IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁙ Peer-reviewed journal & Refereed journal ⁙ Vol. 12, Issue 11, November 2023

DOI: 10.17148/IJARCCE.2023.121103

A Comprehensive Survey on the Current State of the Art Technologies used for Live Environment Assistance

Prof. M. P. Shinde[1], Shreyash Dhurupe[2], Viraj Karanjavane[3], Sanna Shaikh[4], Abhishek Suryawanshi[5]

Department of Computer Engineering, SKNCOE, Savitribai Phule Pune University, Pune 411041, India[1]

Undergraduate Research Paper, Department of Computer Engineering, SKNCOE, Savitribai Phule Pune University, Pune 411041, India[2,3,4,5]

Abstract: A thorough examination of the most recent cutting-edge technologies utilized to support live environments is provided in this survey report. The main technologies covered are Text-to-Speech (TTS) systems for auditory feedback; Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks for object detection and live scene description; real-time processing techniques for continuous feedback; and Optical Character Recognition (OCR) for live text recognition. The significance of user-friendly interfaces in improving user experience and the relevance of different datasets in training these algorithms are also covered in the study. The purpose of this survey is to present a thorough summary of the body of literature, address the benefits and drawbacks of the approaches used now, and make recommendations for possible future research avenues. This work is an important tool for scholars who want to explore and improve existing technologies in live environment assistance systems.

Keywords: Object Detection, Optical Character Recognition, Convolutional Neural Network, Text-to-Speech, Computer Vision, Neural Networks.

I. INTRODUCTION

Live environment support systems have advanced significantly since the invention of technology. These systems combine a number of technologies to offer a user-friendly interface. The purpose of this study is to provide an overview of these technologies, their uses, and possible future advancements.

Technologies such as Text-to-Speech (TTS) systems for audio feedback and Optical Character Recognition (OCR) for live text recognition are used by live environment aid systems. For object detection and live scene description, deep learning models such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks are employed. Continuous feedback is ensured using real-time processing techniques, which enable the system to react quickly to environmental changes. One other important component of these systems is the user interface (UI).

A system that is easy to use and has an intuitive interface improves the user experience. This work reviews the literature in existence, emphasizing the specifics of contemporary approaches. A fair assessment of the industry is given by discussing the benefits and drawbacks of various technologies. The role of different datasets in training these algorithms is also covered in the paper. The study concludes by outlining a possible system for further investigation. It points out gaps in the existing literature and suggests solutions. For academics and developers, this article offers an extensive overview of the state of the technologies now employed in live environment assistance systems.

II. RELATED WORK

Karez Abdulwahhab Hamad and Mehmet Kaya provide a thorough examination of optical character recognition technology in their paper titled "A Detailed Analysis of Optical Character Recognition Technology." The process of converting any type of text, including handwritten text, printed text, and scanned text pictures, into an editable digital format for additional processing is known as optical character recognition, or OCR. Text in these kinds of papers can be automatically recognized by a machine thanks to optical character recognition technology. The accuracy of this technology is dependent upon the caliber of the input that is supplied, such as scanned documents and photos. [1]

In their paper "Object Detection Using Convolutional Neural Networks," Galvez, Reagan & Bandala, Argel & Dadios, Elmer & Vicerra, Ryan & Maningo, Jose Martin gave a thorough overview of CNN and included two state-of-the-art

DOI: 10.17148/IJARCCE.2023.121103

model investigations. According to the study, Faster-RCNN with InceptionV2 has lower speed but higher accuracy than SSD (Single Shot MultiBox Detector) with MobileNetV1, which has high speed detection but low accuracy. The trials' findings indicate that there is a trade-off between speed and precision. Use SSD with MobileNetV1 if you want quick detection, especially for real-time applications. To get a high level of accuracy in detection, utilize Faster-RCNN in conjunction with InceptionV2. [2]

In Venkateswarlu, S. & Duvvuri, Duvvuri B K Kamesh & Jammalamadaka, Sastry & Rani, Chintala's "Text to Speech Conversion," they used a Raspberry Pi to implement TTS (Text to Speech Synthesizer) and OCR (Optical Character Recognition). This publication provided a thorough introduction to system design and methodology. According to the publication's conclusion, a TTS device can convert text to audio with a readability tolerance of less than 2% and high performance. [3]

The learning-based method for scene labeling was provided by W. Byeon, T. M. Breuel, F. Raue, and M. Liwicki in their publication titled "Scene labeling with LSTM recurrent neural networks". The LSTM network learns the neighboring context information about each pixel and internally models the global dependencies between labels using recurrent connections. A study on layers in 2D LSTM recurrent neural networks for scene labeling was also presented. They provided the system architecture and conductuted some experiments that aid in the information gathering process for scene labeling. [4]

Ahmad, Tanvir & Ma, Yinglong & Yahya, Muhammad & Ahmad, Belal & Nazir, Shah & Haq, Amin & Ali, Rahman proposed YOLOv1 neural network-based object detection in "Object Detection through Modified YOLO Neural Network" by altering the loss function and adding a spatial pyramid pooling layer and inception module with convolution kernels of 1 ∗1. The experiment they conducted demonstrated the efficacy of the newly enhanced network, and the publication includes a network analysis and methodology. [5]

Othman, Nashwan & Salur, Mehmet & Karakose, Mehmet & Aydin, Ilhan's paper "An Embedded Real-Time Object Detection and Measurement of its Size" offers an effective real-time object measurement technique that is suggested for industrial systems. The provided system uses computer vision to measure and identify objects. In a real-time video, the system is able to identify and quantify objects. Using OpenCV functions, the size of each object is obtained after it has been detected using the canny edge detector. They improved the algorithm for the canny edge detector by applying morphological operations. The process has the advantage of removing unnecessary sounds. Moreover, it smooths the shape and maintains the dimensions and outline of each object while also removing the unnecessary sounds. As a result, the various objects' outlines in the scene were preserved. Five frames can be processed in a second using the suggested technique, which operates incredibly quickly. Since the Raspberry Pi 3 has great features and is an affordable embedded equipment platform, it is used to implement the systems. [6]

“Real-Time Object Detection and Tracking from a webcam with OpenCV in Python” provides us methodology, system architecture, experiment on Object detection and Tracking from a webcam. OpenCV is used with python and the study also provides insights about computer vision. [7]

“A Joint Convolutional Neural Networks and Context Transfer for Street Scenes Labeling” paper proposes a joint framework of priori s-CNNs and soft restricted context transfer for street scenes labeling. The priori s-CNNs can fully exploit priori information through preserving super pixels location in the image. With the proposed framework, the labeling accuracy of the foreground objects increases significantly. Nevertheless, the missing and false labeling phenomena are common in our results. Thus, we will focus on integrating objects detector into our model to enhance the labeling accuracy in the future. [8]

“Blind and Visually Impaired User Interface to Solve Accessibility Problems” by Azeem Shera, Muhammad Waseem Iqbal, Syed Khuram Shahzad, Madeeha Gul, Natash Ali Mian, Muhammad Raza Naqvi and Babar Ayub Khan aimed to address interface accessibility challenges faced by blind and visually impaired (BVI) users when using mobile applications. By identifying and addressing organizational, presentation, and behavioural (OPB) issues based on the users' mental model, the study developed and evaluated a smartphone application specifically designed for BVI users. The research involved 56 BVI participants, who performed tasks related to organization, avoidance of redundant information, serialization of content, and style and text presentation. The results demonstrated 100% effectiveness in the organization of the application for both blind and visually impaired participants, with visually impaired users exhibiting better efficiency.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed journal & Refereed journal ⁜ Vol. 12, Issue 11, November 2023
DOI: 10.17148/IJARCCE.2023.121103

The study concluded that the developed application significantly reduced OPB problems, leading to higher usability, satisfaction, effectiveness, and efficiency for BVI users. The findings provide valuable design and development guidelines for creating accessible and user-friendly smartphone applications for the BVI community. [9]

"Smartphone-Based Cognitive Assistance of Blind People in Room Recognition and Awareness" study addresses the challenges faced by blind individuals in activities like localization, navigation, and social interaction, with limited focus on room environment judgment and communication with specific people. The proposed solution aims to offer cognitive assistance to blind individuals by predicting room types and locating intended communication partners based on ambient environment and social cues like age, gender, and the number of people present.

Utilizing a combination of microphone, speaker, and camera, the system conveys information through haptic feedback. Evaluation metrics, including movement, sounds, orientation, and position, were analysed, with an extensive user study achieving 87.64% accuracy in room type recognition, 64.57% in gender recognition, 61.73% in age group identification, and 61.71% in correctly identifying the number of people. The study highlights the importance of considering furniture and objects within a room and aims to address privacy concerns through non-image-based solutions for counting people. Future work aims to extend the system's applicability to controlled outdoor environments. [10]

This paper highlights the advancements in technology that have propelled the development of systems aimed at improving the daily lives of visually impaired individuals. The focus is on leveraging existing technologies, specifically Optical Character Recognition (OCR) and Text-to-Speech (TTS) capabilities on Android smartphones.

The system discussed in this paper utilizes these technologies to automatically identify and recognize texts and signs in the user's environment, offering critical information to facilitate navigation. Through a combination of computer vision and internet connectivity, the proposed system not only recognizes signs but also reconstructs sentences and converts them into speech.

The design flow and experimental results of the project are presented, emphasizing the implementation of a Text-to-Speech algorithm that enables visually impaired individuals to interpret signs in their surroundings. The experiments demonstrate the feasibility of the concept on an Android smartphone, using still images, with potential for future real-time implementation. [11]

III. PROPOSED SYSTEM

We suggested a methodology for an Android application that is intended to help people who are visually impaired based on the survey results. With voice commands facilitating smooth interaction, the application functions as a hands-free user interface, allowing users to confidently and easily navigate their surroundings.

The three unique modules that make up the system architecture of the suggested application are all designed to improve the user experience. The first module creates a detailed scene description by utilizing Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) techniques. This enhances the user's spatial awareness by giving them specific information about the area they are currently in.

The real-time object detection and obstacle alert system are the focus of the second module. By guaranteeing that users obtain prompt notifications regarding possible impediments in their way, this improves their mobility and safety. The real-time text and sign detection is the focus of the third module. It speaks detected text and signs to the user audibly by combining Text-to-Speech (TTS) conversion with Optical Character Recognition (OCR) technology.

This feature enables users to engage with their surroundings more successfully. When combined, these modules create an integrated system that uses cutting-edge technologies to provide a voice-command-operated, hands-free user interface. This novel strategy improves visually impaired people's mobility and ability to interact with their surroundings, greatly raising their quality of life. Thus, the suggested methodology serves as evidence of the potential of technology in creating inclusive solutions for individuals with visual impairments.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁘ Peer-reviewed journal & Refereed journal ⁘ Vol. 12, Issue 11, November 2023
DOI: 10.17148/IJARCCE.2023.121103

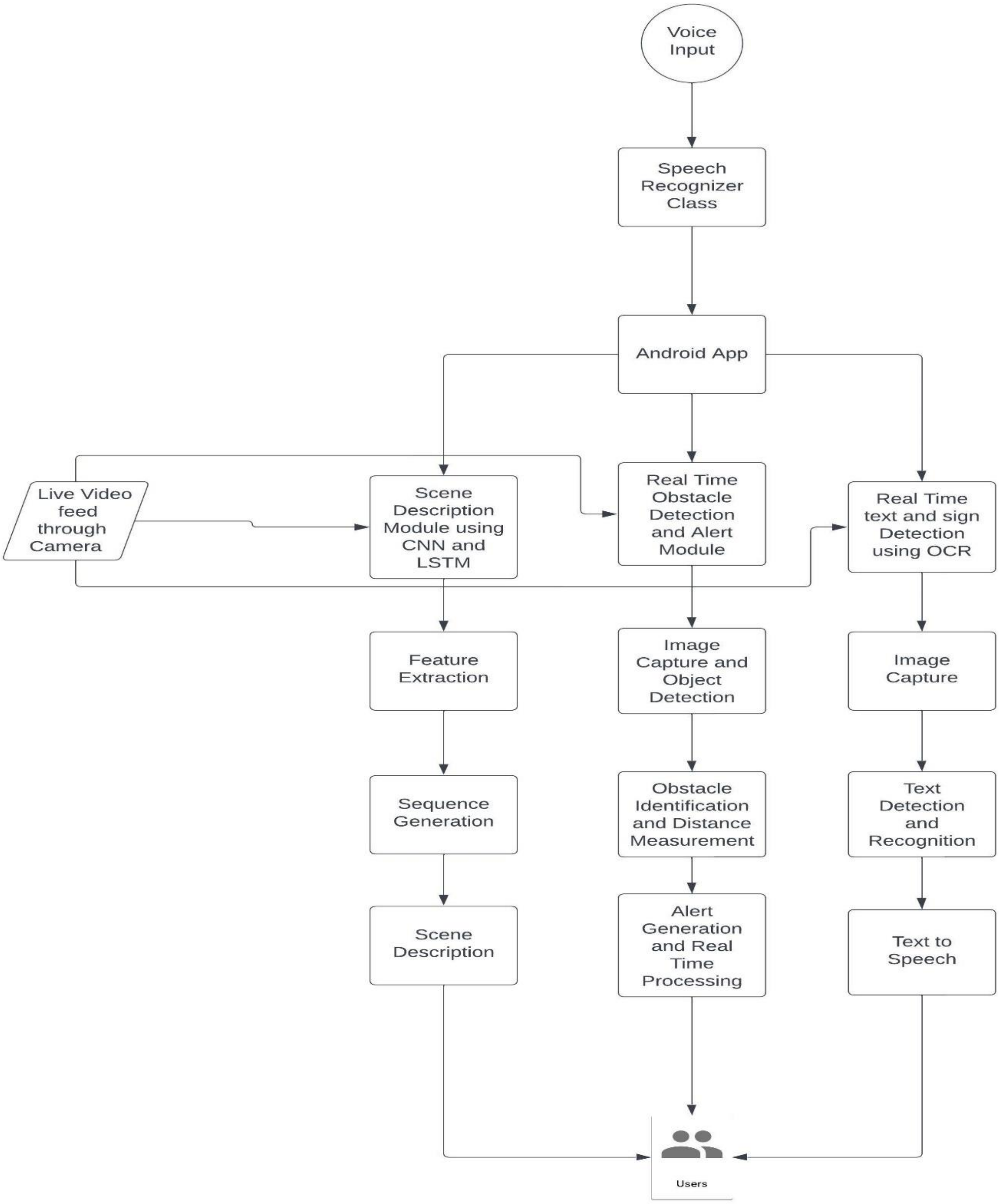

Fig. Proposed System

IV. CONCLUSION

An extensive summary of the most recent cutting-edge technologies utilized in live environment assistance systems has been given in this survey paper. We have investigated several technologies, including real-time processing methods, text-to-speech (TTS) systems, convolutional neural networks (CNNs), optical character recognition (OCR), and long short-term memory (LSTM) networks. Every single one of these technologies is essential for improving user experience and offering help in real time. We have emphasized the specifics of these technologies by carefully going over the body of existing literature. This fair assessment of the field enables us to pinpoint areas in need of development as well as possible lines of inquiry for further study. To improve the capabilities of live environment assistance systems, the paper emphasizes the significance of ongoing innovation in this field.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁚ Peer-reviewed journal & Refereed journal ⁚ Vol. 12, Issue 11, November 2023
DOI: 10.17148/IJARCCE.2023.121103

Live environment assistance systems are a rapidly developing field that frequently sees the emergence of cutting-edge technologies and applications. We should anticipate major developments in this field as long as researchers and developers keep coming up with new ideas. In the hopes that these technologies will continue to improve and enhance people's lives all over the world, this paper serves as a guide for those interested in conducting more research in this field.

The suggested methodology shows how technology can be used to develop inclusive solutions that can greatly enhance the lives of those who are blind or visually impaired. It opens the door for more developments in the field of live environment assistance systems and serves as a monument to the ability of creativity to solve problems in the real world. Thus, this suggested system serves as a promising model for future research and development in this area.

REFERENCES

[1] Hamad, Karez & Kaya, Mehmet. (2016). A Detailed Analysis of Optical Character Recognition Technology. International Journal of Applied Mathematics, Electronics and Computers. 4. 244-244. 10.18100/ijamec.270374.
[2] Galvez, Reagan & Bandala, Argel & Dadios, Elmer & Vicerra, Ryan & Maningo, Jose Martin. (2018). Object Detection Using Convolutional Neural Networks. 2023-2027. 10.1109/TENCON.2018.8650517.
[3] Venkateswarlu, S. & Duvvuri, Duvvuri B K Kamesh & Jammalamadaka, Sastry & Rani, Chintala. (2016). Text to Speech Conversion. Indian Journal of Science and Technology. 9. 10.17485/ijst/2016/v9i38/102967.
[4] W. Byeon, T. M. Breuel, F. Raue and M. Liwicki, Scene labeling with LSTM recurrent neural networks, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 3547-3555, doi: 10.1109/CVPR.2015.7298977.
[5] Ahmad, Tanvir & Ma, Yinglong & Yahya, Muhammad & Ahmad, Belal & Nazir, Shah & Haq, Amin & Ali, Rahman. (2020). Object Detection through Modified YOLO Neural Network. Scientific Programming. 10.1155/2020/8403262.
[6] Othman, Nashwan & Salur, Mehmet & Karakose, Mehmet & Aydin, Ilhan. (2018). An Embedded Real-Time Object Detection and Measurement of its Size. 10.1109/IDAP.2018.8620812.
[7] Shamim, Rejuwan & Agarwal, Trapty. (2023). Real-Time Object Detection and Tracking from a webcam with OpenCV in Python. Juni Khyat Journal. 13. 64.
[8] Wang, Qi & Gao, Junyu & Yuan, Yuan. (2017). A Joint Convolutional Neural Networks and Context Transfer for Street Scenes Labeling. IEEE Transactions on Intelligent Transportation Systems. PP. 1-14. 10.1109/TITS.2017.2726546.
[9] Azeem Shera, Muhammad Waseem Iqbal, Syed Khuram Shahzad, Madeeha Gul, Natash Ali Mian, Muhammad Raza Naqvi and Babar Ayub Khan. (2021). Blind and Visually Impaired User Interface to Solve Accessibility Problems. Intelligent Automation & Soft Computing
DOI:10.32604/iasc.2021.018009
[10] Muhib Ullah, Shah Khusro, Mumtaz Khan, Iftikhar Alam, Inayat Khan, and Badam Niazi.(2022). Smartphone-Based Cognitive Assistance of Blind People in Room Recognition and Awareness.
[11] Hao Jiang, Thomas Gonnot, Won-Jae Yi and Jafar Saniie. Computer Vision and Text Recognition for Assisting Visually Impaired People using Android Smartphone