INTERNATIONAL SCIENTIFIC AND VOCATIONAL JOURNAL (ISVOS JOURNAL)

Vol.: 8 Issue: 1 Date: 30.06.2024 Received: 20.06.2024 Accepted: 30.06.2024 Final Version: 30.06.2024

ISVOS Journal, 2024, 8(1): 87-93 – DOI: 10.47897/bilmes.1502873

Performing Distance Measurements Of Fixed Objects Detected With Yolo Using Web Camera

Volkan Karaca [a,1], Ebubekir Yaşar [b]

[a] Tokat Gaziosmanpaşa University, Tokat, Turkey
ORCID ID: 0009-0007-5002-7227
[b] Mechanical Engineering, Tokat Gaziosmanpaşa University, Tokat, Turkey
ORCID ID: 0000-0002-0780-893x

Abstract

Today, cameras are used for many purposes such as image analysis and synthesis. Technological developments have also made the use of image processing techniques more widespread and distance measurements with cameras have become more precise. In addition, examinations show that the cameras are also used to determine precise target distance or depth mapping studies. In this study, using the Yolo v8 model with a single web camera, the doors in a closed area whose positions did not change were detected, and then the distance of the detected object to the camera was tried to be measured. A comparison was made between the actual distances of the camera to the detected object and the measured distances. For distance measurement, measurement was carried out with the Euclidean method using OpenCV libraries. The study was carried out to show how effectively web cameras can be used at short distances and how much deviation occurs in the measurements. It is thought that measurement errors will be minimized using a webcam in future studies.

Keywords: "Deep Learning, Distance Measurement, Camera."

1. Introduction

The image processing process includes processes such as determining various features and details of an image, making changes to the image, and converting the image to other formats. Image processing is applied in many fields such as computer vision, medical imaging, satellite imaging and photography[1-2]. An image is a valuable data source that contains a lot of information. This information is obtained from the color, shape, size and many other characteristics of the image. Factors such as camera resolution, image quality, and sharpness of details on the image determine the amount and quality of information contained in an image. Therefore, image analysis and processing play a critical role in many applications and research fields[3-4]. Many researchers accept that distance information can be extracted by using some information in the image [5-6]. One of the frequently used and preferred methods in distance measurement systems with cameras is the field variation method. This method focuses on how the image of the target object taken at specific time intervals changes over time. In particular, this method examines the change in how much area the target object's image perceived by the camera covers over time. This provides important information about how the object's distance from the camera and therefore the camera's field of view changes. This information is used to measure and analyze distance more accurately. Therefore, the field variation method is considered a very valuable and effective method for mono camera distance measurement systems [7-8]. On the other hand, parallel to the studies in this field, there are also distance measurement systems using deep learning models. These models make it possible to make more accurate and precise distance measurements with information obtained from camera fields of view, often using complex algorithms and large data sets. Such systems generally have higher accuracy rates compared to traditional distance measurement methods and are generally preferred in more complex and challenging application areas[9-10].

In camera distance measurement systems, the use of equipment or the need for extra information about the target object and the varying measurement sensitivity depending on distance are among the important problems faced by these systems. Providing such information may require both time and additional resources, which may negatively impact the overall efficiency and speed of operations[11-12]. However, the application of pixel-oriented studies to such systems may cause measurement sensitivity to vary depending on distance. This can cause sensitivity to drop significantly and seriously affect the accuracy of the results, especially when measuring at longer distances. For this reason, the difficulty of obtaining accurate results at long distances can be a significant obstacle that pixel-oriented studies may encounter in the use of these systems [13-14]. Pixel-based methods are a

[1] Corresponding Author
E-mail Address: volkan.karaca@gop.edu.tr

widely used technique, especially in image processing and analysis systems. In these methods, the displacement of an image's pixel within the screen is used for various calculations and evaluations. Camera Field of View (FoV), which changes depending on distance, plays an important role in these calculations. FoV determines the width and height of the field that a camera can image. In fixed focus cameras, the resolution remains constant. This affects Pixels Per Meter (PPM). PPM refers to the number of pixels at a given distance in an image and is often used to determine the level of detail of an image. Therefore, pixel-based methods and the systems in which they are implemented play an important role in both image analysis and general image processing processes [15].

In our study, the doors in a closed area (corridor) whose positions did not change were detected using the Yolo v8 model with a single web camera, without using any additional equipment, and then the distance of the detected object to the camera was tried to be measured. The most important reason for using the Yolo v8 model is that it has a faster decision-making structure and produces more successful results compared to other models. A comparison was made between the actual distances of the camera to the detected object and the measured distances. For distance measurement, measurement was carried out with the Euclidean method using OpenCV libraries.

2. Material and Method

2.1. Yolo (You Only Look Once)

Object Detection(Object Detection); It is a computer technology related to computer vision and image processing that focuses on detecting images in photos, videos, and real-time images. Yolo is a cleaner that separates object using convolutional neural networks (CNN). It is the first object detection model to combine bounding box (bounding box) prediction and object signatures into a single end-to-end differentiable network. In other words, the objects in the images or videos and these distribution coordinates are detected simultaneously. “You Look Once”, meaning “Just Look Once”. The reason why this name was chosen is that it is fast enough to detect objects in one go[16].

The only difference between video and image processing is that images consist of a single frame, while videos consist of many frames. While the algorithm works for a single frame in images, it runs repeatedly for all frames in videos. The YOLO algorithm first divides the image into regions. It then draws the bounding boxes surrounding the objects in each region and calculates the probability of finding objects in each region[17].

It also calculates a trust score for each bounding box. This score tells us the percent probability that that object is the predicted object. For example, if the confidence score for a found car is 0.3, this means that the probability of that object being a car is very low. It applies a technique called non-maximum suppression to the objects inside the bounding boxes. This technique removes objects with a low confidence score from evaluation and checks for the presence of a bounding box with a higher confidence score in the same region.

It is searched for objects in each region. If an object is found, the midpoint, height and width of that object are found and then a bounding box is drawn. In order to do this, a number of sub-operations must be performed. A prediction vector is created for each region, and these vectors include the confidence score. If the confidence score is 0, there is no object there, if it is 1, it means there is an object there. More than one bounding box can be drawn for the same object within the same box. To get rid of this problem, the non-maximum suppression technique that I mentioned before is used. What is done with this technique is simply to keep the bounding box with the highest confidence score and remove the others from the image.

2.2. Preparing the Data Set and Creating the Data Label Map

While preparing the data set, 900 photographs were taken in the school corridor. The photo was taken with a mobile phone camera with a resolution of 2280x1080 pixels and a focal length of 26 mm. A few of the photographs taken are given in Fig. 1. These photographs were reproduced with various filters and 3600 photographs were obtained. The resolutions of the reproduced photos have been reduced to the resolution of the webcam to be used (640x480 pixels). Thus, it is aimed to eliminate errors that may arise from the resolution difference between the camera used for shooting and the webcam used to measure distance.

LabelIMG 1.8 program was used while preparing the label map. A label map of the objects identified in the photograph has been created and some of the marked objects are shown in Fig. 2 below.

Fig. 1. Corridor photos

Fig. 2. LabelImage program interface and feature mapped objects

The ".txt" extension maps of some of the objects marked in the corridor with the LabelIMG program are shown in Fig. 3.

```
69 0.292058 0.366667 0.087555 0.210000
65 0.085679 0.359167 0.043777 0.141667
20 0.032833 0.346667 0.063164 0.076667
70 0.196998 0.318333 0.077548 0.083333
71 0.201376 0.422500 0.082552 0.065000
23 0.202001 0.385000 0.067542 0.030000
54 0.479987 0.406667 0.124453 0.300000
56 0.545341 0.262083 0.036273 0.039167
57 0.666979 0.250000 0.028143 0.040000
55 0.796123 0.431667 0.266417 0.446667
```

Fig. 3. Label map of several objects

2.3. Measuring Distance with Webcam

In the study, measurements were made by taking into account the distance of a known object and the area it occupies on the camera. A frame is created for a referenced image. While creating the frame, the frame size is given according to the size of the detected object. An attempt was made to determine the unknown distance by considering the known frame size and known distance. Distance measurement with WebCam was tried to be prepared using the OpenCV library. Since the camera viewing angle is fixed, the viewing angle of the reference image and the detected image will not change, so distance determination was tried to be made from the same angle. Distance calculations are shown in Fig. 4 and Fig. 5. Mathematical calculations are given in equations 1, 2, 3, 4, 5, 6, respectively.

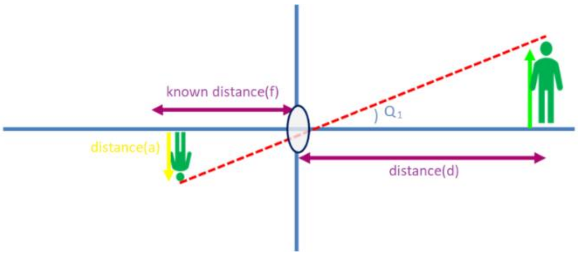

Fig. 4. Distance Calculations-1

TanQ1 is calculated in the equation (1) given below.

$$\frac{a}{f} = \tan Q1 = \frac{h}{d} \tag{1}$$

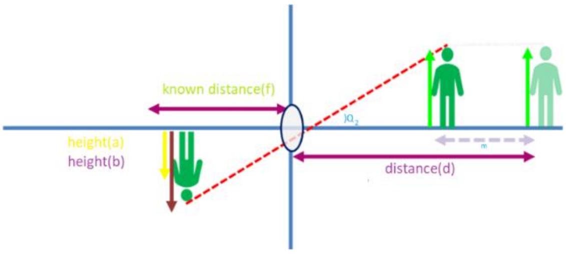

Fig. 5. Distance Calculations-2

TanQ2 is calculated in the equation (2) given below.

$$\frac{d}{f} = \tan Q2 = \frac{h}{d-m} \tag{2}$$

From the two equations given above, the following equations (3,4,5,6) arise, respectively.

$$\frac{a}{b} = \frac{h}{d} * \frac{d-m}{h} \tag{3}$$

$$\frac{a}{b} = \frac{d-m}{d} = 1 - \frac{m}{d} \tag{4}$$

$$\frac{m}{d} = 1 - \frac{a}{b} \tag{5}$$

$$d = \frac{m}{1 - \frac{a}{b}} \tag{6}$$

2.4. Object Detection and Distance Measurement

Below are shown a few distance measurements of the detected object in Fig. 6. An attempt was made to measure the distances of the detected object at a fixed frame width. Distance measurements were performed with a CMOS webcam with

350K color support and 640x480 pixel resolution. While measurements were carried out, 15x15cm landmarks were placed on the doors. Measurements were made using these landmarks as reference.

Fig. 6. Object Detection and Distance measurements

3. Findings and Conclusion

The measurement results made with the camera placed at a real distance between 80cm-20cm at the same height and at frame refresh rates of 1000ms-1ms are given in Table 1.

When the measured distances are evaluated according to frame rates;

• Among all measured frame refresh rates, the error rate is highest at 80cm and 20cm distances,

• Average error of 4.79% at 60-40-20cm at 1000ms frame refresh rate,

• Average error of 6.71% at 500ms frame refresh rate, 60-40-20cm,

• Except for the frame refresh rate of 1000ms and 500ms at a distance of 20cm, it shows the same distance (23.01cm) in all frames with a constant error rate (15.05%)

• Average error of 8.98% at 250ms frame refresh rate, 60 and 40cm,

• Average error of 4.73% at 100ms frame refresh rate, 60 and 40cm,

• Average error of 4.57% at 50ms frame refresh rate, 60 and 40cm,

• Average error of 6.47% at 25ms frame refresh rate, 60 and 40cm,

• Average error of 4.96% at 10ms frame refresh rate, 60 and 40cm,

• Average error of 3.74% at 5ms frame refresh rate, 60 and 40cm,

• At 1ms frame refresh rate, an average error of 2.51% was detected at 60 and 40cm.

When the measured distances were examined, it was seen that the ideal distance measurements were made from 40cm, which has the least error rate. At the same time, the error rate was measured to be minimum at 1 and 5ms frame rates. In addition, when the actual distance is greater than 80 cm, deviations increase further and more effective measurement can be performed at short distances.

As a result, it is thought that this study will facilitate applicability by making more accurate choices with the results obtained according to the ideal frame refresh rate in studies and projects that require distance measurement in real-time and bright environments.

Table 1. Actual distance, measured distance and error rates with web camera at various frame refresh rates

Frame Refresh Rate	Actual Distance(cm)	Measured Value(cm)	Error(cm)	Error(%)
1000ms	80	93,53	13,53	16,91
	60	62,97	2,97	4,95
	40	41,81	1,81	4,52
	20	20,98	0,98	4,9
500ms	80	92,17	12,17	15,21
	60	63,98	3,98	6,63
	40	42,12	2,12	5,3
	20	21,64	1,64	8,2
250ms	80	92,17	12,17	15,21
	60	67,14	7,14	11,9
	40	42,43	2,43	6,07
	20	23,01	3,01	15,05
100ms	80	93,44	13,44	16,8
	60	62,97	2,97	4,95
	40	41,81	1,81	4,52
	20	23,01	3,01	15,05
50ms	80	95,28	15,28	19,1
	60	63,98	3,98	6,63
	40	41,01	1,01	2,52
	20	23,01	3,01	15,05
25ms	80	92,91	12,91	16,13
	60	66,24	6,24	10,4
	40	41,02	1,02	2,55
	20	23,01	3,01	15,05
10ms	80	99,67	19,67	24,58
	60	65,91	5,91	9,85
	40	40,03	0,03	00,75
	20	23,01	3,01	15,05
5ms	80	99,07	19,07	23,83
	60	62,97	2,97	7,42
	40	40,03	0,03	0,075
	20	23,01	3,01	15,05
1ms	80	92,17	12,17	15,21
	60	62,97	2,97	4,95
	40	40,03	0,03	0,075
	20	23,01	3,01	15,05

References

[1] A. Vyas, S. Yu, and J. Paik, "Fundamentals of digital image processing," in *Multiscale Transforms with Application to Image Processing*, pp. 3-11, Singapore, 2018.

[2] E. Avuçlu, "Examining The Effect Of Pre-processed Covid-19 Images On Classification Performance Using Deep Learning Methods," *International scientific and vocational studies journal*, vol. 7, no. 2, pp. 94–102, Dec. 2023, doi: 10.47897/bilmes.1359954.

[3] Ç. Kurnaz and B. K. Gül, "Determination of the relationship between sodium ring width on iris and cholesterol level," *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, vol. 2018, no. 2018, Apr. 2018, doi: 10.17341/gazimmfd.416452.

[4] O. Ülkir, İ. Ertuğrul, and N. Akkuş, “Characterization of electrothermal micro-actuator using image processing algorithm,” *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, vol. 36, no. 2, pp. 1159–1170, Mar. 2021, doi: 10.17341/gazimmfd.683386.

[5] M. Takatsuka, G. A. W. West, S. Venkatesh, and T. M. Caelli, “Low-cost interactive active range finder,” *Machine Vision and Applications*, vol. 14, no. 3, pp. 139–144, Jul. 2003, doi: 10.1007/s00138-003-0129-y.

[6] M. F. A. Hassan, A. Hussain, M. H. M. Saad, and K. Win, “3D distance measurement accuracy on low-cost stereo camera, *Science International*,” vol. 29, no. 3, pp. 599-605, 2017.

[7] J. H. Oh, J. Park, S. H. Lee, B. H. Lee, and J. I. Park, “Error modeling of depth measurement using FIR stereo camera systems,” in *The Third International Conference on Digital Information Processing and Communications (ICDIPC2013)*, 2013.

[8] M. A. Mahammed, A. I. Melhum, and F. A. Kochery, “Object distance measurement by stereo vision,” *International Journal of Science and Applied Information Technology (IJSAIT)*, vol. 2, no. 2, pp. 05-08, 2013.

[9] H. Tsung-Shiang and W. Ta-Chung, “An Improvement Stereo Vision Images Processing for Object Distance Measurement,” *International Journal of Automation and Smart Technology*, vol. 5, no. 2, pp. 85–90, Jun. 2015, doi: 10.5875/ausmt.v5i2.460.

[10] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, “Depth from combining defocus and correspondence using light-field cameras,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 673-680, 2013.

[11] M. Soyaslan, “Object distance detection through disparity map in stereo camera system”, *Sakarya University Journal of Science (SAUJS)*, vol. 20, no. 2, pp. 111–119, 2016, doi: 10.16984/saufenbilder.77095.

[12] R. Srijha, “Methodology for distance measurement : A Comparative Study,” *International Journal of Electronics, Electrical and Computational System IJEECS*, vol. 6, no. 8, pp. 451–457, 2017.

[13] K. Y. Chen, P. K. Tsung, P. C. Lin, H. J. Yang, and L. G. Chen, “Hybrid motion/depth-oriented inpainting for virtual view synthesis in multiview applications,” in 2*010 3DTV-Conference: The True VisionCapture, Transmission and Display of 3D Video IEEE.*, pp. 1-4, 2010.

[14] A. Rahman, A. Salam, M. Islam, and P. Sarker, “An Image Based Approach to Compute Object Distance,” *International Journal of Computational Intelligence Systems*, vol. 1, no. 4, pp. 304–312, Dec. 2008, doi: 10.1080/18756891.2008.9727627.

[15] O. Uslu and E. Özkan, “IP CCTV sistemlerinde pixel (PPM) hesaplaması ve doğru çözünürlük tespiti.” Accessed: May. 20, 2024. [Online]. Available: https://www.emo.org.tr/ekler/a48e0098ca1a33f_ek.pdf

[16] “Gerçek Zamanlı Nesne Takibi(Real-Time Object Detection w/Yolo-Python),” *Smartera*, Dec. 23, 2020. https://smartera.com.tr/gercek-zamanli-nesne-takibireal-time-object-detection-w-yolo-python/ (accessed Jun. 01, 2024).

[17] D. Çakmak, “YOLO (You Look Only Once) Algoritması,” *Medium*, Oct. 14, 2020. https://deryacakmak.medium.com/yolo-you-look-only-once-algoritmas%C4%B1-9e8819c5e4cf (accessed Jun. 01, 2024).