Indian Journal of Natural Sciences www.tnsroindia.org.in ©IJONS

Vol.13 / Issue 75 / December/ 2022 *International Bimonthly (Print)* *ISSN: 0976 – 0997*

RESEARCH ARTICLE

Neural Network System based on Real time Object Detection and Recognition for Video Surveillance Systems

Raja Reddy Duvvuru[1]*, K. Anitha Reddy[2], P. Chandana Priya[3], K.Vimala Kumar[4] and B. Sudhrshan Reddy[5]

[1]Associate Professor, EEE Department, Malla Reddy Engineering College, Secundrabad, Telangana, India.
[2]Assistant Professor, EEE Department, Malla Reddy Engineering College, Secundrabad, Telangana, India.
[3]Assistant Professor, Department of CSE, Malla Reddy University, Secundrabad, Telangana, India.
[4]Assistant Professor, Department of EEE, JNTUKUCEN, Narasaraopet, Guntur, India.
[5]Professor, CIVIL Department, Malla Reddy Engineering College, Secundrabad, Telangana, India.

Received: 19 Aug 2022 Revised: 16 Sep 2022 Accepted: 22 Oct 2022

***Address for Correspondence**
Raja Reddy Duvvuru,
Associate Professor,
EEE Department,
Malla Reddy Engineering College,
Secundrabad, Telangana, India
Email: rajajntuacep@gmail.com

ABSTRACT

In this work, one of the crucial issues that many of the previous works are concentrating on is the recognition and tracking of the item. Due to the increasing demand for video surveillance system applications including traffic control, medical image processing, and satellite image processing, object recognition and tracking are particularly well-liked. This method is also among the most potent ones used in applications based on artificial intelligence, machine learning, and computer vision. Understanding the type of images, attributes, locations of each image in space, and tracking the movements of each object while it is moving are the main goals of these fundamental object recognition based systems. Since human identification has received so much attention in the research that has already been done, many object detection apps focus primarily on it. This section outlines a novel method for object recognition that uses the CNN methodology to recognise both live and non-living objects. The major goal of this section is to provide a framework for classifying items into living and non-living categories. Once they have been discovered, they can then be categorised using the support vector machine technique, which is useful for identifying theft utilising surveillance systems.

Keywords: CNN, Image processing, SVM.

INTRODUCTION

In modern Object recognition is one of the main challenges facing most computer vision work. Increasing demands on surveillance, security, traffic management and medical imaging are particularly popular in object detection and tracking [1]. It is also a product of strong algorithms in machine learning, computer vision and hardware advances that enable a couple of minutes of highly data-intensive calculations. The ultimate aim of the vision-based detection is to understand the type of objects in the picture, their characteristics, and their location in the space and to move or track the object. In the area of image processing, numerous research projects have been undertaken and successfully carried out. Automatic surveillance systems based on real-time videos of public spaces are becoming necessary due to growing concerns about public safety and security. These surveillance systems must be installed in busy and critical locations such as markets, malls, renowned eateries, train stations, etc [2-4]. They are also most in demand for traffic control and examination, activity recognition and tracking, fault detection in industrial applications, and semantic video indexing, without restricting their application to security beneath public places. The approach utilised is to initially detect the target of interest in individual frames in order to do the high level objectives of categorization or tracking a target from video stream [5-7].

Background subtraction is a surveillance technique that is employed in various works. By separating the background and foreground pixels in the frame being processed, this approach extracts the foreground object, or the target that is moving. Many researchers have taken full use of this method's advantages, particularly its performance when a stationary video camera is present and its lighting invariance. A crucial factor to take into account is creating a background model of the video frame that was collected. For feature extraction for detection of objects, textures such as Local Pattern Binary (LBP) [8-10] of the image have been considered. Pixel neighbourhood operations are used to compute LBP characteristics. Histogram of oriented gradients is a common feature descriptor for object detection. HOG features are shape descriptors that represent an object in specific directions in terms of intensity gradients. In [11], the researchers took advantage of HOG [12] functions, citing their invariance properties in regard to transformations such as rotation, deformities and conditions of illumination.

System Design for the proposed method

The system design phase gives the proposed research project an abstract representation that outlines the entire workflow of the study and how each module must be executed and integrated by the effective application development.

Design diagram- High level

This design diagram describes the representation of all the modules in the proposed technique and provides solution for the services offered by the system to produce the high quality design for the research work. In a multi-project model, such an outline is important to ensure that each supporting element design is consistent with the neighboring designs and the large picture. All the services of the proposed system, the platform used and the process of implementation should be described in the brief manner and any important change needed to be done and integrated to be specified in this stage. Furthermore, all major commercial, legal, environmental, safety and safety matters should be considered briefly [13-15].

Design Diagram-High level

The bubble graphs used in this diagram can be classified as dfds. One of the simplest graphical representations, as seen in figures 1 and 2, is the data flow diagram [16].

Case Diagram

Case diagram is shown in the figure below. The functionality for the communication will be documented and carried out in accordance with how 3 explains the interaction between the application framework and the end user.

Characters who participate in the process are referred to as on-screen characters, while those who perform outside the parameters are referred to as performing artists. The primary goal of this design diagram is to explain how each module communicates with the others in a way that aids in the execution of the work [17-19].

Design Diagram-Activity

The activity diagram describes in the figure .4, presents the important activities carried out in the research work. In this diagram the circles represents the start of the activity and the end of an activity and the rectangle boxes defines the modules of each proposed research work [20-21].The purpose of activity diagram in the proposed framework is to make sure the workflow of the application development is implemented according to the desired requirements provided by the end users. The developer will refer these design diagrams for the implementation of the work.

Process of identification of objects and the classification technique

Step 1: Input the video that contains both human and non-living moving objects.

Step 2: Pre-processing the input video: The input video to be pre-processed is two different steps:

- Divide the input video into frames and store individually.
- Once frames are generated apply the morphological operations over the input video.

Image Pre-Processing and Annotation

Image Pre-Processing involves processing or cleaning of images. This step focuses on removal of noise and distortion, sharpening, intensity normalization, etc. The VOC dataset is refined with only person images and annotated according to the format of YOLO model. A text file is created for each image in the same directory with the same name that contains object number and object coordinates on this image, for each object in new line. The object numbers an integer number of object from zero to total number of classes – 1, and object coordinates are float values relative to width and height of image, it can be equal from (0.0 to 1.0]. The ID card images are only pre-processed [22-25].

Training the YOLO model and testing

After pre-processing and annotation, the person dataset is divided into training and testing datasets. We train the YOLO model using training dataset until we get a better mean Average Precision (map). After the training, it is tested with testing dataset. YOLO is a full convolution network consisting built using darknet-53. It detects objects at three different strides (8, 16 and 32) which help to detect smaller objects. The provided input image will be divided into S*S grid and each of the cell will be made of the coordinated of (x, y, w, h) and the confidence of the object. The representation of the coordinates x, y defines the position of the boundary box which is relatively grid in nature. The coordinates with w, h is represented as the width and height of the detected boundary box. The probability of each grid is predicted for C categories. The confidence is determined by the probability model that includes the target and the prediction detection box [26]. The object is defined as Pr which stands for the target object that falling under the cell. If the confidence is presents then it is defined using:

$$C\ (Object) = Pr\ (Object) * IOU\ (Prod, Truth) \quad (1)$$

The cell doesn't contain any kind of the object and the confidence values is defined as C (object)] =0. The IOU is defined as the overlapping rate for which the bound of candidate and the truth value of the ground can be defined as the ration of union and the intersection of the grounds [11-13].Then the classification of the objects is achieved by classifying them into their respective categories. Here, multi class SVM classification is used by supervised learning for the output of the object class. The quality of CNN classification is determined over video data set checking: The category output is the class to which an object is identified. In a single frame, a vector containing all classes detected is generated as an output for multiple objects of different lasses.

Extraction of features for the identification of Non-Living objects

This section presents the detailed study of how the extraction of features has been implemented and the mathematical model related to the techniques used for feature extraction technique used will be described [27].

Corner Detector using Shi-Tomasi Technique

The Harris Corner Detector is an experimental model that detects the corner features from the video frames that provides higher texture features with minimal propositional changes. The harries technique is mainly used to detect the corners of the frames and it is carried out as follows:

The intensity of the pixel proposition is defined in I(x, y) for the position (x, y) of each window frame of the input video, and if the window moves by small margin the shift (e, v), will be marked with I(x+u, y+v). Since the main objective is to locate regions or windows with small displacements in the image, the intensity is expressed mathematically.

$$E(u,v) = \sum_{x,y} w(x,y)[I(x+u, y+v) - I(x,y)]^2 \quad \text{..... (1)}$$

The weight function is defined with 'W' and the high intensity variation in the window frames may lead to the result of E (u, v).

The Taylor's series and simplification technique of Equation 1, provides the results in,

$$M = \sum w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \quad \text{.... (2)}$$

The Eigen values of matrix is defined as M is used to find out suitable corners sing the score value,

$$S = |M| - k(trace(M))^2 \quad \text{..... (3)}$$

If λ_1, λ_2Are Eigen values of M, then |M| = $\lambda_1\lambda_2$andtrace (M) = $\lambda_1 + \lambda_2$ Each corners of the frame is represented using S, to predict the high score value. The smaller change in the calculation is calculated using the Shi-Tomasi technique, that leads to the determining the most suitable corner defined in 'n', rather than identifying the each and every corner of the feature [28].

Shi-Tomasi score for corner detection follows-

$$S = min(\lambda_1, \lambda_2). \quad \text{..... (4)}$$

If the score, S exceeds a threshold it is considered as a corner.

Determining the optimal flow using Lucas-Kanade technique

The optical flow of the detected fame has to be determined, that track the movement of the objects in frame when the object is moving or during the rotation of the camera. The optical flow of any kind of entity from one frame to another frame in a video is to be determined. The proposed algorithm Lucas-Kanade [29-31] is mainly based on the type of optical flow theory, where all the video frames remain in the similar kind of intensity and the pixels in the frames have similar kind of movements and the pixels rate between successive each frames smaller in nature. A pixel rate [32] with the intensity rate defined as I(x,y,t) in a frame at time interval defined as t after the movement with a small displacement defined as (d_x, d_y) in the consecutive frames with a difference of time d_tIs expresses as:

$$I(x,y,t) = I(x + d_x, y + d_y, t + d_t) \quad \text{..(5)}$$

The flow of an optimal equation for the movement of objects in an image is given as-

$$\frac{\partial_f}{\partial_x}\frac{d_x}{d_t} + \frac{\partial_f}{\partial_y}\frac{d_y}{d_t} + f_t = 0 \quad \text{..... (6)}$$

The optical flow motion of the vectors using Lucas-Kanade method is defined by solving Equation (6):

$$\begin{bmatrix} u \\ v \end{bmatrix} = \Sigma_i \begin{bmatrix} f_{x_i}^2 & f_{x_i} f_{y_i} \\ f_{x_i} & f_{y_i}^2 \end{bmatrix}^{-1} \Sigma_i \begin{bmatrix} -f_{x_i} f_{t_i} \\ -f_{y_i} f_{t_i} \end{bmatrix} \quad \text{..... (7)}$$

The representation of the displacement defined I (u, v) represents the of the object between consecutive frames.

RESULTS AND DISCUSSIONS

The framework can be integrated within the Mat lab tool kit that makes it possible to use its toolboxes for the computer vision and machine learning to easily integrate mathematical computations for the identification and classification of artifacts. After extraction, the video processing is carried out smoothly on each frame. The toolbox for image processing includes several filtering and refining functions required to process an identified image before processing. A maximum of 15 videos containing standard products of various classes are collected. The dataset has been divided into three groups–5 video training datasets and 8 video test data sets required for vector support machine technique. Figure 6 below shows the histogram of frames extracted from video 1 in which different objects are recognized in the same way after extracting characteristics, as in Figure 7 and Figure 8 various objects detected from video 6 and video 7 respectively. The final results are drawn using the proposed CNN framework where the input is given to the framework that contains both human and non-living objects in the video dataset. The input video will be processed and detects the background and fore-ground of the objects. Initially pre-processing is applied to remove noise in the video and then morphological operations are applied to analyze the color pixel of the detected frames.

CONCLUSION

This work presented a new approach for object recognition using Vector Machine based classification in Video Surveillance Systems and Lucas-Kanade technique. In this article, the artifacts are correctly identified and their position from an unknown location is calculated. First, object recognition using Shi-Tomasi and Lucas-Kanade techniques will be stored, and the context subtraction will be applied when an object from extracted frames of the input video is recognized. Then the classification of the objects in their individual categories is accomplished by supervised learning with the help vector machine classification. The precision of the technique being proposed is analyzed by the total number of frames detected by object compared to the total number of frames. In this chart four input videos from different sources of various sizes and backgrounds were taken, and for each video we should achieve 92 percent accuracy. Where frames vary from 500 to 1500 for each video, the exactness of the identification of the objects is 80 to 95% for each video.

REFERENCES

1. C S Pillai, AnandaBabu J, “Object Recognition using Lucas-Kanade Technique and Support Vector Machine Based Classification in Video Surveillance Systems”, International Journal of Engineering and Advanced Technology(IJEAT), Vol. 9 Issue -1,October-2019, ISSN:2249-8958.
2. Elliott, D.: Intelligent video solution: A definition, Security, pp. 46–48, 2010.
3. Avidan, S.: Ensemble tracking, IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 2, pp. 261– 271, Feb. 2007.
4. Kushwaha, A., Sharma, C., Khare, M., Srivastava, R., Khare, A.: Automatic multiple human detection and tracking for visual surveillance system. 2012 International Conference on Informatics, Electronics Vision (ICIEV), pp. 326-331.
5. B. D. Lucas, T. Kanade, "An iterative image registration technique with an application to stereo vision," in IJCAI, 1981.

6. M. Everingham, L. Van Gool, C. K. Williams, J. Winn, A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge", IJCV, 2010, pp. 30.
7. Avidan, S.: Ensemble tracking, IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 2, pp. 261– 271, Feb. 2007.
8. Khan, Z., Gu, I.: Joint feature correspondences and appearance similarity for robust visual object tracking, IEEE Trans. Inf. Forensics Security, vol. 5, no. 3, pp. 591–606, Sep. 2010.
9. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection, in Proc. CVPR, 2005, pp. 886–893.
10. Wang, L.: Abnormal walking gait analysis using silhouette-masked flow histograms, in Proc. ICPR, 2006, vol. 3, pp. 473–476.
11. Wang, S., Lee, H.: A cascade framework for a real-time statistical plate recognition system, IEEE Trans. Inf. Forensics Security, vol. 2, no. 2, pp. 267–282, Jun. 2007.
12. Yu, X., Chinomi, K., Koshimizu, T., Nitta, N., Ito, Y., Babaguchi, N.: Privacy protecting visual processing for secure video surveillance, in Proc. ICIP, 2008, pp. 1672–1675.
13. Park, U., Jain, A.: Face matching and retrieval using soft biometrics, IEEE Trans. Inf. Forensics Security, vol. 5, no. 3, pp. 406–415, Sep. 2010.
14. Cong, Y., Yuan, J., Luo, J.: Towards scalable summarization of consumer videos via sparse dictionary selection, IEEE Trans.Multimedia, vol. 14, no. 1, pp. 66–75, Feb. 2012.
15. Cong, Y., Gong, H., Zhu, S., Tang, Y.: Flow mosaicking: Real-time pedestrian counting without scene-specific learning, in Proc. CVPR, 2009, pp. 1093–1100.
16. k. Cheng, Q. Liu, R. Tahir, L. K. Eric and L. He, "Multi-Camera Logical Topology Inference via Conditional Probability Graph Convolution Network," *2021 IEEE International Conference on Multimedia and Expo (ICME)*, 2021, pp. 1-6, doi: 10.1109/ICME51207.2021.9428335.
17. L. H. Jadhav and B. F. Momin, "Detection and identification of unattended/removed objects in video surveillance," *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, 2016, pp. 1770-1773, doi: 10.1109/RTEICT.2016.7808138.
18. R. Alimuin, A. Guiron and E. Dadios, "Surveillance systems integration for real time object identification using weighted bounding single neural network," *2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, 2017, pp. 1-6, doi: 10.1109/HNICEM.2017.8269461.
19. G. Pavithra, J. J. Jose and T. A. Chandrappa, "Real-time color classification of objects from video streams," *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, 2017, pp. 1683-1686, doi: 10.1109/RTEICT.2017.8256886.
20. K. S. Kumar, S. Prasad, P. K. Saroj and R. C. Tripathi, "Multiple Cameras Using Real Time Object Tracking for Surveillance and Security System," *2010 3rd International Conference on Emerging Trends in Engineering and Technology*, 2010, pp. 213-218, doi: 10.1109/ICETET.2010.30.
21. B. Hdioud, A. Ezzahout, Y. Hadi and R. O. Haj Thami, "A real-time people tracking system based on trajectory estimation using single field of camera view," *2013 International Conference on Computer Applications Technology (ICCAT)*, 2013, pp. 1-4, doi: 10.1109/ICCAT.2013.6522038.
22. S. Sen, A. K. Das and S. P. Chowdhury, "Saving Electrical Power in a Surveillance Environment," *2009 Seventh International Conference on Advances in Pattern Recognition*, 2009, pp. 274-277, doi: 10.1109/ICAPR.2009.38.
23. V. Mathew, T. Toby, A. Chacko and A. Udhayakumar, "Person re-identification through face detection from videos using Deep Learning," *2019 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, 2019, pp. 1-5, doi: 10.1109/ANTS47819.2019.9117938.
24. J. Tang, R. Ding and X. Tian, "Real-Time Target Detection Algorithm Based on Background Modeling," *2013 Fourth International Conference on Digital Manufacturing & Automation*, 2013, pp. 970-974, doi: 10.1109/ICDMA.2013.227.
25. S. Alfasly, B. Liu, Y. Hu, Y. Wang and C. -T. Li, "Auto-Zooming CNN-Based Framework for Real-Time Pedestrian Detection in Outdoor Surveillance Videos," in *IEEE Access*, vol. 7, pp. 105816-105826, 2019, doi: 10.1109/ACCESS.2019.2931915.

26. L. Qu, J. Wang, S. Xin, M. Qin and J. Dong, "A System for Detecting Sea Oil Leak Based on Video Surveillance," *2011 Third Pacific-Asia Conference on Circuits, Communications and System (PACCS)*, 2011, pp. 1-3, doi: 10.1109/PACCS.2011.5990183.
27. A. Singh, T. Anand, S. Sharma and P. Singh, "IoT Based Weapons Detection System for Surveillance and Security Using YOLOV4," *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, 2021, pp. 488-493, doi: 10.1109/ICCES51350.2021.9489224.
28. X. Song, L. Wang, H. Wang and Y. Zhang, "Detection and identification in the intelligent traffic video monitoring system for pedestrians and vehicles," *The 7th International Conference on Networked Computing and Advanced Information Management*, 2011, pp. 181-185.
29. B. Ashwini, B. Deepashree, B. N. Yuvaraju and P. S. Venugopala, "Identification of vehicles in traffic video," *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, 2016, pp. 588-593, doi: 10.1109/SCOPES.2016.7955507.
30. C. -Y. Wang, P. -Y. Chen, M. -C. Chen, J. -W. Hsieh and H. -Y. M. Liao, "Real-Time Video-Based Person Re-Identification Surveillance with Light-Weight Deep Convolutional Networks," *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2019, pp. 1-8, doi: 10.1109/AVSS.2019.8909855.
31. Z. Zhen-Jie and W. Qi, "Research on Detection and Tracking of Moving Vehicles in Complex Environment Based on Real-Time Surveillance Video," *2020 3rd International Conference on Intelligent Robotic and Control Engineering (IRCE)*, 2020, pp. 42-46, doi: 10.1109/IRCE50905.2020.9199246.
32. Y. Chen, B. Wu, H. Huang and C. Fan, "A Real-Time Vision System for Nighttime Vehicle Detection and Traffic Surveillance," in *IEEE Transactions on Industrial Electronics*, vol. 58, no. 5, pp. 2030-2044, May 2011, doi: 10.1109/TIE.2010.2055771.

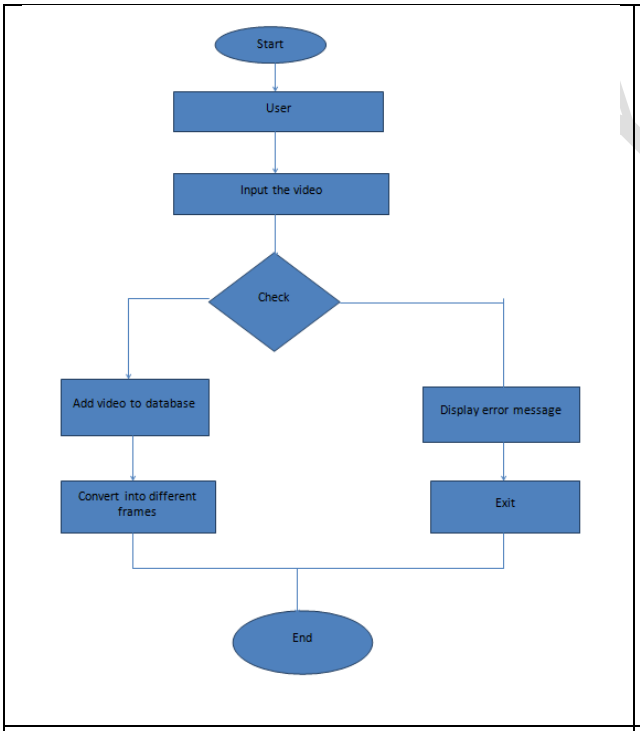

Fig. 1: Data Flow Diagram for Add Database

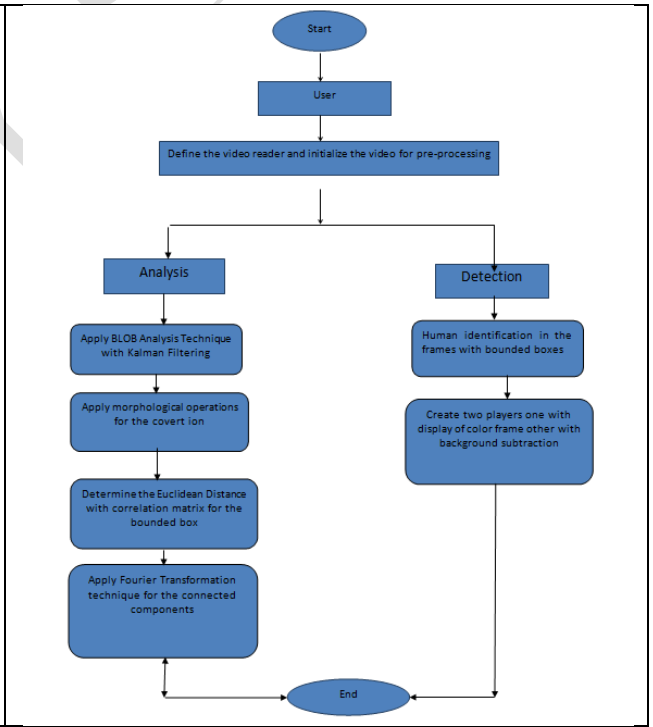

Fig. 2: Data Flow Diagram for Proposed Technique

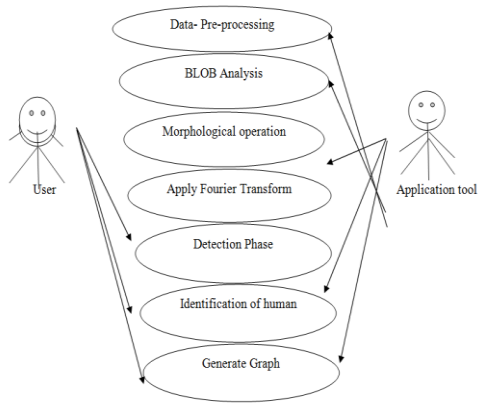

Figure 3: The Use case diagram for Proposed Technique

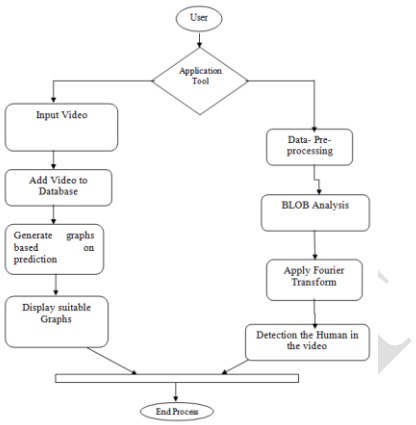

Figure 4: The activity diagram for Proposed Technique

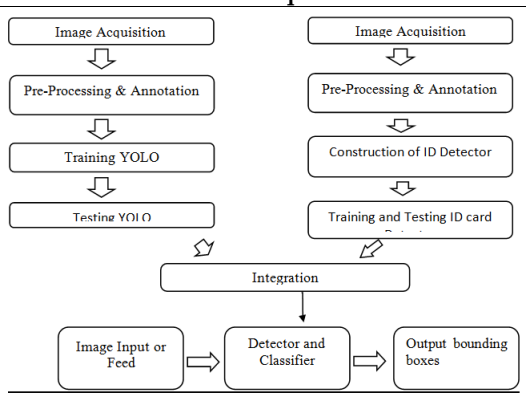

Figure 5: System Architecture of proposed Framework

Figure 6: Frame extraction for the detected objects in the input video 1

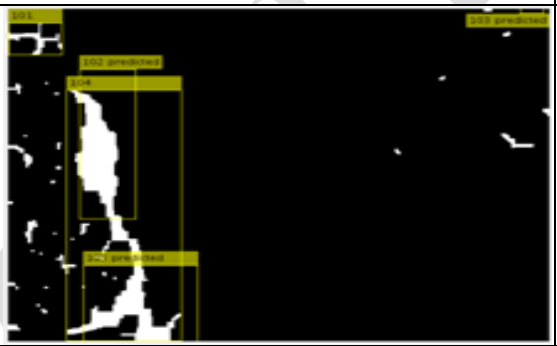

Figure 7: Frame extraction for the detected objects in the input video2

Figure 8: Frame extraction for the detected objects in the input video3