

# Object Detection Distance Estimation

<sup>1</sup> K Usha Rani,<sup>2</sup> Puram Shravya, <sup>3</sup>Kotam Sreeja Reddy

<sup>1</sup>Associate Professor and HOD, Department of CSE, Bhoj Reddy Engineering College for Women, Hyderabad, Telangana, India

<sup>2,3</sup> Students, Department of CSE, Bhoj Reddy Engineering College for Women, Hyderabad, Telangana, India

## ARTICLE INFO

### Article History:

Accepted: 01 July 2023

Published: 10 July 2023

### Publication Issue

Volume 9, Issue 4

July-August-2023

### Page Number

94-101

## ABSTRACT

Artificial intelligence technology is developing rapidly in recent years. The purpose of this paper is to measure the distance to an object using this. In order to measure the distance, two separate pictures from same angles of the object will be taken. It extracts sizes for the same object in two pictures. In order to do this in real time, object detection technology of Artificial Intelligence on mobile phone was used. In this paper, a method for measuring the distance from two pictures is presented. The proposed method was implemented as a prototype on iOS. In order to measure the performance of distance measurement, experiments were conducted in various environments. In the experiments, the empirical data yielded some discrepancies with the actual measurement. This was a result of errors occurring in the object detection process where the actual size of the object was calculated. Despite these discrepancies, this method of object detection may be widely used in instances where accurate measurements are not necessarily required such as guidance systems for the visually impaired.

**Keywords:** Artificial Intelligence, Object Detection, Camera, Focus Distance, Distance.

## I. INTRODUCTION

Artificial Intelligence has been developing with great speed and there are already numerous real-life cases in diverse industries where Artificial Intelligence is used to facilitate their work. This is because Artificial Intelligence is yielding actual, tangible results. Deep Learning, an Artificial Intelligence technology that uses artificial neural networks, is a lead in the industry. There are already many researches conducted to create application software that applies such Deep Learning

technology. One of the most representative researches regarding the Deep Learning Model is the Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) [1-4]. CNN is used primarily to recognize images. Technologies that use CNN include, but are not limited to, image classification, object detection within a given image, style transfer, and object segmentation [1]. RNN is used to recognize sequential information such as letters or voice [1,2]. RNN is structured in a way that a previously entered value/information affects whatever value/information

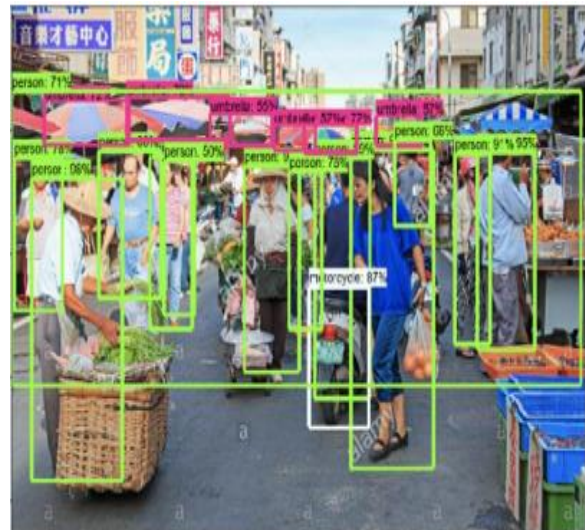
is entered next [2-4]. This is because, for instance, when there is a need to recognize/analyze a sentence, words must be interpreted based on the context of the words before and after it. Thus, RNN is used primarily for, but not limited to, voice recognition, machine translation, and image descriptions. These technologies are being actively applied in the development of apps for mobile phones [5-7]. This is because mobile phones nowadays have surpassed its conventional purpose as a mere telephone and have evolved into a small computer. Also, while Artificial Intelligence is indeed used to provide solutions to a more macro-level problems such as weather forecasts, it is also used to provide useful information and solve every day, micro-level problems of individuals. Thus, various Artificial Intelligence based mobile phone apps including object detection, measurements of objects are being actively developed. Since mobile phones have cameras that can help assist software, the development of such apps is expected to grow even further. This paper aims to develop methods of measuring objects placed in front of the observer (the camera) using object detection, which is an Artificial Intelligence technology. This can be achieved simply by using the camera. If the actual size of the object and the size of the object within the photographed image is given, it is generally very easy to calculate the distance between the object and the camera [8-10]. However, unfortunately, in many cases the actual size of the object is unknown. This paper will introduce a method to calculate the distance from a given object when the actual size of the object is unknown. Since the actual size of the object is unknown, an alternative set of information about the object is required. This paper uses the camera's movement as the alternative [9]. That is, rather than a single photo shot of the object, multiple photo shots of an object at a certain distance can be used to calculate the distance between the camera and the object. This paper uses a camera with a fast rate of reaction for the sake of real-time calculations, and also an Artificial Intelligence based object detection technology to measure real time the size of the object within the

photographed image. Thus, the proposed system was brought to life in an iPhone 8 plus, and to verify the effectiveness of the system, the distance between the camera and several objects in various environments were measured. We were able to infer from the empirical data that there are still somewhat substantial errors in the measurements. This is because Artificial Intelligence based object detection is still not yet detailed and accurate enough. However, despite these errors, the convenience and real-time measurements may very well be applied in various fields of studies. These may include guidance systems for the visually impaired.

## II. RELATED WORK

Recent developments in the Artificial Intelligence technology are truly amazing. And with such rapid developments in the field, various researches to utilize Artificial Intelligence are being conducted every day. Machine Learning is one of the most representative Artificial Intelligence technologies. Just like humans that grow more intelligent as new information is studied, Machine Learning enables the computer to study and accumulate knowledge to make better and educated decisions. Numerous studies and researches have been conducted in the area, including the development of Deep Learning. Deep Learning is a new technology created in light of the development of Artificial Neural Networks. This functions like the human brain. CNN and RNN are one of the most well-known products of the Deep Learning Model [1,2]. CNN is used primarily to identify images. Images can easily be altered into completely new and different files in the eyes of the computer with even the most minor tweak in direction or location. CNN is capable of recognizing images regardless of these alterations. It applies continuously the process of Convolution and Polling to extract abstract information about the image. RNN is used primarily to recognized sequential information such as letters and voices. The RNN is programmed so that a previously entered

value/information affects the interpretation of the value/information entered afterwards. This is because, for instance, when there is a need to recognize/analyze a sentence, words must be interpreted based on the context of the words before and after it. Thus, RNN is applied in various technologies such as voice recognition, machine translation, and image descriptions. Conventionally, CNN has always been used for image classification [5]. That is, when a certain image is given, the CNN classifies whether the object in the image is a dog or a cat, etc. Unlike the conventional system that analyzes all the pixels in the image to produce a result, CNN extracts certain key features of the image into the Neural Network to classify what the image is. It is up to the user to determine which parts of the images are key features to be used in the process of classification. For instance, one may want to set the color, length, and shape as key features to differentiate a banana from an apple. CNN automates this process. Within the many models of CNN, the object detection model is used most widely. This model detects objects within one image, as can be shown in figure 1. It can also identify and classify what the detected objects are. In figure 1, it has identified several people and a basket. The numbers in the figure indicate the certainty of the classified results. There are numerous libraries of object detection technology including tensor flow, YOLO, Image AI, Detectron2: PyTorch-based modular object detection library [5, 7]. The tensorflow is the most widely used [5].



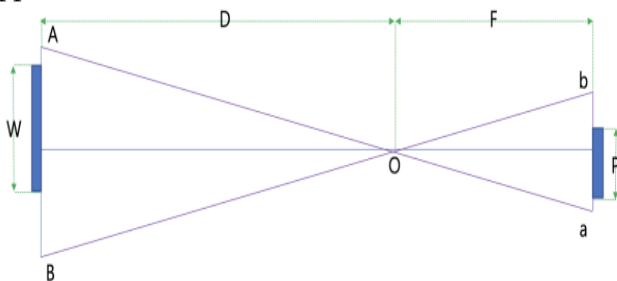
**Figure 1 An example of object detection**

Autonomous vehicles, multi drones, and robots are all new technologies experiencing a boom in light of recent developments in Artificial Intelligence [11, 12]. All these new technologies require essentially the technology for the computer to recognize its surroundings. Amongst the technologies that enable them to do so, one of the most pivotal technologies is one that allows the computer to recognize/measure the distance with other surrounding objects. Currently, Radar and Lidar are most commonly used to achieve this. Radar calculates distances through reflected electromagnetic waves. Radar allows the computer to calculate distances to a fairly accurate degree even in night times where there is little light. However, the further away objects are, the more errors there are in Radar measurements. Lidar determines the distance, shape, and material of its surroundings through the reflection of its beams [13-15]. Lidar is resistant to harsh weather conditions and also has a shorter wavelength than Radars. Depending on where/what these technologies are applied to, the accuracy and speed of measurements required may differ greatly. This paper aims to analyze images inputted by the camera -with the help of Artificial Intelligence technology- to measure distances between objects and the camera.

### III. PROPOSED SYSTEM

Recently, there have been many attempts to use the camera as a means of measuring the size of an object or the distance between the camera and the object. This is because, as most mobile phones come with a camera nowadays, cameras have become increasingly accessible every day. This paper proposes methods to calculate distances with certain objects using the cameras attached on mobile phones and conduct several experiments to verify its effectiveness. Artificial Intelligence is used to identify objects and calculate distances in front of the camera.

Measurement of the distance between the camera and a given object is calculated through the principles of how a camera works. If the actual size of the object is given, it is very simple and easy to calculate the distance between that object and the camera. This is because it is very easy to understand how a camera works. figure2 is a simple explanation of how a camera works. The left side of figure 2 indicates an actual object, and the right side indicates how that object appears on the camera.



**Figure 2 Relationship between an object and it's reflection on the camera**

Point O in figure 2 is where the camera's lenses are. AB is where the actual object is located, and ab is the image of the object appearing on the camera. In this arrangement, triangles ABO and abO are similar figures.

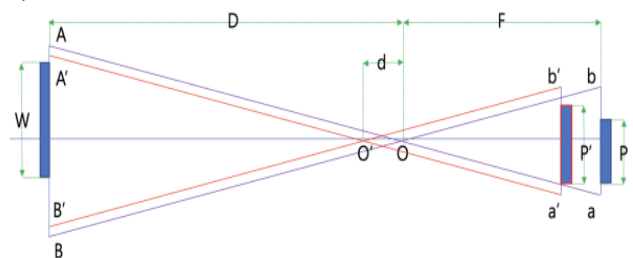
Thus, the following equation can be inferred.

$$W : D = P : F \quad (1)$$

$$D = W \times F / P \quad (2)$$

W is the actual size (height or area) of the object. P is the size of the image of the object reflected on the

camera. F is the focal distance of the camera. Since ABO and abO are similar figures, the equation 1 definitely holds. Equation 2 can easily be inferred from equation 1. The value of F in equation 2 is usually a constant. If the value of P and W are given, the value of D can easily be calculated. This paper, under the assumption that F and P can effectively be measured, aims to measure the distance between the camera and the object. The problem with equation 2 is that, in order to measure distance real-time, the size of the object must always be given. This is impossible in a real-life environment. Since the actual size of the object is unknown, an alternative set of information about the object is required. This paper proposes the camera's movement as the alternative. That is, rather than a single photo shot of the object, multiple photo shoots of an object at a certain distance can be used to calculate the distance between the camera and the object.



**Figure 3 Relationship between image and object after movement**

Consider figure 3. It shares the same settings as figure 2, but with the camera having moved towards the object by the amount of d. Thus, everything else remains the same, including all the previous values of the object, but the camera only is closer by d distance than the object. In figure 3, triangles abO and a'b'O are the exact same triangles and triangles ABO, A'B'O, abO, a'b'O are all similar figures. We can infer from this fact the following equations.

$$P' > P \quad (3)$$

$$W : (D - d) = P' : F \quad (4)$$

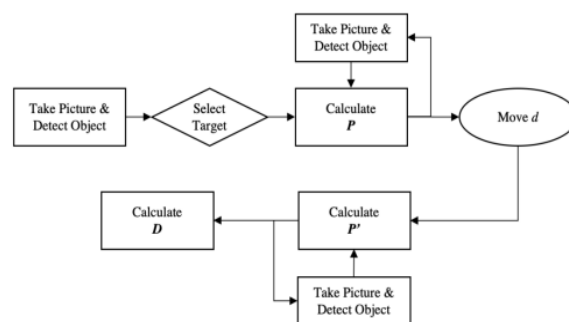
$$D = d + W \times F / P' \quad (5)$$

Using equations 2 and 5, the following equation can also be inferred.

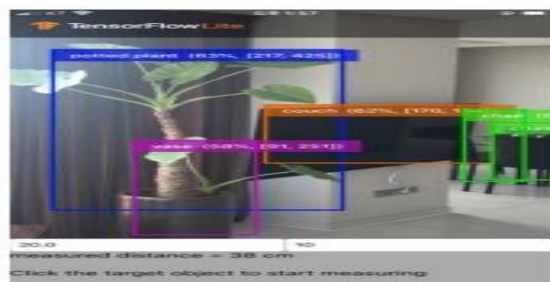
$$D = d \times P' / (P' - P) \quad (6)$$

Equation 6 does not require  $W$  which is the actual size of the object. Thus, even without the value of  $W$  the system is still capable of calculating the distance between the object and the camera. This does, however, require two more measurements of  $d$  and  $P'$ .

It is fairly easy to measure the size of an object in an image. It can be achieved through the simple process of measuring the object within the image and displaying the value through pixels. As can be seen from (6), the measuring units of  $P$  and  $P'$  are not important unless they use a single common unit of measurement such as pixels or inches. It was the general understanding that measuring objects in images and videos is extremely difficult due to the various other objects that are also in the image or video. This, however, has changed as recent developments in Artificial Intelligence have achieved great success in image classification. Image classification is the technology to identify and recognize certain objects within images and comparing them to pre-learned data to classify what the objects are. This technology extracts object information such as size and names from images through pre-learned data. This usually takes around 300msec per image. Since there is no preemptive process required to analyze the image, this process of image classification can be considered real time. figure 1 is a screenshot of object detection. This paper aims to measure  $P$  and  $P'$  in figure 3 using such object detection technologies. To measure  $D$  in equation 6, the value of  $d$  must be measured. This can be measured through the simple process of moving the camera back and forth to a desired amount. For the purposes of this paper,  $d$  (the amount of which the camera moves) is set to 20cm. That is, the object is identified at the current location to measure  $P$ . Then, the camera will move 20cm forward and identify the object again at that location.



**Figure 4 Algorithm for measuring the distance**



**Figure 5 Screen shot for measuring the distance**

A measuring software was developed to evaluate the accuracy of the proposed method. An iPhone 8 plus running at iOS 13.1 was used as the carrier for the software, with the software itself being created using Xcode's swift. Google's tensorflow-lite was used as the object detection model. In this particular experiment, it is important to select the right object. As can be seen in figure 1, an object detection software tries to identify as many objects as possible in any given moment. This test aims to measure the distance between the camera and the object closest to it.

The general algorithm of the developed software is as shown in figure 4. It starts by taking a photo, and then identifying objects through the object detection module. This module identifies any and all objects within the picture and calculates and displays the rectangles surrounding the object as well. The analysis module attempts to sort out the target object. In this paper, objects in the front are selected as target objects. If the selection of target objects is unclear, the software requests another photo shot or object detection. It analyses any new input and will request movement in order to measure the value of ' $d$ ' from (6). After having moved the amount of ' $d$ ', another picture is captured and the size of the target object is measured. figure 5 is

a screen shot of the software in operation. As can be seen in the figure, the large rectangle in the left is tagged “potted plant” with an indication that it is 63% sure that the object is in fact a potted plant. Also, the object has a height of 425 and a width of 217, both in pixels. There are various methods to selecting a target object, but it is not easy. The object detection modules that we currently use are extremely fluctuating in that the target object could change at any given point even when the camera is standing still.

#### IV. RESULTS AND ANALYSIS

In order to evaluate performance, several photos of various real-life settings were taken through the developed app whilst it also measuring P and P'. The application takes photos every 300msec and detects objects within those images. Because the detection process is so fast, even the smallest change in images can yield sensitive changes in detection results. Thus, it is more likely for errors to occur if object detection were to be conducted only once. In table 1, it shows the measurement of the distances according to d at each distance.

The d values were fixed at 10 cm, 20 cm, and 30 cm. These are distances measured at actual distances of 1m, 1.5m, and 2 m. In order to measure the size of the pixels displayed on the screen, it was measured by repeating 10 times in figure 4 and calculated by averaging them.

Actual distance	d = 10cm	d = 20cm	d = 30cm
100cm	91cm	120cm	98cm
150cm	132cm	135cm	159cm
200cm	138cm	245cm	220cm

**Table 1 Different measurement sizes for object at the same distance**

As shown in the table, the measured distances are not accurate. Measurements at short distances are relatively accurate, but measurements at long distances are inaccurate. This is because the accuracy of object detection is more accurate at short distances. If the value of d is larger than the short case, the

measurement distance is more accurate. This is considered to be because the relative error decreases when the value of P'-P in (6) is larger. From results of those experiments, distance measurements using object detection modules are not yet precisely accurate. Compared to the conventional methods of Radars and Lidars, they lack significantly in accuracy. One of the main reasons for this is that object detection technology cannot yet measure accurately the size of detected objects. Thus, when the value of P and P' in (6) is inaccurate, measurements are bound to be inaccurate as well. However, Artificial Intelligence is developing rapidly. Object detection technology still holds great potential, and in the near future, more accurate measurements using the technology will be made possible. Despite its inaccuracy, the results of this paper may be used in various areas. Because it detects objects real time, it can be used to warn people of objects approaching them, as well as a guidance system for the visually impaired. As mobile smart devices are becoming increasingly accessible, most people - including the visually impaired - will carry around at least one smart device at all times. The proposed system can be of great help if it is linked to such smart devices to continuously monitor and alert its user of any obstacles and its distance with the user. This paper is aimed partially at developing such assistance systems for the visually impaired. Accuracy is not the primary focus in such cases. All that the user needs are a rough measurement of distance that will allow him/her to move away from obstacles that are in the way.

#### V. CONCLUSION

This paper proposes a new method of distance measurement using cameras and Artificial Intelligence technology. It is generally very easy to calculate the distance from a camera to an object, if the actual size of the object and the size of its reflection on the camera is given. Based on such facts, we have developed a system where photos of an object are taken from two separate locations, and object detection - a sort of Artificial

Intelligence technology - is used to measure the distance between it and the camera. Several equations were induced mathematically to ensure the accuracy of the proposed method. The iPhone 8 plus was used to ensure that the process was real-time. Google tensor flow-lite object detection model was used. Several experiments in various environments were conducted. Empirical data had illustrated that such distance measurements using object detection modules are not yet precise and accurate. One of the main reasons for this is that object detection technology cannot yet measure accurately the size of detected objects. This is an issue to be resolved relatively easily with the advancement of Artificial Intelligence in the near future. The proposed method and developed system can be used to aid the visually impaired. If the system is incorporated in guidance systems for the visually impaired, it could effectively alert its user of obstacles that are in the way. The system may not provide a precise and accurate measurement of distance, but the information it provides will be more than enough for the user to avoid colliding into obstacles. The results of this paper will be used as a piece in developing guidance systems for the visually impaired.

## VI. REFERENCES

- [1]. SHIN HC, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*. 2016;35(5): 1285-1298. DOI: 10.1109/TMI.2016.2528162.
- [2]. Heng S, Minghao X, Ran L. Deep learning for household load forecasting—A novel pooling deep RNN. *IEEE Transactions on Smart Grid*. 2017;9(5): 5271-5280. DOI: 10.1109/TSG.2017.2686012.
- [3]. Shaoqing R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*. 2015: 91-99.
- [4]. Ross G. Fast r-cnn. *Proceedings of the IEEE international conference on computer vision*. 2015: 1440-1448.
- [5]. Jeff T. *Intelligent Mobile Projects with TensorFlow: Build 10+ Artificial Intelligence Apps Using TensorFlow Mobile and Lite for IOS, Android, and Raspberry Pi*. Packt Publishing Ltd, 2018.
- [6]. Joseph R, et al. You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016: 779-788.
- [7]. Rachel H, Jonathan P, Cuixian C. YOLO-LITE: a real-time object detection algorithm optimized for non-GPU computers. *IEEE International Conference on Big Data*. 2018: 2503-2510.
- [8]. Masahiro K, et al. Axi-Vision Camera (real-time distance-mapping camera). *Applied Optics*. 2000;39(22): 3931-3939. DOI:10.1364/AO.39.003931.
- [9]. Abir RK, et al. Person to camera distance measurement based on eye-distance. *Third International Conference on Multimedia and Ubiquitous Engineering, IEEE*. 2009: 137-141. DOI: 10.1109/MUE.2009.34.
- [10]. Arturo F, et al. Camera distance from face images. *International Symposium on Visual Computing*. Springer, Berlin, Heidelberg. 2013: 513-522.
- [11]. Zaarane A, et al. Distance measurement system for autonomous vehicles using stereo camera. *Array*. 2020 Mar; 5; 100016. DOI:10.1016/j.array.2020.100016.
- [12]. Robert J W, Xiang L, Charles XL. Pelee. A real-time object detection system on mobile devices. *Advances in Neural Information Processing Systems*. 2018: 1963-1972.
- [13]. Yan W, et al. Pseudo-lidar from visual depth estimation: Bridging the gap in 3d object detection for autonomous driving. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019: 8445-8453.

- [14]. Satapathy, S.K., Mishra, S., Mallick, P.K. et al. ADASYN and ABC-optimized RBF convergence network for classification of electroencephalograph signal. *Pers Ubiquit Comput* (2021). <https://doi.org/10.1007/s00779-021-01533-4>
- [15]. Bisoy, S. K., Mallick, P. K., & Mishra, A. Fairness Analysis of TCP Variants in Asymmetric Network. *International Journal of Engineering & Technology*, 7(2.12), 231-233.

**Cite this article as :**

K Usha Rani, Puram Shravya, Kotam Sreeja Reddy, "Object Detection Distance Estimation", *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, ISSN : 2456-3307, Volume 9, Issue 4, pp.94-101, July-August-2023.