ALL VISION: A Novel Security Architecture and Alarm system Using Object Detection And Computer Vision Technology For Large Campuses

Rishabh Panchal, Devanshi Rajput, Shreya Patel, Shival Patel

B. Tech student, B. Tech Student, B. Tech Student, B. Tech Student
Computer Science & Engineering,
Parul University, PIET, Vadodara, India

Abstract : The primary objective of the ALL VISION is to reduce the cost and human effort required for the task of constantly monitoring the CCTV feed in any institutes or organization that has large campus. ALL VISION uses modern computer science innovations such as Object-Detection, Image Processing and Data mining to alert the nearby guards with the help of the alarm about present danger/threat or illegal activities prohibited by administration. Data storages are also maintained to keep logs of suspicious activities that might be needed to reference some day. It will serve all the organization with fully automated security environment. Situational detection works in regards to humans and their activities to provide notification and provide red screen in case of violence is detected.

Index Terms **– Computer Vision, Object-Detection, Violence Detection, Image Processing.**

I. INTRODUCTION

Object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos.

In general, we come across thousands of items in our daily lives and our mind processes the basic knowledge and meaning of the items. The human brain understands this by prior knowledge of the items or items related to the one just seen. Our brain, characterizes all of this with the help of its dimensions, colours, weight, flexibility and many other properties related to such objects. Computer Vision uses same concept and rely on the datasets which is machine equivalent of brain memory. The prior knowledge of such characteristics are known as training and thus CPUs and computers become able to identify objects as humans do.

The object detection is combination of both Computer Vision and logical algorithms that deciphers the image processed via camera and then applies prior knowledge of what the thing presented against it should be. Thus, giving efficient results of how the particular image results in the outcome of the possible object. As established earlier, the object detection can be used for identifying many of the objects in the surrounding of particular camera.

However, there are many objects that could be illegal in the campus/grounds of specific organisation and might even posses some dangers. In current days, CCTV cameras are attached everywhere and can be monitored for such activities. However, that task requires maximum effort from at least 3-4 security personnel and even then there is always possibility of something going unobserved.

Thus, by using the object detection with logical algorithms that can raise alerts when illegal objects or threatening condition arise. There shouldn't be need of constant monitoring the feeds of camera when the task can be achieved automatically with computers. Even activities that are less dangerous but more suspicious can also be observed and referenced for future.

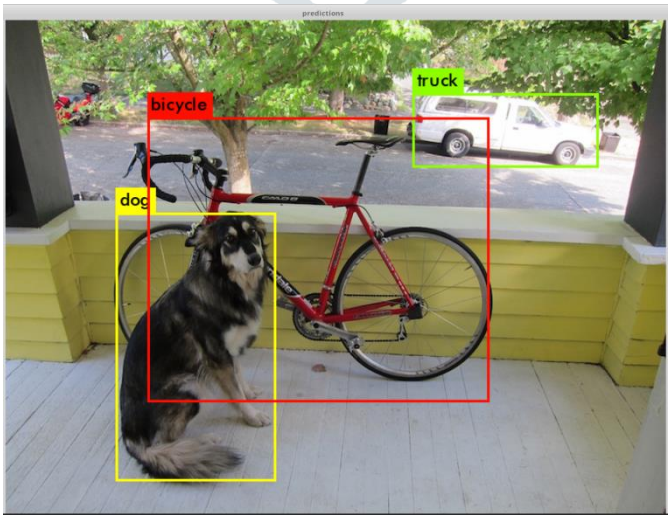

Fig 1: YOLO Prediction example with objects detected from image

II. RESEARCH METHODOLOGY

Object detection method combining top-down recognition with bottom-up image segmentation. There are two main steps in this method: a hypothesis generation step and a verification step.

In the top-down hypothesis generation step, we design an improved Shape Context feature, which is more robust to object deformation and background clutter. The improved Shape Context is used to generate a set of hypotheses of object locations and figure ground masks, which have high recall and low precision rate. In the verification step, we first compute a set of feasible segmentations that are consistent with top-down object hypotheses, then we propose a False Positive Pruning(FPP) procedure to prune out false positives.

2.1 Object Detection Procedure

Object Recognition can be explained as a procedure where identification of a specific object in an image or video sequence is performed. They rely on various techniques such as matching, learning, or pattern recognition algorithm using appearance-based or feature-based techniques. In order to do the analysis on object recognition and segmentation .They are classified on the basis of static objects in images and moving objects in images. The techniques based on static objects are related to the mathematical model and graphical models which includes the techniques such as; Spectral Graph Partitioning, Spin Image Recognition algorithm, Sift and Graph cuts. The techniques based on moving objects are basically Soft computing and algorithmic based approaches such as; Gray Scale Video Imagery, Block Partitioning Algorithm, MLS point clouds etc.

However, more sophisticated techniques can be used for object recognition and segmentation such that they deal well with occlusions and luminous effects and give better results. Various hardware based techniques such as; POSS-v, MLS point cloud, video imagery can use better processors for more precise and accurate results. Algorithmic techniques can be improved in future if we work upon reducing their complexities.

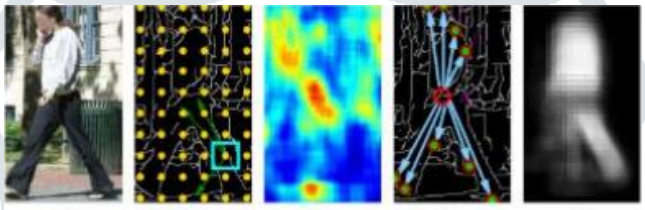

Fig 2.1: Top-down recognition. (a) An input image;(b) A matched point feature votes for 3possible positions; (c) The vote map V. (d) The hypothesis traces back find its voters. (d)Each points predicts the figure-ground configuration. (Google Images: Tracking human movements)

2.1.1 Used Technologies

There are three technologies which are used: YOLO, Tensor Flow and OpenCV camera. YOLO, a new approach to object detection. The images that fall into same class group are identified. Frame object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. It looks at whole image at taste time so its prediction are informed by global context in the image.

Our model is simple to construct and can be trained directly on full images. Tensor Flow is a computer vision library developed by Google for task of repairing the GPUs and CPUs for training the datasets and running the application based on videos at runtime. It is an open source software library for high performance numerical computation. It provides platforms for deployment of computation and from desktops to clusters of servers to mobile and edge devices. Brain team within Google's AL organization, it comes with strong support for machine learning and deep learning and the flexible numerical computation core is used across many other scientific domains.

Open CV camera is a library of programming functions mainly aimed at real-time computer vision. Original developed by intel. It includes area likes gesture recognition, facial recognition system, human computer interaction , motion understanding, object identification , structure for motion and motion tracking . It is written in c++, all of the new development and algorithms in open CV are developed in c++ interface.

2.1.2 Framework Detection

ALL VISION is a simple and manageable desktop application which is divided into two modules. Firstly, it detects objects around the surrounding such as, any suspicious objector device, it is for hazardous situation and another module is of violence detection which separates by fight and non-fight datasets.

2.1.3 Flowchart and Use Case Diagrams

2.1.3.1 Flowchart

As the diagram represent, ALL VISION GUI provides access to the two of the most important modules of the framework; Object scenario such as this, footage in actual scenario), therefore the next stage is of dividing each video in to individual frames.

Scenario 1: The object detection module uses the YOLO API specially designed to make the boxes for the objects available to user on the screen. YOLO is framework designed with C++ and python to make the object detection directly with access to modules and models pre-trained with machine learning

Scenario 2: The violence detection modules uses Tensor Flow API developed by Google in python to make the more and more easier. Provided by the datasets of movie fights and hockey fights, machine is trained to find the situation similar to fighting in the environment of video or camera.

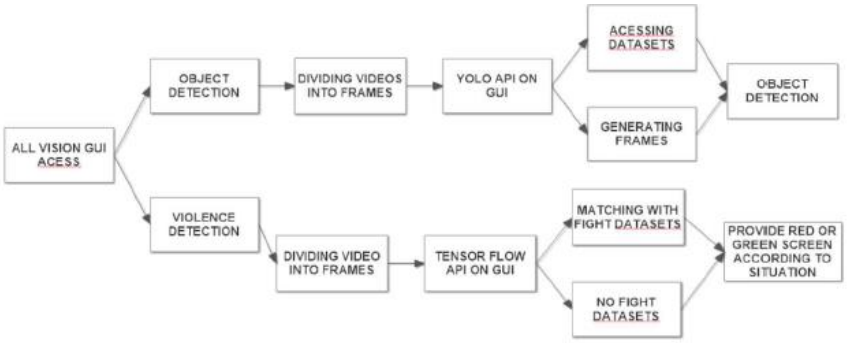

Fig 2.2: Flowchart of ALL VISION access mode

2.1.3.2 Use Case Diagram

While the Use Case seems quite a simple and small, it defines the user definition for the use of Frame work quite easily. Security Admin or the person in charge of the security systems provided in campus generally will access the modules of ALL VISION. Both the modules of Object Detection and Situational Detection is available for the security in-charge to run at same time. This is achieved with the functionality of enabling multiple instances of GUI running at same time. While the system defines automatically according to the campus's need the tasks that will needed to be achieved, we have provided administrator user in case system requires any management and maintenance services. The framework while automatically is trained with suited model, we have also provided script that can enable the running of training process once again to re-train or update the dataset.

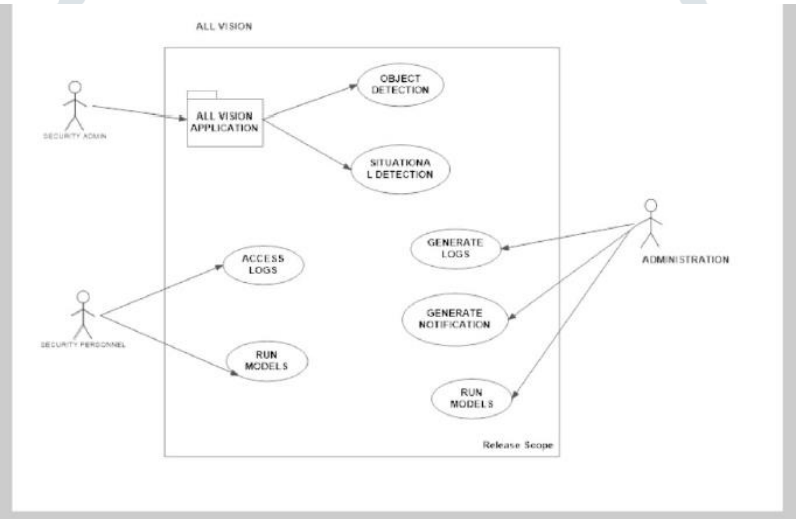

Fig 2.3: Use Case: Object vision user access

2.1.4 Control Flow of TensorFlow algorithm

CNN is Convolutional Neural Network used for decision making. YOLO also uses the updated version of neural network knows as R detection decision making. Following describes the steps of how the training is achieved.
Step 1: Providing two set of video input (converting C++ to the pb and python).

Step 2: Darknet also uses the darkflow to generate weights/data files. (Final model of the decision making). The result of this process generates neural network

Step 3: Generated CNN which is enough to describe the tree and network by which the final decision is achieved, the computer can only understand the final process in the form of binary files. This binary file is known as Final Checkpoint (FC).

Step 4: The resultant checkpoint is provided used in deep learning to provide the storage for comparison between various results occurred during the stage of training. This networks provide basic functionality of data mining such as classifying, processing, and making predictions. It's main work isto refine the generated Final Checkpoint to better suit the needs of framework. It also control how the training procedure is achieved with the help of R-CNN (Region – CNN) specifically aimed for the object sets of video input (fight and non-fight) to the Darknet to Long Short-Term Memory (LSTM) networks.

III. GUI DEVELOPMENT WITH C#

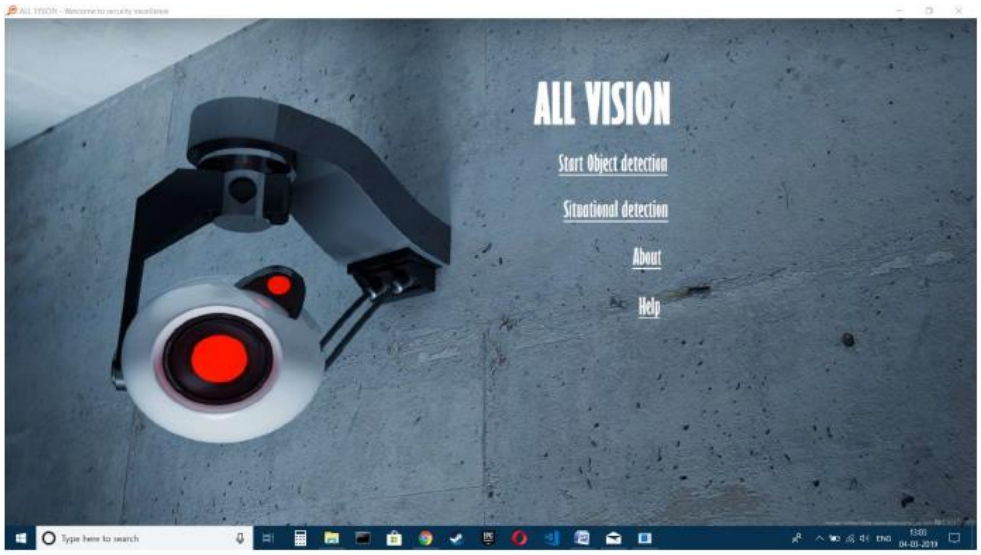

Fig 3: ALL VISION homepage

Menu 1 (Start Object detection): This starts the YOLO API and enables the python script that will detect objects from the video. When found images/video which contain the items such as knife, guns etc which can be very harmful in the scenario; will lead to immediate notification.

Menu 2 (Situational detection): As the name suggest, it is the simple UI that shows the current video going on in front of the eye of camera. The original frame shows the green light on the system which fighting or harmful detection is detected, it will turn the lights into red thus attracting attention of the security personnel.

Menu 3 (About): Describes what is ALL VISION and how to run the modules and help functions.

Menu 4 (Help): Gives the contact information and provides various help pointer to describe and detail in what scenario ALL VISION works best.

IV. RESULTS, ADVANTAGES & LIMITATIONS

The results of the structure is mentioned in each of the chapter generated. However, following are the main analysis we discovered.

4.1 Results

- The object detection module developed in ALL VISION is used for detecting harmful objects and will lead to the notification if the harmful object is detected.
- The violence detection module developed in ALL VISION is used for detecting scenarios which can lead to fights and will generate red light on screen when violence is found.
- The performance and real time analysis is possible with the help of CUDA and NVIDIA GPU.
- An overall secure environment is generated when both modules are run together.

Fig 4: As violence is detected in video stream, frame turns red

4.2 Advantages and Uses

- Primary uses inside the large campuses containing CCTV cameras.
- Any public camera stated by the government and constantly monitored.
- Provides non-volatile cache support for logs and videos can be stored as output when situation is located.
- Faster communication support with YOLO & TensorFlow
- Real-Time detection with GPU Trained data set with algorithm available for training repeated data and new data.

4.3 Limitations

- If GPU is not available, real time detection will not work on live videos, rather it will work on the video provided as input.
- Current GPU is only supported with NVIDIA.
- GUI supported with Windows only (although code can run on linux or any other platform).
- The overall efficiency is dependent on the datasets and training, video faster than 45 FPS may lead to omitting of some objects that was on the screen.
- If the human is not properly detected, fights might not be detected.

V. CONCLUSION

With the implementation of this framework, we have successfully analyzed the requirements, both technical and financial for the implementation of our software/desktop application knows as ALL VISION. The project implements our selected method from various technologies that we have used to implement inclusive of YOLO framework as well as Situational Object detection. The report aims towards giving users of this application a technical view point regarding methodologies and fundamentals used for ALL VISION.

VI. REFERENCE

[1] A survey on Object Recognition Methods, Sukanya CM, RoopaGokul, Vince Paul, IJCSET January 2016
[2] An active Search Strategy for Efficient Object Class detection, Abel Gonzalez Garcia, Vittorio Ferrari, Alexander Vezhneverts, CVPR 2016 Paper
[3] Zero Shot Object Detection, AnkanBansal, Karan Sikka, Gaurav Sharma, Ajay Divakarna, ResearchGateNet Publication
[4] A review Paper on Object Detection for Classification Accuracy and Robustness, Divya Patel, International Journal of Computer Applications
[5] Cluster Analysis of Cyber Crime data using R, RadhaMothukuri, International Journal of Computer Science and Mobile applicationsR.
[6] A Novel Sorting algorithm for data analysis, T. Kesava, G. Mohan Ram, International Journal of Scientific Research in Computer Science Volume3, Issue 1
[7] Evolving Database for New-Gen Big Data application, Ronald Barber, Rene Mueller, IBM Research – Almaden
[8] IMAGING: In-Memory AlGorithm for Image processiNG, Ameer haj Ali, Ronny Ronen, Nimrod Wald, Article in Circuits and systems I:Regular Papers, IEEE Transactions on – June 2018
[9] A Survey on Object recognition and segmentation techniques, Anshika Sharma, 3rd 2016 International Conference for Sustainable Global Development, March 2016
[10] Object Detection combining and segmentation, Liming Wang, Gang Song, ResearchGate net publication – 2017
[11] Use of image processing through emotion detection, Chandan Koner, Manas Karmarkar, International Journal on Applications of Information and Communication Engineering February 2015
[12] Person Re-Identification by Deep Learning Multi-scale representation, Xiatian Zhu, ICCV workshop paper, 2018
[13] Vehicle Detection from satellite Images in digital Image Processing, Anand Sharma, International Journal of Computational Research, November 2017
[14] Synoptic Study of Abandoned Object Detection, RutujaPatil, Prof. KishorPandyaji, IJESC research paper volume 7
[15] YouOnlyLookOnce: Unified, Real-Time Object Detection, Joseph Redmon, Ali Farhadi, ResearchGate net publication, 2016