

3D Object Detection from Multi-View Images by Using OpenCV in Virtual Reality

Raghavendra Reddy¹

Samuel Nitesh John²

Shashank MS³

Pradeep D⁴

VireshSharnappa Karadi ⁵

[¹raghavendra.reddy@reva.edu.in](mailto:raghavendra.reddy@reva.edu.in)

[^{2,3,4,5}Niteshjsamuel@gmail.com](mailto:Niteshjsamuel@gmail.com)

Shashumadikar1997@gmail.com

Pradeepkumar9051692@gmail.com

viresh.karadi.97@gmail.com

ABSTRACT

The current Scenario of the day-to-day development of the computer technology has led to the Rapid increase in the use of 3D Modelling based Computer Graphics or we can say virtual reality / Augmented reality, Computer Based Video Games and 3D Movies and many More Application. This concept creates the new world to human kind where everything can be built, make and used. There are several technologies available in the market which help to create the 3D object in the virtual world. Our concept is to detect the object in the virtual world and tell what that object refers too. We use the OpenCV for detecting the object and find it's coordinated respectively. The 3D data to given through the computer window which is unity simulator. The accuracy of proposed system depends on frame (i.e., 10- 13, 17-20, 27- 30, 45- 70, in 0~0.3s, 0.1~0.4s, 0.8~1s, >1s time windows respectively). The accuracy even depends upon the distance away from the object (i.e., 60.4%, 73.9%, 89.3%, 95.2%, and 62.2% in >15, 15~10, 10~6, 6~0.5, <0.3 meters respectively). The result of the data been train are varieties that way the model will understand object in 3D world. The accuracy will be very bases on to the distance away from the object since it's 3D world the vast world will has to be consider. The future enhancement is vast by using this module example like identifying the 3D module of the human and find the its joints, arms, legs etc. to make the 3D module move on its own.

KEYWORDS: *object detection, 3D module, OpenCV, Unity 3D world, 3D object detection, virtual Reality, Augmented Reality*

1. INTRODUCTION

The innovation of neural network has inspired and made a way to solve many real time problems occurring naturally. every neutral network will self-learn and that helps to repair the picture. Neutral network will self-train and self-learn and run the image appearing, which helps extraction of features to the competition, So the effect and the effectiveness of their approach Based on the discussion above, this paper proposes a new template editing method that uses 2D images or is not as good as our approach. geometric shapes that incorporate the users' visualized thoughts in hand-drawn sketches or the visual feature of the 2D images selected by the users.[1] This is because the network performs multiple operations of max-pooling and down-sampling, which is originally constructed for abstraction task. The feature dimension is largely reduced. This results in the weak spatial description with drastically reduced spatial resolution. Convolutional Neural

Network cannot learn the description of the structural relationship effectively. Moreover, due to the high and variable flight altitude and multi-angle rotate shooting of Unmanned Aerial Vehicle views, the ground objects with the same semantics are usually with small size, multiple scales, and high density spatially close to others, a cross-matching module is employed to avoid identity switch based on the proposed appearance model[12].

Now days the virtual Reality and Augmented Reality are major technology that are becoming the very popular to the user. The interface which it provides is fascinating and attractive to the users which will attract more user to use it and even in the game industry also those technology emerging. If the machine Learning to be combine this will create a vast door to new ideas for developer as well as gamers. The OpenCV is the image processing tool which will help us to achieve this goal frankly the data which are generated for the computer vision can be converted into the frame that way it will ease to identified the object.

The main contributions present in this paper are as follows:

- We have proposed a method where the 3D Object can be detected in the 3D or virtual world.
- We have designed a system for detecting objects which is represent in the 3D world to detect what the object represents and the give ability to move on its own.

To provide a system to communicate Unity window screen to OpenCV module to detect the object.

In the proposed system we will detect the 3D object in the 3D Environments using the OpenCV and CNN algorithm to detect the object in the 3D world which will be the Unity 3D the platform for the 3D modeling and object render. The current technology that works on the real-world object detection our system will detect the object in the 3D world the since the data are all related to real world, we can make use of the data to tell object in the 3D world.

This paper is structured in the following way: Section II outlines the associated studies in the field of activity detection research. Section III explains the new method's ultimate perspective. Section IV describes the implementation specifics, followed by Section V conclusion and future enhancement work.

2. LITERATURE SURVEY

Because of deep learning and Neural Network, a lot of progress has been made in machine vision. As a result, stronger and more reliable algorithms have emerged. In the field, certain algorithms were used to solve a wide variety of complicated computer vision problems. Some of the papers discussed here will help us to identify the research gaps.

Weizh NIE et. al.[10]proposed /suggested3d modelling in such a way by using Computer graphics and there by using Computer graphics he produced a 3d object representation in a surface the author make used of an special software to identify the open points in an 3d virtual space to form a mesh like structure. This collection of vertices is later grouped to form an object. The important characteristics of this 3d objects are it can be either created automatically or manually by deconstruction of the mesh like structure. This are used in various fields like movies

Xinying wang et. al.[11]proposed /suggestedthe innovation of neural network has inspired and made a way to solve many real time problems occurring naturally. every neural network will self-training and self-learning and that helps to repair the picture. Neural network will self-train and self-learn and run the image appearing, which helps extraction of features to the competition Nevertheless.

Chengji Liu et. al.[7]proposed /suggestedYOLO algorithm help to identify the object in the image first it will take photos of height and width after it will try to identify the object in the image the identified image will be marked out with the green box benchmark marked box take as an output layer help to identify the box, the prepared output will be the input image for pipeline to process image objects classification and determination of object location inside the box. The limitation is that it takes lot of time to train the module to the custom data.

Cong Tang et. al. [1]proposed /suggested suggest ImageNet Large scale recognition challenge (ILSVRC), Deformable part Model (DPM), the image it will categorize with region selection, future extraction, classification adopted object detected by DNN. DNN had a subnetwork location and subnetwork recognition, it will compute feature in CNN. Limitation is that the latency in the data which accrued.

Bernardo Augusto Godinho de Oliveira et. al. [4] proposed/suggested The Neural networks are before due to more powerful GPUs, multicore CPUs and numerous algorithms have been introduced, reducing the complexity of time. ImageNet makes it possible to store pictures for machine learning. Because of this, it has become easy to form large and complex machine learning. The conventional neural networking algorithm assists in the identification of the object.

Table [1] shows the summary of various existing work approaches, advantages, limitations, and accuracy.

Few limitations are existing in the previous work. From the literature survey, few models even failed to perform predictions with good accuracies. Limitations towards the datasets even prevent getting good results quickly and precisely.

Table 1: Summary of the various papers studied under literature survey.

Ref No.	Approach	Advantages	Limitations	Accuracy
[10]	PointNet+ + pattern for the shape of a 3D object, K-means Approach for image processing.	It is able to detect missing functions.	harder to follow potentially useful impulses when your characters have a date with the next big event	83.2%
[6]	Hyperas, which chooses the best performing parameters	Due to 2D slice every corner of the 3-D image can be scanned	Fails while larger data.	79.2%
[11]	repairing 3D algorithm,	repaired region to maintain the detailed characteristics of its region through the local GAN and the global GAN	Not apply for the small holes but also large holes or even multiple holes.	88.5%
[3]	Deep Convolutional Neural Network technology for Object detection	Repair defective pixels in object detection.	It tracks down only with human beings.	94.2%
[5]	OpenCV: for Object detection and tracking	It will provide high-level precision with a less prominent image.	It is possible that the SSD will not generate high level accuracy with high level features.	85.3%

3. METHODOLOGY

The 3D object in the 3D world or the Virtual world are different compare to the real world. For system to understand the real-life object it uses the camera or pixel field to evaluate or process the data but it the 3D or virtual world is made of the

numbers and float values to make its environments look real, every position in 3D world are made of float numbers and pixel this way it become harder for system module to Understand. To this solution is to make the window vision to feed it to the OpenCV to Understand what that represent in the real word by object detection. The 3D scanned model delivered to OpenCV for image processing. OpenCV assists in detecting objects. OpenCV is a library of programming features aimed at real-time computation vision, OpenCV can be used for different programming languages to facilitate the development with python, Java, MATLAB, and openCV.js for the web platform. Captured 3D data first it will extraction of environment information feature it will comparison of observation The object will be observed if it does not exist it will take it as new observation, it will pass to change the route it will be updated in route.

Meanwhile the datasets to train the module can be the images of the real world pitchers the data can be normal process as the typical object detection work either it will be the YOLO modules or OpenCV to Train the disadvantage of the YOLO module which the training time will high which vary to 4 days to 15 days or moths according to the computation power there are the some light weight module which can be use or directly feeding the data to the R-NN or CNN which will reduce the computation power as well improve the performance[2]. The data will we feed will have the result on the terminal.

The data to which we can get from the Unity will every coordinate where the character move and interactive with the 3D world the only data we required to detect the object is the screen which the character can see in the 3D world. This way we can get the screen shot of the scene and give to module to tell the what is the character is seeing. The module will take the image and run the object detection module in it. and module return the value of array which contain the data label, confidence parameter, coordinates.

The Architecture of Proposed System:

The features used in the proposed model need to be pre-processed and divide into various stages. And each data has to come from the unity 3D world to process it. It has several steps from OpenCV to ML module to detect object.

1. Preprocessing unity 3D window

The object which are present in the 3D world are all the 3D objects the unity provides the custom-made 3D object which are available in Unity Assets store. After add the objects while running the computer vision will override with the character view in 3D world. This way we get the vision what the character seeing in the virtual world and then take those data and send it to the OpenCV to further processing as in figure [1].

2. OpenCV with object detection

The data Image which we get from the Unity will bel feed into the OpenCV dll module to detect the object which will effectually executed and print in the console the OpenCV will work as the computer vision to classified or cascade the object base on Haar cascade which is base on negative feeding and positive feeding to the module to cascade and train to detect which belong to the which class object label.

Deep neural networks, in particular, need a huge volume of training data. Furthermore, the images used in the training model make up a significant portion of the final model's output. It necessitates the collection of objective data of good quality. There are many datasets available for emotion recognition studies, ranging from a few thousand high-resolution photographs to hundreds of thousands of smaller images.

We train the network with GPU for 20 epochs to ensure that the precision converges to the optimal level. The more the data which feed the better the result will be to detect the object the improving the module will affect the result of the module. every second of time the module we feed the 20 frames to the module this we it can learn.

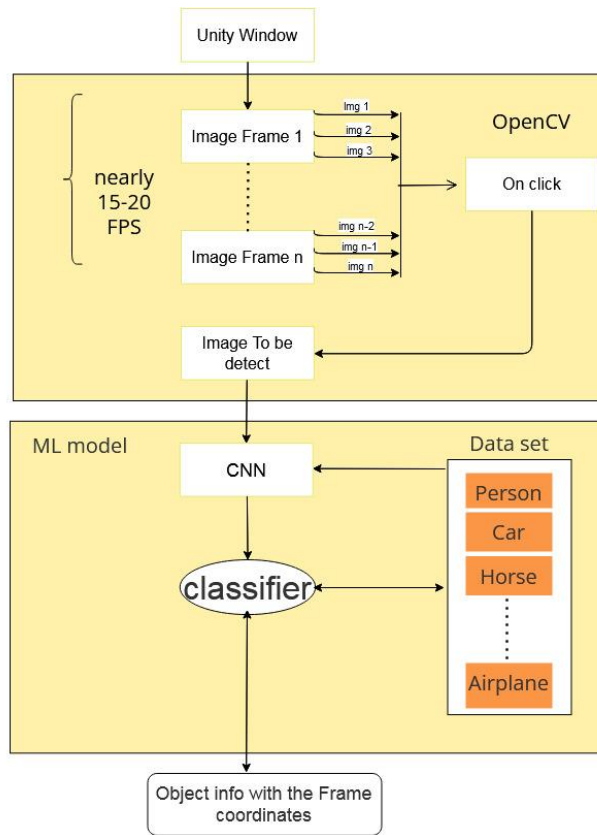


Figure 1: Simplified flow-chart for the entire system

Neural networks have proved to be inspiring in terms of computer vision tasks and performance as well. The performance of the module depend on the system GPU to train and evaluate the 3D object has been seen in the Unity Window and to be further processed to evaluate the data base on Image recognition to detect the object after the detecting the object we can train that object to move on the 3D module by using the MLAgent which is provided by the unity to train the object in the 3D environment according to conditions .

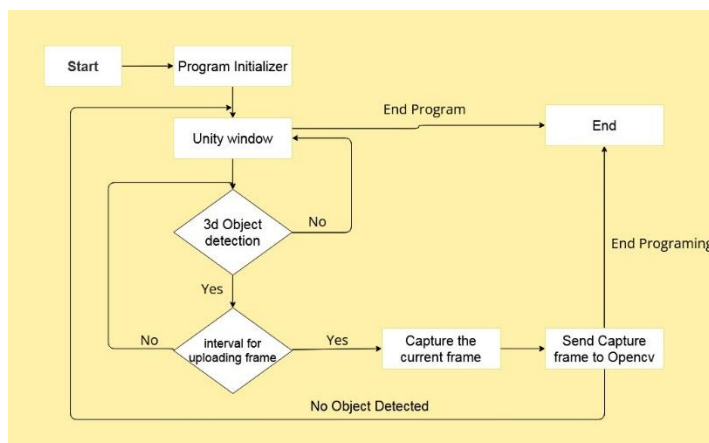


Figure 2: process of program structure

OpenCV,InitiallyOpenCV work on the Haar cascade where we feed the negative feedback as well as positive feedback to train the module to detect the object the data which is classified are will directly feed into the module to train then base on the it can differentiate as shown in the Figure [1], the object since we doing in the 3D world. the object that we detect

should be come from the 3D world and according to the screen it displays the value that it detects with the coordinate and confidence value. After detecting the object, we can make that object to move base on the ML agent which we train for basic move. Which work like chasing the target without hitting the wall which will done automatically to visualize the data which we train to move the object in the Unity world. The process of the module can be further hence by using the behavior module. Which will tell the module base on the label what to do walk move fly etc. this is the further enhancement.

Algorithms:

Algorithm for object detection from a 3D world/virtual world.

- Step 1: start the Unity Engine to enter the 3D world
- Step 2: Search for the object to be detect
 - 2.1 align the window to the object
 - 2.2 take the screenshot
 - 2.3 send to the OpenCV module
- Step 3: Feature Extraction from Images
 - 3.1 run the trained module
 - 3.2 detect the object from the frame
 - 3.3 revert back the value from the frame
 - 3.4 display in the user screen with
- Step 4: make the object move
 - 4.1 the object search for the target
 - 4.2 move till the target has achieved
- Step 5: Final end the simulation

The movement of the object has been trained by the ML agent to find the target and move accordingly the Brain. The OpenCV not only detect the single object at a time but also multiple objects as well. The data which in feed to the OpenCV while training is multiple objects present in the images. The process starts from the top layer which is unity window as shown in Figure [2] where we feed the unity window to our module where it detects the object if yes then take the screenshot of the frame and capture the frame to send to the OpenCV module to check if successfully detect the object, then exit the process.

4. RESULTS

The proposed system is to detect the object in 3D world but the Unity does not allow the python 2.7 above since because of the draw back we cannot directly implement the OpenCV rather we use the DLL for to do our work effectively the data which will the object can be detect are pretty good result like are 90 ~ 98 which is good interims of other algorithm or method like YOLO v3 or V4 etc. while making the object moving training we use the reinforcement learning to train the module which in below will be seen in Table[2]. While training we divide the module to train faster in train steps every step, we can see how many rewards it gets by mean reward.

Table 2: observations from the proposed method while training

Train Step	Step	Time Elapsed	Mean Reward
1	50000	110.343	0.549
2	100k	233.055	0.921
3	150k	352.577	0.971
4	200k	473.802	0.992
5	250k	599.837	0.996
6	300k	731.005	0.998
7	350k	124.877	0.766

8	400k	176.532	0.122
9	450k	234.865	0.654
10	500k	100.654	0.964
11	550k	444.821	0.992
12	600k	510.213	0.897
13	650k	123.999	0.156
14	700k	874.111	0.774
15	750k	735.032	0.987
16	800k	743.921	0.367



Figure 3: 3D object detection(Airplane)



Figure 4: 3D object detection (car)

Figure 3,4 shows the snapshot of different object present in the 3D world and detect base on the screen which the character is seeing. The character can move freely without the restriction and make and just have to click on in the screen and the object can automatically detect the object and give the result in the console.

Model accuracies are calculated based on the probability of correctness and perfectness in the prediction of tasks assigned as the 3D object detect by the OpenCV module the movement of the object is done by the MLAgent which use the reinforcement learning which means base on its environment it moves according and find the target to get the reward as show in Figure[5] since there is reward there has to be punishment as we can say loss which is show in the Figure[6].

Since as we can see the previous the reward is less as the training improve reward accumulated and the loss or punishment will reduce. This way the MLagent can learn to move according the environment.

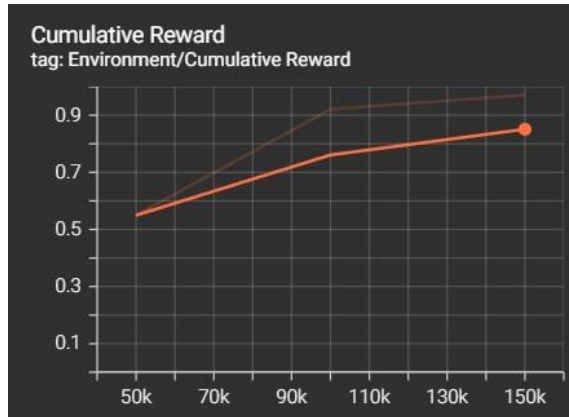


Figure5: Reward of the MLagent accumulated



Figure 6: Loss of the MLagentdissipate

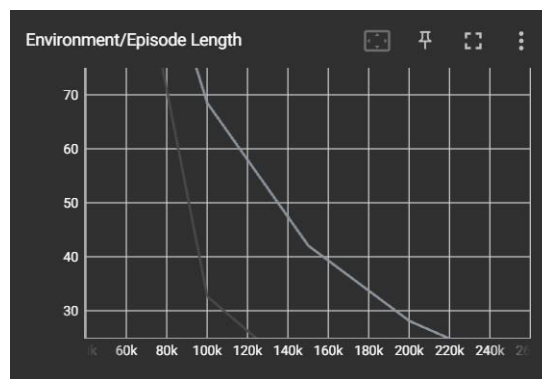


Figure 7: Final Model Accuracy and Losses at each Epoch

The final module which is already successfully train can able to detect the object and tell which categories it belongs as shown in the Figure[7] the more the training episode gets the better the module perfume in detecting the object around so

the down fall of the line shows that every episode the module reduces it false value. Table 2 shows Accuracy comparison between the proposed system and the existing system.

Table 2. Accuracy comparison between the proposed system and the existing system

Author	Methods used	Accuracy(%)
Weizh NIE [10]	PointNet++, K-means	83.2
IlyassQuazzaniTaybi [6]	2D slice,CNN	79.2
Xinying wang [11]	3D-DCGAN	88.5
Xudong Li [3]	CNN	94.2
Chandan G [5]	OpenCV, Deep Learning	85.3
Proposed Method	OpenCV, unity3D	96

5. CONCLUSION & FUTURE ENHANCEMENT

Finally, we conclude that our model is able to detect the object and finally able to tell which that object represents in the 3D world the accrue of the module will around 88~94 which is pretty good compare to other module and the requirement for training the data set are also light wait which take less time to train the module. In upcoming world virtual reality will take on the world this way the module will use and can be improve in the other fields of vision to make the user experience high and effective.

immense of this module is limit less this module can be used in other way, like for Example like 3D human object detecting the object joints legs, arms, head we can make the module train to move on its own with action and movements on its own which will open the door for new world to be expectance even not only in the game field as well in the health and industry field the user experience will be greatly improve by 70% which is terrifying number this will also reduce the code to develop the object movement in the 3D world.

6. REFERENCES

- Kenichi Sugihara, Yoshitugu Hayashi, "Automatic generation of 3D building models with multiple roofs", IEEE, Volume:13, 2008.
- Zuopeng Zhao, Zhongxin Zhang, Xinzheng Xu, Yi Xu, Hualin Yan and Lan Zhang, "A Lightweight Object Detection Network for Real-Time Detection of Driver Handheld Call on Embedded Devices", IEEE, Volume:7,2012.
- Xudong Li, Mao Ye, Yiguang, Ce Zhu, "Adaptive Deep Convolutional Neural Networks for Scene-Specific Object Detection", IEEE, 2017.
- Bernardo Augusto Godinho de Oliveira, Flavia Mangalhaesfreitasferreira, "Fast and Lightweight Object Detection Network: Detection and Recognition on Resource Constrained Devices", IEEE, Volume:6,2018.
- Chandan G, Ayush Jain, Harsh Jain, Mohana, "Real-Time Object Detection and Tracking Using Deep Learning and OpenCV" IEEE, Volume:7, 2018.
- IlyassQuazzaniTaybi, Taoufiq Gadi, Rachid Alaoui, "DSlicesNet: A 2D Slice-Based Convolutional Neural Network for 3D Object Retrieval and Classification", IEEE, Volume:9, 2018.
- ChengjiLiu,,Yufan Tao ,Jiawei Liang ,Kai Li,Yihang Chen "Object Detection Based on YOLO Network", IEEE , 2018.

8. Minyoung Park, Hyuntaek Lee, Hee-Jin Choi, "Augmented Reality Lightfield Display for a Smart Window Using an Active Pinhole Array" IEEE, Volume:7, 2019.
9. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", IEEE, Volume:13, 2019.
10. Weizhi NIE, YA WANG, DAN SONG, "3D Model Retrieval Based on a 3D Shape Knowledge Graph", IEEE, Volume:8, 2020.
11. Xinying wang, Dikai Xu, fangming Gu, "3D Model Inpainting Based on 3D Deep Convolutional Generative Adversarial Network", IEEE, Volume:8, 2020.
12. Min pang, fenggang Xiong, Xiaowen Yang, Xiehan, "Developing an Image-Based 3D Model Editing Method", IEEE, Volume:8, Issue:10, 2020.
13. Haidhi Zhu, Xin Yan, Hongying Tan, yuchoa Chan, "Moving Object Detection with Deep CNNs", IEEE, Volume:6, Issue:10, 2020
14. Mate kristo, Marina ivasic kos, Miranpober, "Thermal Object Detection in Difficult Weather Conditions Using YOLO", IEEE, Volume:8, Issue: 06, 2020.
15. Virginia Mamone, Vincenzo Ferrari, Sara Condino, Fabrizio Cutolo, Dipartimento di Ingegneria dell'Informazione, "Projected Augmented Reality to Drive Osteotomy Surgery: Implementation and Comparison with Video See-Through Technology", IEEE Conference, Volume:8, 2020.
16. Xingyu Pan Mengya Zheng, Xuanhui Xu, Abraham G. Campbell, "Knowing Your Student: Targeted Teaching Decision Support Through Asymmetric Mixed Reality Collaborative Learning" IEEE, Volume:9, 2021. .
17. Yushan Siriwardhana, Pawani Porambage, Madhusanka Liyanage, Mika Ylianttila, "A Survey on Mobile Augmented Reality With 5G Mobile Edge Computing: Architectures, Applications, and Technical Aspects" IEEE, Volume:23, 2021.
18. Jan Čejka, Marino Mangeruga, Fabio Bruno, Dimitrios Skarlatos, Fotis Liarokapis, "Evaluating the Potential of Augmented Reality Interfaces for Exploring Underwater Historical Sites" IEEE, Volume:9, 2021.
19. Udaya Dampage, D. A. Egodagamage, A. U. Waidyaratne, D. A. W. Dissanayaka, A. G. N. M. Senarathne, "Spatial Augmented Reality Based Customer Satisfaction Enhancement and Monitoring System" IEEE, Volume:9, 2021.
20. Joel Weijia Lai, Kang Hao Cheong, "Adoption of Virtual and Augmented Reality for Mathematics Education: A Scoping Review" IEEE, Volume:10, 2022