JOURNAL OF INNOVATION AND TECHNOLOGY
eISSN:2805-5179 Vol. 2024, No.03

Augmented Reality Application for Optical Character Recognition

Nandan Kumar N[1], Wan Nor Al-Ashekin Binti Wan Husin[2*]

[1] Dayananda Sagar Academy of Technology and Management, Karnataka, India.
[2] Faculty of Data Science and Information Technology, INTI International University, 71800 Nilai, Negeri Sembilan, Malaysia.

[*]**Email:** wanashekin.whusin@newinti.edu.my

Abstract

Augmented Reality (AR) technology has become popular for improving user experiences by superimposing virtual features on the real world. OCR is another recent method for extracting text from photos or real-world items. AR and OCR are combined in a new software that provides an immersive and engaging experience. The proposed AR-based OCR system uses Firebase as a backend. Users can point their smartphones at papers, signs, or other textual material to use AR, which will automatically recognize and extract the content. This extracted content can be translated, converted to text-to-speech, or shared on social media. Storage and management of recognized text data is reliable and scalable with the Firebase database connector. The Firebase Realtime Database can immediately sync extracted text across several devices for user collaboration and sharing. Firebase Authentication can authenticate and authorize users for safe OCR access. The program uses image processing for text extraction, OCR models for accurate recognition, and AR frameworks like ARCore (Android) and ARKit (iOS). The application will be linked to the Firebase backend using SDKs and APIs for real-time data synchronization and safe data storage. The AR-based OCR application has great promise in education, logistics, retail, and other industries. It can extract text from physical documents, increase accessibility for visually challenged people, and translate foreign language text in real time. Firebase's backend database solution meets the application's needs for scalability, dependability, and data security.

Keywords

Augmented Reality, Optical Character Recognition, OCR models, Image Processing, Text Extraction

Introduction

Augmented Reality (AR) is a technology that merges the real and virtual worlds to improve the user's perception and interaction with the environment. OCR (Optical Character Recognition) is a technique that extracts text from pictures or real-world objects. By combining these two

Submission: 15 May 2024; **Acceptance:** 27 June 2024

technologies, sophisticated apps that can recognize and interact with text in the actual environment may be created.

Google Firebase is a platform for developing mobile and web applications. It offers a number of features, including a real-time database, authentication, hosting, and cloud functionality. The Firebase Realtime Database is a NoSQL database hosted in the cloud that allows developers to store and sync data in real time. An Augmented Reality application for Optical Character Recognition that uses Firebase Database may take advantage of both technologies' capabilities. The program may take photos or live video feeds from a mobile device's camera and process them using OCR algorithms to extract text from the acquired information. This extracted text can subsequently be utilized for a variety of tasks like translation, information retrieval, and interaction with other digital elements.

By offering a scalable and real-time data storage solution, Firebase Database may play an important part in this application. The recognized text may be saved in the Firebase Realtime Database, allowing for real-time synchronization between devices or users. This enables collaborative or shared experiences in which numerous users interact with the same text at the same time. The Firebase Database also has extensive querying and indexing features, this makes it easier to search for and retrieve text data stored in the database. This is important for creating features like text-based search and filtering furthermore; to manage user authentication and access control, Firebase Authentication may be linked into the application. This enables safe user-specific data storage as well as personalized experiences. In conclusion, an Augmented Reality application for Optical Character Recognition powered by Firebase Database can provide real-time text recognition and storage, collaborative experiences, and seamless synchronization across various devices. Developers may create new and dynamic apps that span the physical and digital worlds by harnessing the capabilities of both AR and Firebase.

Literature Review

OCR technology has been present for some decades and has grown in popularity in recent years. Computers can recognize printed or handwritten text and transform it to a digital version that can be modified or processed by software using OCR technology. Augmented reality (AR) technology, on the other hand, is a relatively new idea that has received a lot of interest in recent years because of its capacity to superimpose digital information on the actual environment. The purpose of this literature study is to investigate the possible benefits and drawbacks of adopting AR technology for OCR.

The researcher as (1) investigated the viability of employing AR technology for OCR. The scientists created an AR-based OCR system that allowed users to use a smartphone camera to capture a picture of a printed document, and the system would subsequently extract the text from the image and display it in an augmented reality environment. The study's findings demonstrated that the AR-based OCR system could accurately extract text from printed documents, and that the augmented reality environment provided an intuitive and immersive manner to display the recovered information. (2) has investigated the efficacy of an AR-based OCR system for recognizing handwritten text. The authors created an AR-based OCR system that recognized handwritten text and displayed it in an augmented reality environment using machine learning techniques. The study's findings revealed that the AR-based OCR system could recognize handwritten text with an accuracy of 91.6%, and the augmented reality environment provided a more natural and intuitive method to interact with the recognized text. However, adopting AR technology for OCR has several drawbacks. One key drawback is that AR technology needs specialized gear, such as a smartphone or tablet with a camera, which limits the technology's accessibility. Furthermore, lighting conditions, camera quality, and the quality of the printed or handwritten text being recognized can all impact the accuracy of AR-based OCR systems.

Finally, the use of augmented reality (AR) technology for OCR has the potential to give a more immersive and natural approach to extract and display text from printed or handwritten documents. While technology has significant limitations, such as the requirement for specialized hardware and possible accuracy difficulties, greater study in this field might lead to the creation of more effective and accessible AR-based OCR systems.

Problem Statement

The combination of augmented reality (AR) with optical character recognition (OCR) technology has the potential to revolutionize information retrieval and interaction with the physical world. However, certain problems must be overcome to effectively harness the advantages of AR OCR applications. The problem statement for this project is to identify and address the major problems in building and deploying effective AR applications for optical character recognition, such as issues relating to accuracy, real-time processing, user interface design, and privacy/security concerns. By solving these issues, our research intends to improve the usability, reliability, and practicality of AR OCR systems, allowing for their wider use across numerous businesses and areas.

Methodology

Implementation

Optical Character Recognition (OCR)

OCR is a technique that enables users to transform text or documents acquired in photos by an input device into an editable, searchable, and reusable data format for further image processing. This technique enables a machine to recognize letters automatically using an optical system, similar to how humans use their eyes to detect objects in the environment. Several issues arose during the early stages of implementing OCR, such as restrictions in the amount and complexity of the hardware and the algorithm (3). However, OCR has been widely employed in a variety of applications such as cheque processing, digital libraries, recognizing text in natural surroundings, comprehending hand-written office forms, and so on. With the progress of technology and the contributions of well-known firms like as IBM, HP, Microsoft, Google, and others via continued research, OCR has grown and become more and more sophisticated over time. An OCR system is made up of numerous subsystems, each of which is dedicated to solving a certain problem. difficulties and play various functions in image processing (Laine & Nevalainen, 2006). Although there are different algorithms available, many of them follow the essential processes outlined which include setup firebase database, setup the AR-OCR Application, implement of OCR technology, perform OCR on captured images, authenticate and connect to firebase, text segmentation, store and retrieve the OCR data in Firebase Database, enhance the AR-OCR Application and finally deployment

1.

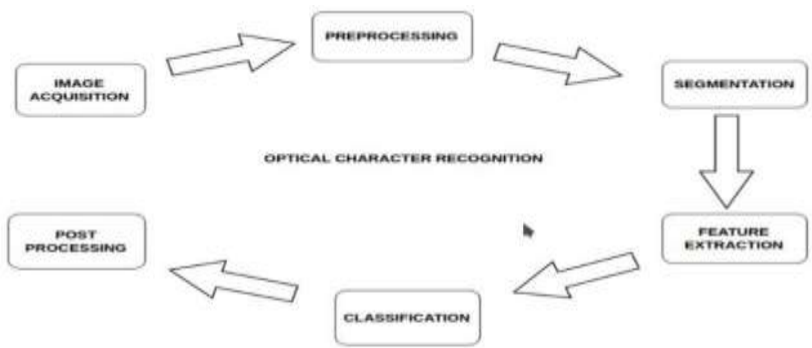

Figure 1. OCR Process

Related Works

According to Smith, (4), AROCR and AR-based OCR system that uses Firebase as its backend database. The authors cover the development process, including picture capture, text recognition, and interaction with Firebase. They assess the system's correctness and real-time processing performance. Lee S. et al., (2020) have suggested an AR-OCR system for mobile device real-time translation. To store and retrieve translation data, they use Firebase as the backend database.

On the other hand, Chen (5) introduced the AROCR-DB, an AR-OCR system that combines a Firebase cloud database. The authors go over the architecture of the system, concentrating on image processing, word recognition, and Firebase database interaction. They assess the system's performance and demonstrate its use in realistic settings. The integration of OCR with Firebase for cloud-based processing and storage is investigated in this study. The authors provide a complete system that integrates picture preprocessing, word recognition, and the Firebase database. They go through Firebase's benefits for OCR apps, including real-time data synchronization and offline support (6).

Finally, Li (7) offered AROCR+, an improved AR-OCR system that includes a cloud database powered by Firebase. They offer a comprehensive installation guide that covers picture capture, text recognition, and Firebase connection. The research assesses the system's performance and indicates its possible applications in a variety of fields.

These connected publications offer useful insights into the integration of AR, OCR, and the Firebase database, covering topics such as system design, implementation procedures, performance assessment, and prospective applications. Researchers and developers can draw inspiration and assistance from this research when developing their own AR-OCR apps with Firebase as the backend infrastructure.

Chosen Technology

1. OCR

There are various open source and commercial OCR engines on the market today, each with their own set of strengths and shortcomings. Engines like GOCR, Cuneiform, OCRAD, Tesseract, and OCROPUS are available from several open-source communities. Commercial OCR engines include ABBYY Fine Reader, OmniPage, and Microsoft Office Document Imaging. Tesseract is an open-source engine created by HP laboratories between 1985 and 1995 and subsequently given to Google Inc. in 2006. Tesseract paired with the Leptonica picture Processing library, which can read and transform a broad range of picture files to text in over 60 languages. It is compatible with all PC operating systems as well as the Android and iPhone mobile platforms. Because of the popularity of Tesseract as an open-source engine, many academic experiments and OCR software innovations have been completed successfully. According to research done by OCRAD, GOCR, and Tesseract, Tesseract outperforms other open-source engines. Despite the presence of dirty data, Tesseract proved to be the top free and open-source OCR engine in terms of accuracy and processing speed, as seen in Figure 2.

	cuneiform	gocr	ocrad	tesseract
License	BSD	GPL2	GPL3	Apache 2.0
Recognition rates and time spent:				
courier/black	61% (1.11s)	67% (0.09s)	21% (0.02s)	81% (0.63s)
courier/gray	✗	67% (0.09s)	21% (0.03s)	81% (0.63s)
justy/black	3% (1.14s)	31% (0.11s)	1% (0.02s)	15% (0.61s)
justy/gray	✗	31% (0.10s)	1% (0.02s)	15% (0.60s)
times/black	96% (1.07s)	76% (0.16s)	82% (0.03s)	92% (0.74s)
times/gray	✗	76% (0.16s)	82% (0.03s)	92% (0.74s)
verdana/black	95% (1.07s)	98% (0.10s)	98% (0.03s)	98% (0.45s)
verdana/gray	✗	98% (0.10s)	98% (0.02s)	98% (0.46s)

Figure 2. Comparison of Open-Source OCR Engine

2. Firebase

The Firebase Realtime Database is a JSON-formatted cloud-hosted database (8-10) Every client connected to the network receives real-time data synchronization. When we build cross-platform apps using our iOS and JavaScript SDKs, all our clients share a single Realtime Database instance and receive instant data changes. The Firebase Realtime Database is a NoSQL database that enables us to store and sync data across our users in real time. It's a massive JSON object that developers can work with in real time. Through a single API, the Firebase database provides the application with the current value of the data as well as changes to that data. Because of real-time synchronization, our customers may access their data from any device, whether online or mobile, as seen in Table 1.

Table 1. Comparison between Firebase and SQL

Basis of comparison	**Firebase**	**SQL(RDBMS)**
Data Storage	Stored as JSON Tree	Stored in a Relational Model as Rows and Columns (Tables)
Schema flexibility	Dynamic Schema, data can be added, updated or deleted anytime	Fixed schema. Altering will result in going offline temporarily
Specialty	Data which has no definite type or Structure	Data whose type is known in advance
Technique	Synchronize data	Fire Query

3. System Design

A. Architectural Design

This section describes the architecture for an augmented reality (AR) application with optical character recognition (OCR) capabilities, using the Firebase database as the backend solution. The architectural design addresses the structural components, data flow, and interaction between the application's many levels. It describes the most important architectural patterns, such as client-server communication, data storage, and synchronization. This design is to serve as a model for developers looking to build an AR-OCR application with Firebase as the backend infrastructure.

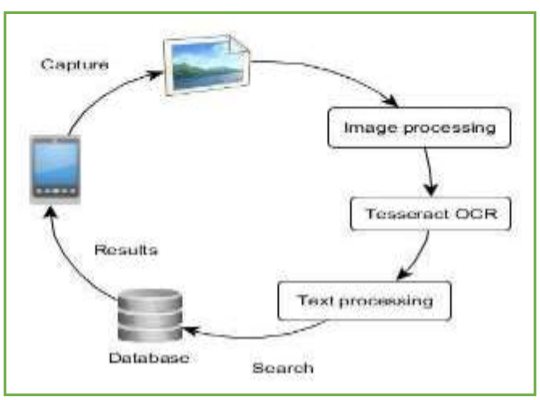

Figure 3. AR-OCR Application with Firebase as The Backend Infrastructure

B. Flowchart

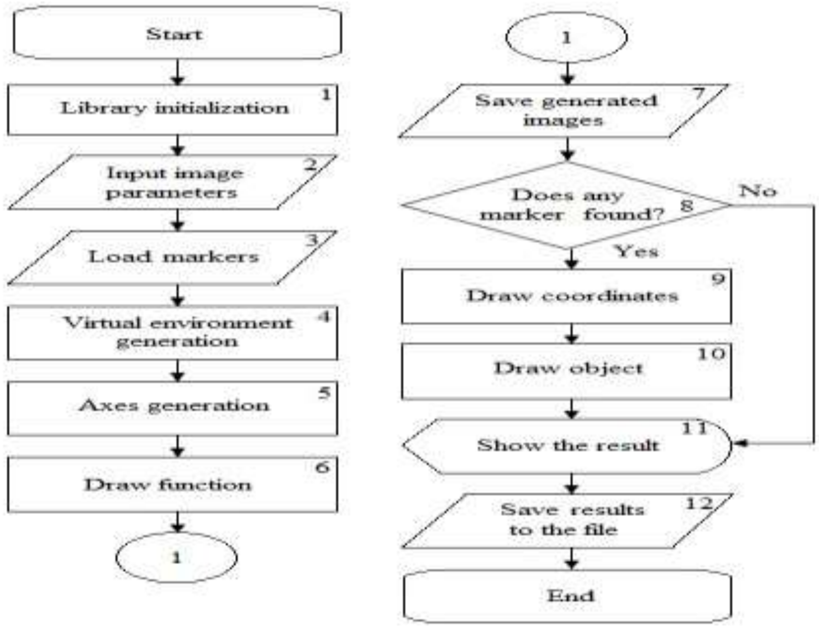

Figure 4. Flowchart

Result and Discussion

This section provides an examination of the results of an augmented reality (AR) application that includes optical character recognition (OCR) capabilities and uses the Firebase database as the backend solution. The analysis focuses on the application's performance, accuracy, and user experience. It analyses the metrics utilized for evaluation, gives the results, and provides insights into the AR-OCR application's strengths, limits, and prospective areas for improvement with Firebase integration.

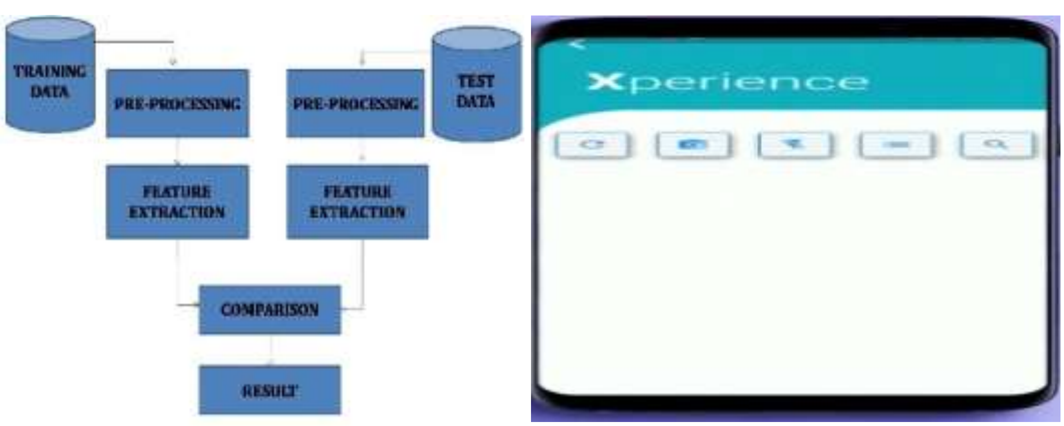

Figure 5. Results of an Augmented Reality (AR) Application

Conclusion

Finally, combining augmented reality (AR) and optical character recognition (OCR) technology with Firebase as the backend database opens up new opportunities for developing unique apps. Using Firebase's sophisticated capabilities, this combination enables real-time text recognition in AR contexts as well as smooth data storage and retrieval. We investigated the principles, advantages, problems, and prospective applications of an AR-OCR application utilizing Firebase in this literature study. We've seen how AR technology improves the user experience by superimposing digital information on real-world items, and how OCR technology extracts and recognizes text from photos. Developers may make use of Firebase's real-time data synchronization, offline support, and authentication features by using it as the backend infrastructure.

This review's system architecture serves as a model for developers to construct their own AR-OCR apps. The architecture describes the major components, such as the user interface, picture capture, OCR module, Firebase database, and data flow between them. It emphasizes the significance of an easy-to-use interface, picture preprocessing, reliable OCR techniques, and safe data storage. The use of Firebase as the backend database guarantees that data is stored and retrieved in real time. Images may be effectively saved and retrieved using Firebase Storage, while OCR results and related information can be stored using Firebase Realtime Database or Firestore. This allows users to retrieve OCR data from many devices by enabling real-time synchronizations between the program and the database. In terms of applications, an AR-OCR system powered by Firebase has several possibilities. It may be used in language translation programs to display real-time translations of text in several languages. It may also be used in real-world text extraction applications such extracting data from signs, papers, or product labels. Additionally, interactive educational experiences that allow users to explore and engage with textual content may be created.

Acknowledgement

The researcher did not receive any funding for this study, and the results have not been published in any other sources.

References

Huang, J., Kinateder, M., Dunn, M. J., Jarosz, W., Yang, X. D., & Cooper, E. A. (2019). An augmented reality sign-reading assistant for users with reduced vision. PloS one, 14(1), e0210630.

Singh, A., Ghasemi, Y., Jeong, H., Kim, M., & Johnson, A. (2021). A comparative evaluation of the wearable augmented reality-based data presentation interface and traditional methods for data entry tasks. International Journal of Industrial Ergonomics, 86, 103190.

Mori, S., Suen, C. Y., & Yamamoto, K. (1992). Historical review of OCR research and development. Proceedings of the IEEE, 80(7), 1029-1058.

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., ... & Monge-Sanz, B. M. (2019). SEAS5: the new ECMWF seasonal forecast system. Geoscientific Model Development, 12(3), 1087-1117.

Li, T., Li, C., Zhang, X., Liang, W., Chen, Y., Ye, Y., & Lin, H. (2021). Augmented reality in ophthalmology: applications and challenges. Frontiers in Medicine, 8, 733241.

Torres, W., van den Brand, M. G., & Serebrenik, A. (2022). Xama: Optical character recognition for multi-domain model management. Innovations in Systems and Software Engineering, 1-25.

Li, B., Wang, X., Gao, Q., Song, Z., Zou, C., & Liu, S. (2022). A 3D Scene Information Enhancement Method Applied in Augmented Reality. Electronics, 11(24), 4123.

Carvalho, R., Melo, J., Graça, R., Santos, G., & Vasconcelos, M. J. M. (2023). Deep learning-powered system for real-time digital meter reading on edge devices. *Applied Sciences*, *13*(4), 2315.

Laine, M., & Nevalainen, O. S. (2006, September). A standalone OCR system for mobile cameraphones. In *2006 IEEE 17th international symposium on personal, indoor and mobile radio communications* (pp. 1-5). IEEE.

Khawas, C., & Shah, P. (2018). Application of firebase in android app development-a study. *International Journal of Computer Applications*, *179*(46), 49-53.