JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

REALTIME OBJECT DETECTION USING OPENCV

Pokala Nithya Sai- Student, Department of CSE – Data Science, Geethanjali College of Engineering and Technology

Muthyala Prashanth- Student, Department of CSE – Data Science, Geethanjali College of Engineering and Technology

Ganta Sathvik- Student, Department of CSE – Data Science, Geethanjali College of Engineering and Technology

Mr. G Vijay Kumar- Sr. Asst Professor, Department of CSE – Data Science, Geethanjali College of Engineering and Technology

Abstract

This article summarizes the background and application fields of Open-Source Computer Vision Library (OpenCV) and deep learning and conducts research based on their object detection and tracking applications. The model algorithm selected in this article is the Convolutional Neural Networks (CNN) algorithm. CNN algorithm can be used for object detection in real-time scenarios. Moreover, the CNN algorithm has good credibility. The Python program developed based on CNN algorithm can effectively achieve real-time object tracking. The model shows good detection and tracking performance for trained targets and can be further applied in more specific scenarios in the future. Because deep learning can process large-scale data and recognize complex patterns, it can automatically learn and extract advanced features. Combining the two can achieve faster and more accurate detection and tracking of target objects. After having a large enough training sample size, it can detect and track the specified object more accurately. However, it is precisely due to the enormous sample size required for deep learning that there are still some difficulties in applying it to real-time object tracking. This article first discusses the use of supervised learning methods for deep understanding. When the input sample capacity is large enough, the machine can better reflect the real-time detection and tracking of the target object. But the time required will significantly increase. This issue still needs to be addressed in the subsequent research process.

1. INTRODUCTION

Real-time object tracking is a critical task in computer vision, which has attracted wide attention in recent years. It involves continuous monitoring and positioning of objects of interest in the scene to achieve applications such as video surveillance, autonomous navigation, and human-computer interaction [1]. Traditional object tracking methods often rely on hand-designed features and explicit models, making it difficult to handle complex scenes, occlusion, and changes in the appearance of objects. To overcome these limitations, advances in deep learning techniques in recent years have revolutionized the field of computer vision. Deep learning, as a subfield of machine learning, has shown remarkable ability in handling large-scale data and complex pattern recognition tasks. It utilizes multilayer neural network models inspired by the structure and function of the human brain to learn and extract advanced features from raw data automatically.

Deep learning models, especially CNN, excel at object detection and classification tasks. By training these models on large-scale data sets, they learn to recognize and understand objects' inherent features and representations [2]. This ability to automatically learn discriminant features makes deep learning an ideal method for real-time object tracking. At the same time, OpenCV became a widely adopted open-source framework for computer vision tasks. OpenCV offers a comprehensive set of image processing and computer vision algorithms that researchers and developers favor. Its modular architecture, cross-platform compatibility, and extensive documentation make it ideal for academic research and industrial applications. OpenCV offers many capabilities, including image acquisition, preprocessing, feature extraction, and visual analysis. It has efficient real-time processing ability and can be used to develop a real-time object-tracking system. In addition, OpenCV seamlessly integrates with deep learning frameworks such as TensorFlow and PyTorch, enabling researchers to fully leverage the power of deep learning within the OpenCV ecosystem [3]. Combining deep learning and OpenCV has great potential to advance the field of real-time object tracking. Using deep learning models trained on large-scale data sets, object tracking systems can learn to generalize and adapt to various object appearances, changes, and complex environments. OpenCV is supporting in this regard, providing a robust real-time video processing framework and implementing feature extraction and tracking algorithms [4].

This paper will delve into deep learning and OpenCV and the integration of the two in real-time object tracking. It will explore methods and techniques for training deep learning models for object tracking, integrating deep learning frameworks with OpenCV, and implementing efficient tracking algorithms. In addition, it will discuss the advantages, challenges, and potential applications of this combined approach, demonstrating the potential for more accurate, robust, and efficient real-time object-tracking systems.

1.2 Scope of the Project:

The project aims to develop a comprehensive real-time object tracking system by integrating deep learning techniques with the OpenCV framework. Through thorough research and analysis, the project will identify requirements, design a suitable system architecture, and gather relevant datasets for training deep learning models. The trained models will be seamlessly integrated into OpenCV, enabling efficient object detection and tracking in real-time scenarios. Extensive testing and validation will ensure the system's accuracy, reliability, and performance across various environments. Documentation will accompany the system, providing users and developers with comprehensive guidance. The project will not only address the limitations of traditional object tracking methods but also offer potential extensions and enhancements to further improve its functionality and applicability in domains such as video surveillance, autonomous navigation, and human-computer interaction.

2. LITERATURE SURVEY

The literature review highlights several key research studies focusing on various aspects of object tracking, deep learning, and their integration with the OpenCV framework. Mangawati et al. [1] concentrate on object tracking algorithms tailored for video surveillance applications, while Biswas et al. [2] explore object classification within video records using neural network frameworks. Noble [3] compares feature detection and matching algorithms in OpenCV, providing insights into their performance and suitability for different computer vision tasks.

Furthermore, U et al. [4] delve into real-time object distance and dimension measurement using deep learning and OpenCV, potentially for applications like industrial automation or augmented reality. Rajesh et al. [5] investigate Quantum Convolutional Neural Networks (QCNN) for enhancing performance in object recognition and classification tasks. Additionally, Pang et al. [6] propose a combined approach of deep learning and preference learning for improved object tracking adaptability and accuracy in dynamic environments.

Farhodov et al. [7] explore the integration of Faster RCNN detection with OpenCV CSRT tracker using drone data to enhance tracking performance in aerial surveillance. Choudhary et al. [8] focus on edge detection of cowpea leaves using OpenCV and deep learning for agricultural monitoring and disease detection. Wang et al. [9] propose an optimization architecture for image processing based on deep learning and big data processing technology, aiming to enhance workflow efficiency.

Tushar et al. [10] investigate object detection using OpenCV and deep learning, providing insights into accurate and efficient object detection capabilities for various applications. Sharma et al. [11] present a practical implementation of object detection using OpenCV and Python, demonstrating the practical application of computer vision techniques. Additionally, foundational works by O'Shea and Nash [12] provide an introduction to Convolutional Neural Networks (CNNs), laying the groundwork for understanding advanced deep learning techniques.

Moreover, Satpute et al. [13] discuss real-time object detection using deep learning and OpenCV, emphasizing the integration of deep learning techniques with the OpenCV framework. Hearst et al. [14] investigate support vector machines (SVMs) for classification tasks, providing an overview of their principles and applications in various domains. Lastly, Chandan et al. [15] focus on real-time object detection and tracking using deep learning and OpenCV, aiming to enhance surveillance and monitoring systems.

3. OVERVIEW OF THE SYSTEM

3.1 Existing System

Technology is increasing day by day, development in object recognition technology includes autonomous vehicles etc. is huge. There is an existing system on object detection are already there but it is not reaching the visually impaired who need it most. Those systems which already existed may be object recognition or voice guidance and that are not much useful for blind. In our system we have both object detection and after that that information is converted into voice that will be useful for blind people.

3.1.1 Disadvantages of Existing System

Limited Accessibility: Existing object detection systems often fail to reach visually impaired individuals who need them the most. This limitation stems from the lack of integration with accessibility features such as voice guidance, making them less useful for the target user group.

Lack of Comprehensive Solutions: Many existing systems focus solely on object recognition or voice guidance individually, rather than providing a combined solution. This fragmentation results in a disjointed user experience and may not adequately address the diverse needs of visually impaired users.

Challenges in Facial Feature Extraction: Extracting facial features to a significant extent remains a challenging task for existing systems. This limitation impacts the accuracy of facial recognition technologies, which are essential for certain applications such as identifying individuals or emotions.

Inability for Voice Recognition: Some existing systems may lack the capability for voice recognition, limiting their interaction with visually impaired users who rely on voice commands or feedback for navigation and operation.

3.2 Proposed System

The Proposed system is to analyze the surroundings for object detection and voice guidance to help the blind people by using various machine learning and deep learning algorithms. By using these we can implement an efficient object-detection system that helps the blind find objects in a specific space without help from others. This system successfully detects objects, labels them and also shows its accuracy. The model also calculates the distance from the object to the camera and gives voice feedback as when the person with the camera is approaching the object.

3.2.1 Advantages of Proposed System

Comprehensive Solution: The proposed system integrates object detection with voice guidance, providing a comprehensive solution for visually impaired individuals. This combination addresses the limitations of existing systems by offering both functionalities in a single platform.

Utilization of Machine Learning and Deep Learning: Leveraging various machine learning and deep learning algorithms enhances the system's ability to accurately detect objects in real-time. This approach allows for continuous improvement and adaptation to different environments.

*Independence for Blind Individuals: B*y enabling blind individuals to detect objects in their surroundings without assistance, the proposed system promotes independence and autonomy in daily activities. This empowerment contributes to their overall quality of life.

Accuracy and Precision: The system not only detects objects but also labels them and calculates the distance from the camera. This information is crucial for blind users to understand their surroundings effectively. The voice feedback provided ensures accuracy and precision in object recognition.

User-Friendly Interface: The system is designed to be user-friendly, with voice feedback providing intuitive guidance to users. This interface reduces the learning curve and makes the technology accessible to a wide range of visually impaired individuals.

Real-Time Feedback: By providing real-time feedback on object detection and distance calculation, the system enhances situational awareness for blind users. This feature is particularly valuable in navigating outdoor environments and during emergencies.

Scalability and Adaptability: The proposed system can be adapted to work with different devices and platforms, allowing for scalability and widespread adoption. This flexibility ensures that the technology can cater to the diverse needs of visually impaired individuals globally.

3.3 Proposed System Design

In this project work, there are five modules and each module has specific functions, they are:

1. Dataset Pre-Processing

2. Object Detection

3. Feature Extraction

4. Classification

5. Voice Generation

3.3.1 Dataset pre-processing

Data pre-processing in machine learning is the process of preparing the raw data to make it ready for model making. It is the first and the most crucial step in any machine learning model process. We have COCO Dataset which consists of more than 200k images for data pre-processing. We had training dataset, public test dataset (which is then used as validation dataset for our project), and further a private test dataset (same size with public test dataset and will be used as data for evaluating the prediction performance).

3.3.2 Object detection

Object detection is the process of finding instances of objects in images. In the case of deep learning, object detection is a subset of object recognition, where the object is not only identified but also located in an image. This allows for multiple objects to be identified and located within the same image.

3.3.3 Classification

Depth estimation or extraction feature is nothing but the techniques and algorithms which aims to obtain a representation of the spatial structure of a scene. In simpler words, it is used to calculate the distance between two objects. Our prototype is used to assist the blind people which aims to issue warning to the blind people about the hurdles coming on their way. In order to do this, we need to find that at how much distance the obstacle and person are located in any real time situation. After the object is detected, rectangular box is generated around that object.

3.3.3 Classification

The dimensionality of data obtained from the feature extraction method is very high so it is reduced using classification. Features should take different values for object belonging to different class so classification in Yolo will be done as a single regression problem.

3.3.3 Voice Generation

After the detection of an object, it is utmost important to acknowledge the person about the presence of that object on his/her way. For the voice generation module GTTS plays an important role. GTTS is a conversion library in Python which converts text into speech. This library works well with both Python 2 and 3. GTTS is a tool which converts text to speech easily. Audio commands are generated as output. If the object is too close then it states "Warning: The object (class of object) is very close to you. Stay alert!". Else if the object is at a safer distance, then then a voice is generated which says that "The object is at safer distance". This is achieved with the help of certain libraries like GTTS, pyttsx3, pytesseract and engine.io.

3.4 Architecture

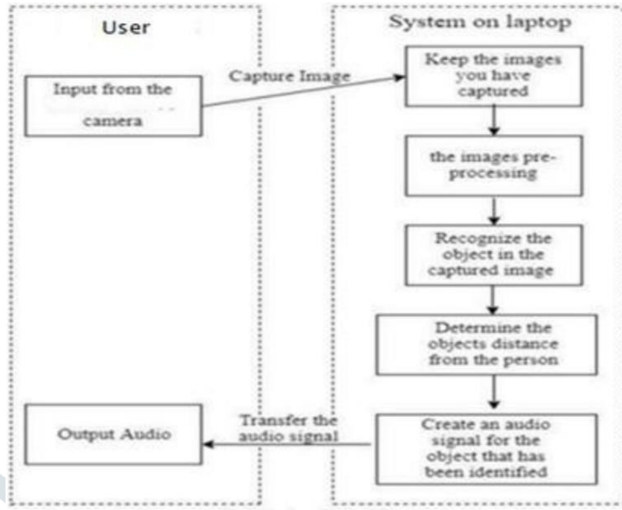

Fig 1: System Architecture

4. RESULT SCREEN SHOTS

5. CONCLUSION

We successfully developed a model for Object Detection using OpenCV (Open Computer Vision) Algorithm which detects the objects more accurately and converts detected information into Voice Alerts with GTTS (Google Text to Speech). Several technologies have been created to aid visually impaired persons. One such attempt is that we would wish to make an Integrated Machine Learning System that allows the blind victims to identify and classify real-time objects generating voice feedback and distance. Which also produces warnings whether they are very close or far away from the thing. For visually blind folks, this technology gives voice direction. This technique has been introduced specifically to assist blind individuals. The precision, on the other hand, can be improved. Furthermore, the current system is based on the Android operating system; it can be altered to work with any device that is convenient. It works better than other existing techniques for visually challenged people who need it the most in their everyday life. This technology gave us the capacity to develop these projects that will be of real benefit to the individual in need. Our system will help the visually impaired to navigate in the outdoor environments independently by detecting the object and the exact location of it with necessary alert systems during navigation as well as during emergencies. Since the objects in the outdoor environment could be related to any subject, the dataset must cover sufficient ground so that the model is trained appropriately. It can provide a low-cost system for many in need.

6. REFERENCES

[1] A. Mangawati, Mohana, M. Leesan and H. V. R. Aradhya, "Object Tracking Algorithms for Video Surveillance Applications," 2018 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2018, pp. 0667-0671, doi: 10.1109/ICCSP.2018.8524260.

[2] A. Biswas, A. P. Jana, Mohana and S. Sai Tejas, "Classification of Objects in Video Records using Neural Network Framework," 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2018, pp. 564-569, doi: 10.1109/ICSSIT.2018.8748560.

[3] F. K. Noble, "Comparison of OpenCV's feature detectors and feature matchers," 2016 23rd International Conference on Mechatronics and Machine Vision in Practice (M2VIP), Nanjing, China, 2016, pp. 1-6, doi: 10.1109/M2VIP.2016.7827292.

[4] B. M U, H. Raghuram and Mohana, "Real Time Object Distance and Dimension Measurement using Deep Learning and OpenCV," 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 2023, pp. 929-932, doi: 10.1109/ICAIS56108.2023.10073888.

[5] V. Rajesh, U. P. Naik and Mohana, "Quantum Convolutional Neural Networks (QCNN) Using Deep Learning for Computer Vision Applications," 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), Bangalore, India, 2021, pp. 728-734, doi: 10.1109/RTEICT52294.2021.9574030.

[6] Pang, S., del Coz, J.J., Yu, Z. et al, "Deep Learning and Preference Learning for Object Tracking: A Combined Approach," Neural Process Lett 47, pp.859-876, 2018, doi: 10.1007/s11063-017-9720-5.

[7] X. Farhodov, O. -H. Kwon, K. W. Kang, S. -H. Lee and K. -R. Kwon, "Faster RCNN Detection Based OpenCV CSRT Tracker Using Drone Data," 2019 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, 2019, pp. 1-3, doi: 10.1109/ICISCT47635.2019.9012043.

[8] V. Choudhary, P. Guha, K. Tripathi and S. Mishra, "Edge Detection of Variety of Cowpea Leaves Using OpenCV and Deep Learning," 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022, pp. 1312-1316, doi: 10.1109/ICAC3N56670.2022.10074348.

[9] W. Wang, J. Wang, Z. Zhang and D. Shi, "OpenCV Implementation of Image Processing Optimization Architecture of Deep Learning Algorithm based on Big Data Processing Technology," 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2022, pp. 65-68, doi: 10.1109/ICSCDS53736.2022.9760795.

[10] Tushar, K. Kumar and S. Kumar, "Object Detection using OpenCV and Deep Learning," 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022, pp. 899-902, doi: 10.1109/ICAC3N56670.2022.10074012.

[11] A. Sharma, J. Pathak, M. Prakash, and J. N. Singh, 'Object Detection using OpenCV and Python', in 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India: IEEE, Dec. 2021, pp. 501–505. doi: 10.1109/ICAC3N53548.2021.9725638.

[12] K. O'Shea and R. Nash, 'An Introduction to Convolutional Neural Networks'. arXiv, Dec. 02, 2015. Accessed: Jun. 25, 2023. [Online]. Available: http://arxiv.org/abs/