ISSN 2278-3091
Volume 9, No.4, July – August 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse357942020.pdf
https://doi.org/10.30534/ijatcse/2020/357942020

Proposed real-time obstacle detection system for visually impaired assistance based on deep learning

Fatima ZahraeAitHamou Aadi[1],Abdelalim Sadiq[2]
[1]Ibn Tofail University, Morocco,fatimazahrae.aithamouaadi@uit.ac.ma
[2]Ibn Tofail University, Morocco, a.sadiq@uit.ac.ma

ABSTRACT

The mobility in an unfamiliar environment independently is one of the biggest challenges for visually impaired and blind people. It becomes difficult for those with this kind of disability to keep a track of their routine environments, and so many of them tend to rely on their sighted friend or family member for assistance. In this paper, the authors are proposing a real time based-vision system using object detection, object tracking and object-camera distance estimation for the visually impaired people assistance. The main purpose is to come up with a low cost system which can detect and track the surrounding obstacles and estimate their risk on the users, in order to facilitate theirmobilityin any kind environment. Firstly, the You Only Look Once approach is the main object detector, by applying YOLO the obstacles will be detected in a frame t and their bounding boxes will be tracked in the coming frames. Secondly, DisNet which is a deep learning based method is applied for the distance estimation.

Key words: Deep learning, DisNet, Mobility, Object Detection, Visually impaired assistance, YOLO

1. INTRODUCTION

According to the World Health Organization (WHO), there are currently 2.2 billion people in the world living with visual impairment; 13\% of them are legally blind while the other 87\% have some form of significant visual disease [1]. Crossing the street: knowing whether security is assured, and being sure to stay on the road, not colliding or interfering with objects and people, are the type of difficulties which keep this category away from leading an independent normal life in the society. Therefore, as researchers, we are concerned by providing an innovative solution that facilitates their mobility in unfamiliar environments. Nowadays, all the technological improvements have made us able to produce powerful solution, like the Internet of Things and Big Data. In this paper, the authors propose an obstacle avoidance system for the blind disabled community, using a smart glasses Furthermore, as part of approaching the human vision of the environment, this research aims to integrate the computer vision context in these glasses, in order to understand the user's environment.

The main idea of this approach is creating an assisting system capable of helping visually impaired people to navigate and succeed all their daily activities independently. To do so, the authors are planning to create an obstacle detection and tracking system able to identify all surrounding risks based on the DisNet [2]approach for the distance estimation, the glasses will support a monocular camera. While the blind person navigation glasses will ensure the collection of information about the surrounding environment. This will enable it to give the users the necessary instructions to succeed during their travels. Depending on the extracted information analysis, the system will not only be able to identify obstacles, but more than that, will carry other useful capabilities to facilitate the user's journey, such as facial recognition and reading instructions, etc. During the trip, the glasses will collect all the meaningful information about the environment so the system can build a ground truth model. This concept is based on how blind people build an imaginary map for all the places that they have visited.
This paper is organized as follow: section 1 introduce the motivation of the paper. Section 2 and 3 present the object detection and tracking chosen methods as well as the distance estimator used in the proposed system which is well explained in section 4. Finally, section 5 conclude this paper.

2. OBJECT DETECTION AND TRACKING

The key problem in computer vision is object detection. Over the past few years many methods have been proposed based on the deep learning concept, not only under the guise of detecting object in images but also for the position and classes estimation. Those methods are mainly divided into two categories the first one present all methods that can estimate the position as well as the class in one stage like YOLO [3] and SSD [4]. The second category gather all those methods like Region based convolutional neural network (RCNN) [5] and its enhanced versions. This categoryrequires a primary stage for the object region proposal in the image and a CNN based classifier in the second stage.
Answering the question of which detector is best may be more difficult. Because a particular algorithm for object recognition cannot be oversimplified so far, to give the best results in all scenarios. Otherwise it is necessary to know which configurations of detectors give the best balance of speed and accuracy required. In the case proposed by the authors and

according to the evaluation of different methods which been described in [6], the YOLO approach has been selected as the main object detector within the proposed system, seeing that fits their configurations.

2.1 You Only Look Once

YOLO [3],[7] is an object detection algorithm based on convolutional neural network (CNN), which examines the whole image, once for all process. The algorithm applies a single neural network to the full image, and then divides the image into regions and predicts bounding boxes and probabilities for each region as shown in figure 1. Unlike RCNN [5], [7] and other algorithms, which needed a multiple parts, starting by region proposals generation and then run an image classifier on them. Therefore, the bounding boxes prediction and the class probabilities calculation uses just one neural network. On the down side, YOLO makes localization errors compared to other systems [3], [7].

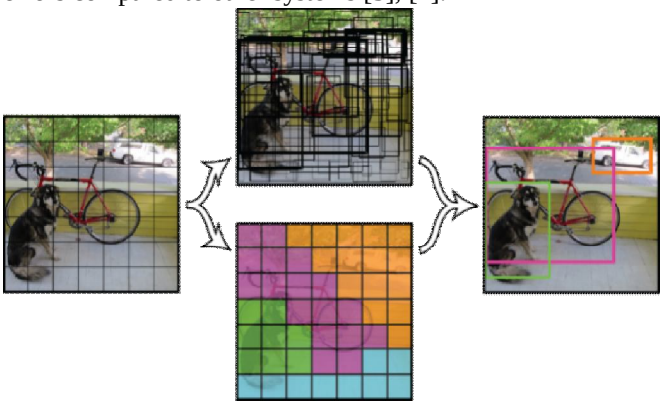

Figure 1: You Only Look Once Process [3]

In the case of real-time object detection system the need of a detector with a high accuracy is highly demanded and also the adaptability with low-cost hardware is important also, in this regard YOLO authors claim a new state-of-the-art accuracy with a highly maintained frame rate processing. The fourth version achieves an accuracy of 43.5% AP (65.7% AP□□) for the MS COCO with an approximately 65 FPS inference speedon Tesla V100 [8].

2.2 Object Tracker

Object tracking: is the process of finding a moving object in a stream video, by determining if the object is the same as the one presented in the previous frame. Among the tracking categories there is point based tracking as a first category, where the object is represented by a feature point in order to be tracked. Some of the methods following this concept we recognize Kalman filter [9] and particle Filtering [10]. The second category gather under the Kernel based tracking name, performed by computing an object in motion, which is denoted as a region of a primitive object, from one frame to the next. Object movement is usually in the form of parametric motion, such as transformation. There are a variety of tracking methodologies present based on this Kernel tracking approach such as Simple Template Matching [11], Mean Shift Method [12] and Support Vector Machine (SVM) [13].

The Shape/Silhouette based tracking present the last category, its main goal is to track complex shape objects, which contain other shapes like human body (hands, fingers, etc). Those techniques identify the moving object state in each frame, by creating an object model from the preliminary frame [11], [14], and [15].

The authors' system needed multiple object tracking method that is robust, fast and optimizes the processing. Seen that, the methods of point trackers need to be detected in each frame. Whereas, kernel based tracking and Shape/Silhouette based tracking require detection of an object in its first appears in the scene. Accordingly, they applied in their system a kernel based tracking technique named the Kernelized Correlation Filters tracker (KFC). This method is able to detect as many existing objects as possible in real-time, without the need to apply the detector each frame.

3. DISTANCE ESTIMATION

The distance estimation is an important step under the pretense of identifying the obstacle risk during the user trip. As known, there are numerous sensors for that mission as Laser, Infrared ray, LIDAR, etc.Considering the fact that the authors are planning to come up with a low cost system for the visually impaired community, it seems reasonable to use the camera itself as the distance estimator sensor instead of adding any of the mentioned sensors, there is no doubt that it gives more accurate results, but at the same time it will make the system more complex and sure expensive.

The authors tested the use of an active triangulation method for distance estimation which emits a signal and then measures the reflected signals [16]. Using this method caused not much delay, but it counts as long as the system is reliable to the real-time processing.

This paper use in this regard a neural network-based object distance estimation from camera method called DisNet. This method was originally proposed by M. A. Haseb and al. in [2]. DisNet is a supervised learning technique was trained on 2000 different objects extracted from a video camera record using YOLO detector. The bounding boxes of those objects as well as their real distance from the camera have been feeded as an inputs to DisNet, in order to train the main model how to estimate the distance between an object and a camera, as shown in figure 2. Furthermore, the real distance related to the 2000 box have been measured using a laser sensor within the same scene recorded.

The vectors given to the model are composed as follows:

$$v = [1/B_h 1/B_w 1/B_d C_h C_w C_d] \quad (1)$$

Where:

B_h : Height of the bounding box / image height (pixels)

B_w : Width of the bounding box / image width (pixels)

B_d : Diagonal of the bounding box / image diagonal (pixels)

C_h, C_w, C_d : Values of average height, width and breadth of an object for a particular class.

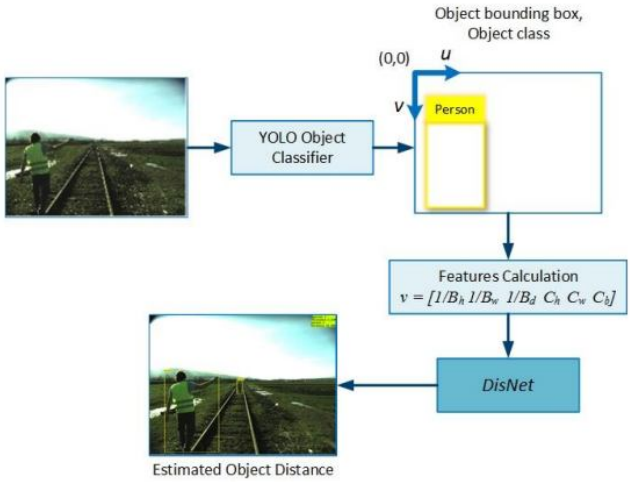

Figure 2: The DisNet-based system used for object distance estimation from a monocular camera [2]

4. PROPOSED SYSTEM ARCHITECTURE

Admittedly, the human brain has a remarkable ability to perceive and understand a real-world scene, going through a very complex process at the level of neurons. This makes simulation a very difficult and complicated. Although, computers exceed most of the big data, human visual system emulation is still primitive and can only capture basic information from visual images such existing objects. Mainly the object detection is the key step in this system. As known, there are numerous sensors for that mission as Laser, radar, etc. [18], [19]. But using a camera it will serve the author's idea of having a low cost system for the visually impaired community.

The main goal is to take this emulation to a next level and pursuing a whole learning process, which groups together the object detection and tracking as well as the risk classification related to all objects around the user, in order to give this assisting system the ability to :

- Identify and distinguish between obstacles in front of the user's body from the ground to the heads.
- Provide information about the obstacle-user distance with essential direction instructions based on a risk classification.
- Afford instructions to the user about surroundings.
- Give the user an ability to create a mental map of the environment to succeed self-orientation.

According to some measure, this blind assisting technology and using a wearing glasses. As shown in figure 3, the system integrated within glasses has different stages starting with a frame collecting stage, passing through the object detection phase. Detected objects are tracked in the next frames. An object is identified by its position, type and its distanceto the userestimated using the DisNet technique. All these information are used to classify the risk related to this object.

For the proposed system the distance between the user and the obstacle is very important stage it become after detecting an object and knowing its direction. Below are the steps shown in figure 3:

1. Get frames in real-time from given source (smart glasses, camera).
2. Reading this dataset in our program.
3. Detecting the objects present in the images using YOLO.
4. Compare the extracted objects with the known dataset
5. Once the object is recognized its Location at the current frame is extracted.
6. The object will be tracked, while computing his risk on the user in the next frame tell it disappear.
7. Repeat 3 to 6 tell the end of the navigation

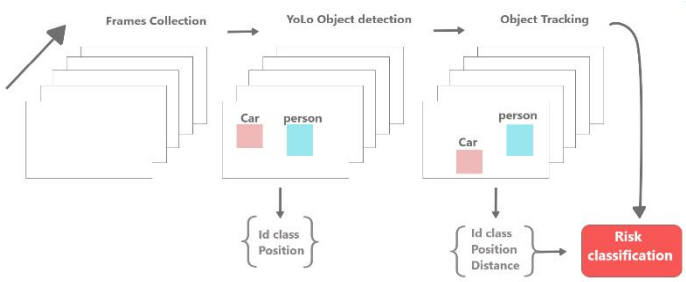

Figure 3: Blind Assisting System architecture within smart glasses

The system only begins to generate warnings of vulnerable objects in the proximities when the distance between the user and the obstacle is less than the thresholding value equal to 0.5 (thresholding-risk). If the distancebetween the object and the person is less than thresholding-risk, then the object is considered as a high risk obstacle, otherwise is classified as a negligible. Meanwhile the user would be informed just by the surrounding obstacle with high risks and their types, consequently the focus on the global environment where the user navigates would be reduced.

The use of YOLO based object detection in the proposed system is reliable in spite of the fact that YOLO classifier was used in its original form trained with COCO dataset [17], without retraining with the images from the test field. The result of object detection and distance estimation in a dynamic environment (moving car and moving object-obstacle) show that achieving distance estimation is not satisfactory in spite of the fact that objects not fully bounded with the bounding box and also changed within the stream video due to the tracking method, however this part needed an improvement as well as the risk estimation in the case of colliding boxes.

5. CONCLUSION

The integration of various solutions in a single assistive system is a huge challenge. This paper present a contribution on this regard by setting out a system architecture based on deep learning techniques that allows blind people or people with low vision to detect and avoid obstacles.

This system is mainly divided into three steps:

- The first step is detecting the surrounding objects using YOLO which gives a high accuracy and speed performing the real time applications
- Second step is applying KFC to track the detected objects.
- Third step reserved to the object distance estimation using DisNet model, this latter was trained on 2000 object bounding boxes and their real distance to the camera measured by a laser

The distance estimation results were mainly used to identify the obstacles risk on the user. The evaluation of this system was been evaluated on a real scene video record, as a future work the authors will test the feasibility of this system on the real field. Furthermore the object mention tracking step still needed an enhancement, in this regard,novel features will be used to improve the accuracy of object tracking and distance calculation for risk estimation and better understanding of the environment.

REFERENCES

1. World Health Organization. "Blindness and vision impairment". 2019.
2. M. A. Haseeb, J. Guan, D. Ristić-Durrant and A. Gräser. **DisNet: A novel method for distance estimation from monocular camera.** *10th Planning, Perception and Navigation for Intelligent Vehicles (PPNIV18)*, IROS, 2018.
3. J. Redmon, S. Divvala, R. Girshick and A. Farhadi. **You Only Look Once: Unified, Real-Time Object Detection**, 21-26 July 2017. https://doi.org/10.1109/CVPR.2016.91
4. W. Liu et al.**SSD: Single Shot MultiBox Detector**, *ECCV*, pp21-37, 2016.
5. S. Ren, K. He, R. Girshick and J. Sun. **Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks**, *Advances in Neural Information Processing Systems 29th Annual Conference on Neural Information Processing Systems*, vol. 28, 2015.
6. M. L Vanishree, S. Sushmitha and B.K. Roopa,**Addressing the challenges of Visually Impaired using IoT**,*In International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC)*, 2017.
7. F. Z AitHAmouAadi, A. Sadiq and A. Souhar. **Introduction to the environment understanding for the visually impaired assistance: Object detection and image segmentation**, *In Third International Conference on Intelligent Computing in Data Sciences (ICDS). IEEE, Marrakech, Morocco*, 26 December 2019.
8. A. Bochkovskiy, C. Wang, H. M. Liao.**YOLOv4: Optimal Speed and Accuracy of Object Detection.** *Computer Vision and Pattern Recognition*,2020.
9. R. E. Kalman. **A New Approach to Linear Filtering and Prediction Problems**,*Journal of Basic Engineering*,pp 35-45, 1960. https://doi.org/10.1115/1.3662552
10. C. Shen, A. Van Den Hengel and A. Dick. **Probabilistic Multiple Cue Integration for Particle Filter Based Tracking**. *In 7th Australian Pattern Recognition Society Conference*. pp. 399-408, 2003.
11. A.Yilmaz, O. Javed and M. Shah. **Object tracking: A survey.***ACM computing surveys (CSUR)*, vol. 4, 2006.
12. D. Comaniciu, V. Ramesh and P. Meer. **Real-Time Tracking of Non-Rigid Objects Using Mean Shift**, *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp 142-149, 2000.
13. G. Ramya and Mrs. Srilatha. **Multiple Object Tracking using Support Vector Machine**, *In Global Journal of Researches in Engineering: J General Engineering*, Vol. 6, pp 34-38, 2014.
14. A. M. Jacob and J. Anitha. **Inspection of various object tracking techniques**. *International Journal of Engineering and Innovative Technology*, vol. 6, pp 118–124, 2012.
15. R.C. Veltkamp. **Shape matching: similarity measures and algorithms**, *IEEE International Conference on Shape Modelling and Applications*,pp 188–197, 2001.
16. B. Drayton. **Algorithm and design improvements for indirect time of flight range imaging cameras**,*Victoria University of Wellington, NZ*, 2013.
17. T. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar and C. L. Zitnick. **Microsoft COCO: Common Objects in Context**, *In ECCV*, 2015.
18. H. Khudov, S. Yarosh, V. Savran, A. Zvonko, A. Shcherba5, P. Arkushenko. **The Technique of Research on the Development of Radar Methods of Small Air Objects Detection**. *International Journal of Emerging Trends in Engineering Research*. Volume 8(7), pp. 3708-3715, July 2020. DOI: https://doi.org/10.30534/ijeter/2020/132872020
19. H. Khudov, I. Khizhnyak, V. Koval, V. Maliuha, A. Zvonko, V. Yunda, V. Nagachevskyi, and V. Berezanskyi. **The Efficiency Estimation Method of Joint Search and Detection of Objects for Surveillance Technical Systems**, *International Journal of Emerging Trends in Engineering Research*, Vol. 8. № 3, 2020, pp. 813–819. https://doi.org/10.30534/ijeter/2020/34832020