International Journal of Future Generation Communication and Networking
Vol. 13, No. 1, (2020), pp. 1289-1299

CANON OF PROPORTION

[1]Dr. R. Janarthanan, [2]Dr.R.Dhanagopal,[3]S.A. Geethalakshmi, [4]K. Nisha
[1,3,4]Department of CSE, Chennai Institute of Technology, Chennai, Tamilnadu, India.
[2]Department of ECE, Chennai Institute of Technology, Chennai, Tamilnadu, India.

Abstract

Humans can distinguish and recognize objects present in an image easily. The visual system of people is quick and precise with the goal that it can perform complex errands, for example, distinguishing various items and identify deterrents with negligible cognizant ideas. Be that as it may, for a machine, it's hard to act as needs be without earlier preparing or model. So, it requires a computerized vision as a medium to achieve the assignment. This paper proposes a model for tackling the issue of distinguishing the articles in casing of live catch, computing the distance between the object detected or distinguished which is having higher probability and the perspective (computerized vision). You Only Look Once (YOLO), is a very quick real time multi object/things recognition it is utilized right now in the proposed model. At that point the distance is estimated utilizing the triangle likeness. This model is significantly proposed for overhauling Drone administration framework (Drone service system) with no given physical mechanical control. Furthermore, the model likewise gives insight to computerized vision which can be applied to any sort of advanced digital vision given that it need not be sound system dreams like stereo visions. Thus, the last order is given to the machine holding advanced vision (Drone) in light of the high likelihood article to move in individual course which doesn't hurt environment.

Keywords: *Object detection, Distance Calculation, YOLO, Drone system*

I. INTRODUCTION

Object detection - most significant, interesting and demanding undergrowth of computer vision, [1-6], [8-22] which has been generally applied in every individual's life, to say as, monitoring security, self-directed driving, etc., to find cases of semantic objects of a specific class. With the fast advancement of profound learning calculations for recognition assignments, the object detector's throughput has been well enhanced. With the accessibility of a lot of information, quicker Graphics Processing Unit (GPUs), and better calculations, it is anything but difficult to prepare Personal Computer (PCs) to distinguish, perceive and arrange numerous articles inside a picture with high exactness regard to time. Image processing is constantly an interesting field as it gives upgraded visual data for human simplification and processing of picture information for transmission and delineation for machine perception. Digital images are processed to give a better solution using image processing. There are limitless circumstances which calls for image processing in the form of software. These are coupled into few gatherings: Verifying the presence, Object discovery and localization, estimation, distinguishing proof and confirmation. Almost, the captured image from the digital vision isn't legitimately processed inside the application. Rather, it is preprocessed, to upgrade the image according to the specific assignments. Here, the preprocessing which happens involves noise lowering, illumination & contrast up gradation. To utilize the camera as an estimating gadget, it must be adjusted to the physical world, where it alludes calibration as geometric and color. Geometric calibration corrects the distortion of lens which mainly ensures accurate reproduction of colors and determines the

ISSN: 2233-7857 IJFGCN

relationship between a pixel and real-world units, such as millimeters or inches. With regards to finding parts, typically coordinating is included: this implies searching for locales that are like or equivalent to a predefined layout which can either be a picture or a geometric example containing data and these techniques are called pattern matching as of two kinds as correlation and geometric individually. Most estimation procedures depend on anxious discovery calculations. An edge is a territory in an image showing a huge change in the picture intensity i.e., the product as software dissects the dark degrees of the image and distinguishes shape. To do so, recognizing the objects is done using the extremely fast real time detection algorithm at multi-scale called YOLO (You Only Look Once). It unifies the different parts of object distinguished into a single neural network. The structure employments highlights from the whole picture to foresee each bounding box over all classes for a picture all the while. The YOLO conformation allows end-to-end training and real-time speeds while keeping up high normal accuracy. This conceptual model is proposed to be implemented in Drone service system where these drones with digital vision can act without any prior physical control required. This relation can be implemented in several ways by applying the solution to the alternatives of methodology wherever the digital vision plays a major role to produce the result such as self-driving cars, robotic motion [22], drone delivery system, traffic management system, license plate scanning in tolls, video surveillance, etc.

II. YOLO

Using this technique, you only look once (YOLO) at a picture to foresee what articles are available and where they are. YOLO is refreshingly basic. The algorithm applies a neural system to a whole picture. The system divides the picture into a S x S grid and comes up with bounding boxes, which are boxes drawn around pictures and anticipated probabilities for every one of these locales. This brought together the model with few advantages over conventional techniques for object discovery. To begin with, YOLO is amazingly quick. Since the frame recognition taken as deterioration problematic don't prerequisite a composite pipeline. It just runs for neural network on another picture at test time to foresee and predict detection. The base system runs at 45 frames every second with no group preparing on a Titan X GPU and a faster version runs at in excess of 150 fps. This implies we can process live streaming video continuously with under 25 milliseconds of latency. Moreover, YOLO accomplishes more than double the mean average accuracy (meticulousness) of other real-time systems. Second, YOLO reasons all-inclusive about the picture when making prediction.

Unlike sliding window and region proposal-based techniques, YOLO sees the whole picture during training and test time so it certainly encodes contextual information about classes just as their appearance. Third, YOLO learns generalizable portrayals of objects. At the point when trained on natural pictures and tested on work of art, YOLO outflanks top discovery strategies like DPM and R-CNN by a wide edge. Since YOLO is exceptionally generalizable it is more averse to separate when applied to new domains or unexpected inputs. YOLO still falls behind state-of-the-art detection systems in precision. While it can rapidly distinguish objects in pictures it struggles to accurately limit a few items, particularly little ones. An assortment of pre-trained models is likewise available to download. One set of class probabilities per grid cell, paying little heed to the quantity of boxes B. At test time it multiplies the conditional class possibilities and the distinct box self-confidence calculations,

$$\mathbf{Pr(Classi\,|Object)*Pr(Object)*IOU = Pr(Classi\,)*IOU}$$

ISSN: 2233-7857 IJFGCN

this will provide us class-specific confidence scores for each box. These scores encode both the possibility of that give the impression in class-specific of the box and how well the prophesied box hysterics the object.

Additionally, YOLO v3 calculation comprises of completely CNN and a algorithm for post-processing output from neural network. CNNs are exceptional design of neural systems appropriate for preparing grid-like information topology. The distinctive feature of CNNs which bears significance in object discovery is parameter sharing. In contrast to take care of forward neural systems, where each weight parameter is utilized once, in CNN design every individual from the piece of kernel is utilized at each position of the input, which means learning one set of parameters for each area rather a different arrangement of parameters. This element plays significant job in catching entire scene on the frame. CNN accepts an image as a input and yields tensor which characterizes:

- Coordinates and positions of predicted bounding boxes which ought to contain objects,
- A likelihood (probability) that each bounding box contains object,
- Probabilities that each article inside its bounding box belongs to a particular class.

This network has a capacity to all the while recognize different items on the single input image. Features are picked up during the network training process when the network analyzes the entire input image and does the predictions. In that manner, the network knows about the entire landscape and objects environment, which helps the network to perform better and accomplish higher accuracy results contrasting with the techniques which utilize the sliding window approach. The idea of separating the images to grid cells is unique in YOLO, when contrasted with other object detection solutions. The input picture is divided into a S x S grid of cells where every framework of grid cell can anticipate 3 bounding boxes.

OPENCV

Open Computer Vision is a computer-based vision system in which the system can able to view the environment as the human could. This makes the system to understand and can able to perform some computations based on the things viewed. OpenCV may be a library of programming features within the foremost aimed toward time period PC computer vision originally developed by way of Intel. The library is cross-platform and free to be used below the ASCII text file BSD license. It is a library which may consists of more than 2500 optimized algorithms. These calculations are frequently want to see and recognize faces, build up objects, order human activities in recordings, pathway camera developments, pathway moving articles, remove 3D models of items, fabricate 3D reason mists from sound system cameras, sew pictures along to give a high goals picture of an entire scene, notice comparable pictures from an image data, remove red eyes from pictures taken exploitation stripe, follow eye developments, recognize view and set up markers to overlay it with expanded authenticity, and so forth.

III. RELATED WORK

According to the research made on the problem space, there exist few methods, techniques and algorithms. These are based on object detection and distance calculation which have been respectively done separately. The object detection is actually relies on few perspectives such as weather condition which plays an important role for detecting and recognizing objects in the vision frame. Also, the existing work depends on some physical sources that acts according to it. One of those includes the viewpoint i.e., camera. The vision system is the most significant part to discover the objects in real-time. Some of the existing system

ISSN: 2233-7857 IJFGCN

uses various sensor technologies for reading the data like LiDAR, radars and ultrasonic devices. These kinds of sensors are quite a bit expensive when compared to the digital visions. Likewise, some of those are applicable only by using stereo types of cameras and other systems are only proved by proposing solutions but unable to achieve in real time. The main factor resides on the accuracy, precision loss and speed. These keep improving better a bit and not with any good averages still.

In paper [1] proposed a real time software for detecting traffic participants as the solution for the next generation of autonomous cars for solving the problem of perceiving and understanding their surroundings in Advanced Driving Assistance Systems by using YOLO v3 algorithm which specialize the neural network and tracks at least more than three objects. The 5 object classes has been trained in the network such as car, truck, pedestrian, traffic signs and lights and also by demonstrating effectiveness of the approach under multiple driving conditions as bright and overcast sky, snow, fog and night. In addition to this, the solution proposed uses input from one camera. But Precision can be enhanced with training on the bigger and more miscellaneous datasets.

In paper [3] devises a solution for object detection and computes the distance between the object and the viewer under foggy conditions by introducing VESY (Visibility Enhancement Saliency YOLO) sensor. This is provided by considering the risk of road accidents which increases as the visibility decreases while driving under foggy conditions. It uses a stereo camera to detect object and the frame applied to dehaze algorithm and YOLO algorithm. Then the distance is calculated via depth map. Only few objects are able to pass high threshold.

Reference [4] proposes a new method for measuring the distance accurately using frequency-domain analysis which also uses stereo camera as an alternative to sensor technologies for calculating distance. But accuracy is improved only ~10% better and limits when applied to shadow regions.

Reference [7] integrates ideas such as attention mechanism and feature pyramid network and proposes an upgraded network as a solution to object detection model in remote sensing image which precisely detects objects of varying sizes by enhancing characteristics abstraction technique on multi scale. Real-time detection speed cannot achieve in this technique.

Reference [9] provides the detailed insight of image processing and various identification techniques.

IV. PROBLEM SPACE

Imagine a situation, In a hospital with 15 floors and a patient in the 12^{th} floor is in need of few medicines immediately and the patient's attender needs to get those medicine soon from the pharmacy within the hospital which is in ground floor while the patient suffering from the need of mental support from their attender, in this case it is impossible to handle both ends.

Considering Drones as a solution to this problem for delivering medicines to the patient's place cam serve as a solution to handle the critical situation. But to get happen this structure, the camera in the drone should be loaded with the technology where it can find the obstacles in its path and makes its path respectively which does not harm the environment. This needs to be solved technically by processing the

ISSN: 2233-7857 IJFGCN

live stream through digital vision. Hence the statement of problem is described using the terms: Object Detection + Distance Prediction + Intelligence.

V. PROPOSED SOLUTION

Canon of Proportion deals with the mathematical measurement that describes the rational approach to constructing beauty in art i.e., A system of mathematical ratios constructed on dimensions of fragments of the human body, designed to create ideal proportions for the human figure in art and highlights that "height and width have a definite geometrical relation to one another". The concept is expressed with the intelligent object detection which is one of the key software components in the next generation of artificial intelligence and machine learning. The above intelligent object detection can be done by perceiving requirements via dataset and understanding the surroundings, initially by recognizing multiple objects in a single frame and then calculating the distance between points is related to the distance prediction between the vision and the object found. This relation can be implemented in several ways by applying the solution to the alternatives of methodology wherever the digital vision plays a major role to produce the result such as self-driving cars, robotic motion, drone delivery system, traffic management system, license plate scanning in tolls, video surveillance, etc. Thus far, distance measurements to objects are stereotypically passed out using light detection and ranging [LIDAR] or by utilizing sensor technologies, such as radar, laser scanners and Global Positioning System (GPS), but these sensors are expensive compared with camera, many suppliers are investigating the development of camera-based distance measurement systems. Also, it is predicted that well-organized technologies to repossess appreciated information from these sensors will be crucial in future generations.

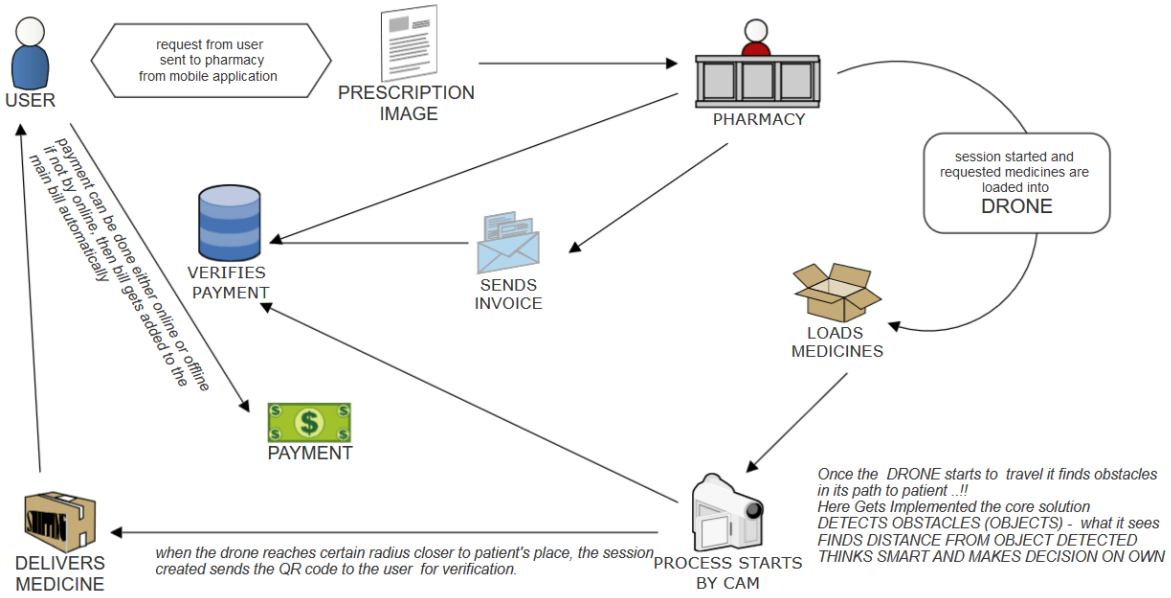

Fig 1. Process flow diagram of the proposed model

The solution is proposed as a software, wherein it can be installed in any of the above applications as per the requirement. Here, the implementation is dealt with the drone delivery service in hospitals. Actually, the drones operate using hardware computing device attached and programmed to it. Whereas, this idea

ISSN: 2233-7857 IJFGCN

provides the generic solution as an alternative to such hardware devices used in drones and without any physical control mechanism which needed to be provided. Everything over the service is pre-programed by inducing intelligence to camera present in drones. The action such as identification of objects in the path which the drone moves YOLO [You Only Look Once] algorithm, then detecting the kind of object identified, predicting the distance between the object and drone are the primary algorithmic working and then comes the important part of providing intelligence to vision of drone. It is given based on the distance measured and the object detected: If the object is stationary, then the command to drone is provided either to move above or below the object, If the object is not stationary and its motion is in the same direction similar to drone or in the opposite direction, then the command is provided accordingly in a manner which do not disturb surrounding by any means. Whenever the request is raised, a unique session id is created and medicines are set ready to deliver via drone. Then the path, the drone travels is extracted using the Google Maps API and sent. Once the drone reaches the patient's place, the session id which is created on the user side, generates a QR code for security verification once the payment is done. By this way, the method of canon of proportion describes its implementation. It can be simulated, trained and can be used for any of the applications listed respectively without using any kind of hardware computing devices.

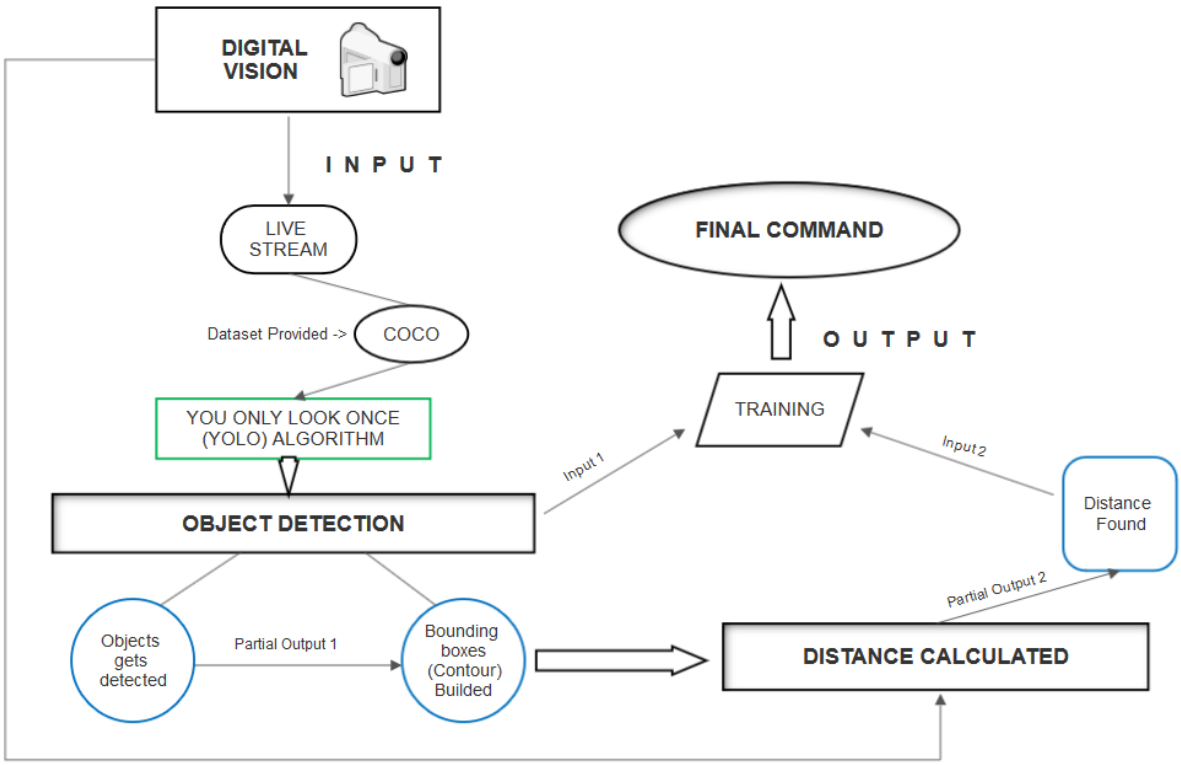

Fig 2. Architecture diagram

OBJECT DETECTION

The Object detection phase in the proposed model is done using YOLO (You Only Look Once) algorithm which is encoded with python-OpenCV to enable the digital vision get accessed when executed. Once the code gets executed the objects which are detected is represented by building contour (bounding boxes)

ISSN: 2233-7857 IJFGCN

around it and displays the text of what object is detected and also with the probability of each discovered element in the frame. The objects are discovered and recognized by the dataset provided. And here the proposed solution uses COCO dataset. The accuracy of recognizing objects are quite improved than the existing ones.

DISTANCE CALCULATION

The distance between the object detected and the viewpoint to be calculated, we use a small calculation. It's basic geometry: you have a right-angle triangle, with half the FOV as one of the angles (a), and half your image size as the opposite side (A).

To calculate the focal length Ƒ , use - **tan(a) = A/ Ƒ, which gives Ƒ = A/tan(a).**

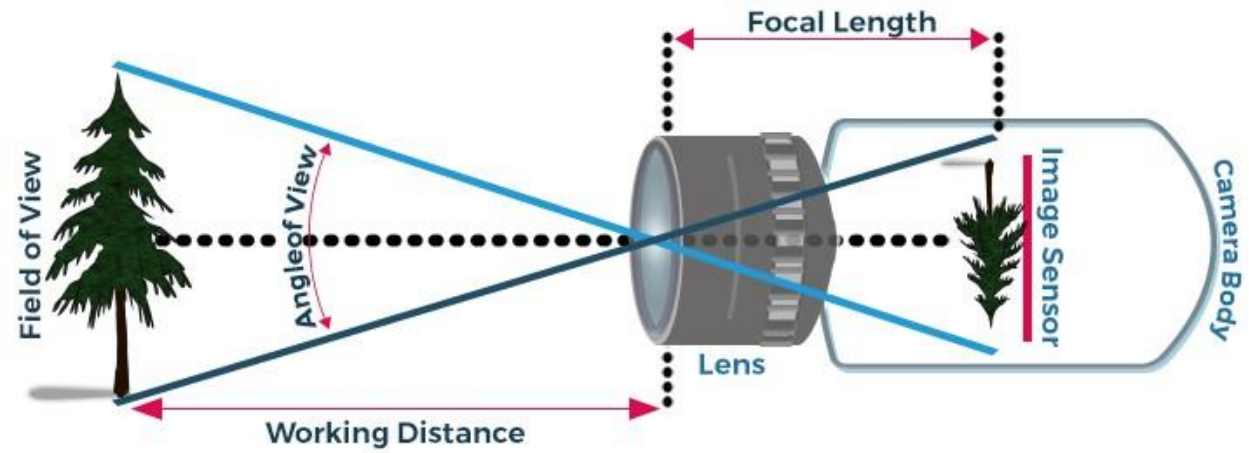

Fig 3. Distance calculation using field of view

As the sensor size is given in pixels (assumed square pixels), the focal length will also be in pixels.
To get it in a more usual unit (m), we need to know the pixel size.
The triangle resemblance goes somewhat like this: Let's say we have an object with a known width W.
Then, we place these objects some distance D from our camera.
We can take a picture of an object using our camera and then measure the deceptive width in pixels P.
This allows us to derive the Ƒ of our camera (Ƒ is the perceived focal length):

Ƒ = (P x D) / W

If the width W is unknown then use the method given below to find the W in pixels:

W = Total width of a frame in pixels – (Number of pixels from left frame to contour of object detected + Number of pixels from right frame to contour of same object detected)

And apply in the formula above to get the focal length.

ISSN: 2233-7857 IJFGCN

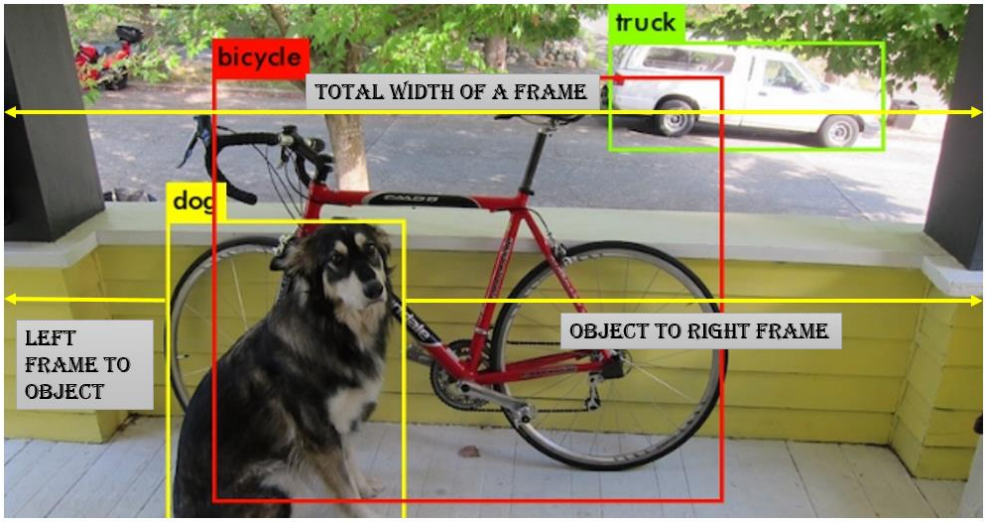

Fig 4. Calculating unknown width (W) of an object detected

VI. EXPERIMENTS

After various experiments and testing of this application, it provides a better output. The aim of our proposed method is to present the solution which could detect and recognize multiple objects in live stream and calculates the distance between the object discovered and the camera (digital vision) and final command is provided which enables the drone to move in direction respectively that doesn't harm the present surrounding.

Here the final command is passed only when the object discovered is found closer to the drone which is established using the distance score. Once the distance score is less than fixed probability of collision it decides to move on the command and takes its appropriate direction.

Fig 5.a

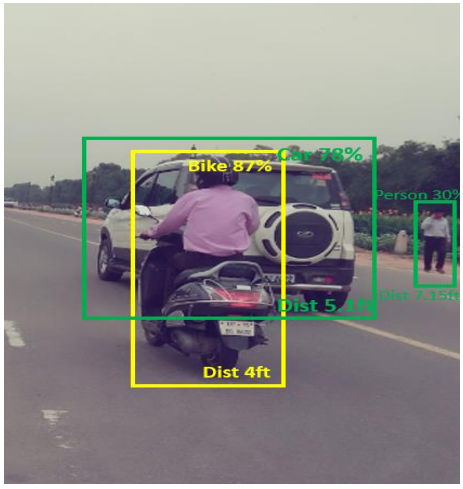

Fig 5.b

ISSN: 2233-7857 IJFGCN

Fig 5.c

Fig 5.d

Fig 5 (a, b, c, d) describes the experiments on object detection and distance calculation on various scenarios. [The below Table - 1 represents the object detected in the above figures in which yellow color bounding boxes represents the main focused object and distance calculated corresponding to it].

Table – 1. Distance prediction of the object detection of the figures

Fig No.	Object Detected	Distance Predicted
5.a	Person	3 ft
5.b	Bike	4 ft
5.c	Person	3.2 ft
5.d	Person	3.2 ft

VII. CONCLUSION

In this paper, the solution has proposed to unravel the problem of real time critical situation happening in hospitals which is detailed in the given problem space. It could gain the mentally weak patients to have their helpers with them and also benefits the core where the digital vision could act smart: that decides on their own based on what it sees. Intelligence in form of inducing programmatically to machines greater the chance of viewing the technology growth that makes routine life of people digital in all. This solution needs a High-resolution camera for the more accuracy and perfect result to be delivered. Future implementations can be done in any industry where digital vision plays a major role. The proposed solution is purely based on the problem space given.

REFERENCES

[1] AleksaĆorović, VeliborIlić, SinišaĐurić, MališaMarijan, and BogdanPavković "The Real-Time Detection of Traffic Participants Using YOLO Algorithm" 26th Telecommunications forum TELFOR 2018, November 2018.

[2] Sandeep Kumar, AmanBalyan and Manvi Chawla "Object Detection and Recognition in Images" 2017 IJEDR, vol. 5, ISSN: 2321-9939.

ISSN: 2233-7857 IJFGCN

[3] SarthakKatyal, Sanjay Kumar, RonakSakhuja, and Samarth Gupta "Object Detection in Foggy Conditions by Fusion of Saliency Map and YOLO" 2018 Twelfth International Conference on Sensing Technology (ICST)
[4] In-Sub Yoo, Seung-Woo Seo "Accurate object distance estimation based on frequency domain analysis with a stereo camera" IET Intelligent Transport Systems: Research Article, March 2017.
[5] Guangbin Wu , Weishan Chen , Hui Cheng , WangmengZuo, David Zhang, and Jane You "Multi-Object Grasping Detection with Hierarchical Feature Fusion" IEEE open Access Journal, April 2019 (Digital Object Identifier 10.1109/ACCESS.2019.2908281).
[6] Nils Bore , Johan Ekekrantz , PatricJensfelt and John Folkesson, "Detection and Tracking of General Movable Objects in Large Three-Dimensional Maps" in IEEE transactions on robotics.
[7] Xiang Ying, Qiang Wang, Xuewei Li, Mei Yu, Han Jiang, JieGao, Zhiqiang Liu, Ruiguo Yu, "Multi-Attention Object detection model in remote sensing images based on multi-scale" IEEE Open Access Journal, July 2019 (Digital Object Identifier 10.1109/ACCESS.2019.2928522).
[8] Licheng Jiao, , Fan Zhang, Fang Liu , Shuyuan Yang , Lingling Li , Zhixi Feng, and RongQu "A Survey of Deep Learning-Based Object Detection" IEEE open Access Journal, September 2019 (Digital Object Identifier 10.1109/ACCESS.2019.2939201).
[9] Sahil V. Khedaskar, Mohit A. Rokade, Bhargav R. Patil, Tatwadarshi P. N. "A Survey of Image Processing and Identification Techniques" VIVA-TECH international journal for research and innovation, 2018.
[10] Felzenszwalb, P. F., Girshick, R. B., McAllester, D., &Ramanan, D. (2010). Object detection with discriminatively trained part-based models. IEEE transactions on pattern analysis and machine intelligence, 32(9), 1627-1645.
[11] Goodfellow, I., Bengio, Y., Courville, A., &Bengio, Y. (2016). Deep learning (Vol. 1). Cambridge: MIT press.
[12] Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).
[13] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).
[14] Redmon, J., Divvala, S., Girshick, R., &Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).
[15] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.
[16] Springenberg, J. T., Dosovitskiy, A., Brox, T., &Riedmiller, M. (2014). Striving for simplicity: The all convolutional net. ArXiv preprint arXiv:1412.6806.
[17] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
[18] Lin, T. Y., Dollár, P., Girshick, R. B., He, K., Hariharan, B., &Belongie, S. J. (2017, July). Feature Pyramid Networks for Object Detection. In CVPR (Vol. 1, No. 2, p. 4).
[19] Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics), 28(1), 100-108.
[20] Yu, F., Xian, W., Chen, Y., Liu, F., Liao, M., Madhavan, V., & Darrell, T. (2018). BDD100K: A Diverse Driving Video Database with Scalable Annotation Tooling. arXiv preprint

ISSN: 2233-7857 IJFGCN

arXiv:1805.04687.

[21] Konar, A., Chakraborty, A., Halder, A., Mandal, R. and Janarthanan, R., "Interval type-2 fuzzy model for emotion recognition from facial expression", In *Indo-Japanese Conference on Perception and Machine Intelligence, Jan 2012,* pp. 114-121, Springer, Berlin, Heidelberg.

[22] Saha, S., Datta, S., Konar, A. and Janarthanan, R., "A study on emotion recognition from body gestures using Kinect sensor", In *2014 International Conference on Communication and Signal Processing,* pp. 056-060, IEEE.

[23] Muthukumar, B., Dhanagopal, R. & Ramesh, R. KYP modeling architecture for cardiovascular diseases and treatments in healthcare institutions. J Ambient Intell Human Comput (2019). https://doi.org/10.1007/s12652-019-01653-z

ISSN: 2233-7857 IJFGCN