer Berlin Heidelberg.

iv. S. Hari Kumar, P.Sivaprakash,” New Approach for Action Recognition Using Motion based Features”, Proceedings of 2013 IEEE Conference on Information and Communication Technologies (ICT 2013).

v. Ikizler-Cinbis, N. and Sclaroff, S., 2010, September. Object, scene and actions: Combining multiple features for human action recognition. InEuropean conference on computer vision (pp. 494-507). Springer Berlin Heidelberg.

vi. Zhou, D. and Zhang, H., 2005, October. Modified GMM background modeling and optical flow for detection of moving objects. In 2005 IEEE International Conference on Systems, Man and Cybernetics (Vol. 3, pp. 2224-2229). IEEE.

vii. Yamaguchi, K., Kato, T. and Ninomiya, Y., 2006, August. Vehicle ego-motion estimation and moving object detection using a monocular camera. In 18th International Conference on Pattern Recognition (ICPR'06) (Vol. 4, pp. 610-613). IEEE.

viii. Sepehr Aslani, Homayoun Mahdavi-Nasab,” Optical Flow Based Moving Object Detection and Tracking for Traffic Surveillance”, World Academy of Science, Engineering and Technology International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering Vol:7, No:9, 2013.

ix. Denman, S., Fookes, C. and Sridharan, S., 2009, December. Improved simultaneous computation of motion detection and optical flow for object tracking. In Digital Image Computing: Techniques and Applications, 2009. DICTA'09. (pp. 175-182). IEEE.

x. Zhan, C., Duan, X., Xu, S., Song, Z. and Luo, M., 2007, August. An improved moving object detection algorithm based on frame difference and edge detection. In Image and Graphics, 2007. ICIG 2007. Fourth International Conference on (pp. 519-523). IEEE.

xi. Han, H. and Tong, M., 2013, July. Human detection based on optical flow and spare geometric flow. In Image and Graphics (ICIG), 2013 Seventh International Conference on (pp. 459-464). IEEE.

xii. Kale, K., Pawar, S. and Dhulekar, P., 2015, September. Moving object tracking using optical flow and motion vector estimation. In Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions), 2015 4th International Conference on (pp. 1-6). IEEE.

xiii. Ms. Shamshad Shirgeri, Ms. Pallavi Umesh Naik, Dr.G.R.Udupi, Prof.G.A.Bidkar4, " Design and development of Optical flow based Moving Object Detection and Tracking (OMODT) System", International Journal of Computational Engineering Research||Vol, 03||Issue, 4||

xiv. Chaple, M. and Paygude, S.S., 2013. Vehicle Detection and Tracking from Video Frame Sequence. International Journal of Science & Engineering Research, 4(3), pp.1-7.

xv. Gonzalez, Rafael C., and E. Richard. Woods, "digital image processing," Prentice Hall Press, ISBN 0-201-18075-8. 2002.