International Journal of Advances in
Soft Computing and Intelligent Systems (IJASCIS)
2024, Vol 03, Issue 02, 253-265
ISSN: 3048-4987

Original Paper

TEXT RECOGNITION AND 3D RENDERING USING OPTICAL CHARACTER RECOGNITION AND AUGMENTED REALITY

Pushpa T S[a], Sumanth V[b], Shreyas C.S.[c*], Prajwal R Bhat[d], Nandan Kumar[e]

[a,b,c,d] Department of Computer Applications, B.M.S. College of Engineering, Karnataka, India, pushpa.mca@bmsce.ac.in, sumanthv.mca21@bmsce.ac.in, shreyas.mca22@bmsce.ac.in, prajwal.mca22@bmsce.ac.in

[e] Department of Computer Applications, Dayanand Sagar Business Academy, Karnataka, India, 15nandankumar@gmail.com

ABSTRACT

Optical character recognition is a modern technique that allows text to be extracted from photographs or real-world objects. This application seeks to provide an AR based OCR system that makes use of the Fire- base database as a database. Using AR, users may aim their smartphone cameras toward actual papers, signs, or any textual material, and the application will automatically recognize and extract the text in the taken image. After extraction, it displays in 3D format if the object is present in the data- base.

This application's integration with Firebase Database is crucial because it provides a scalable and real-time data storage option. Stored text offers smooth real-time synchronization between users and devices by utilizing the Firebase Realtime Database. This makes it easier for people to interact with one text at a time, fostering shared and collaborative experiences. Locating and obtaining particular text material is made easier by the database's powerful querying and indexing capabilities. Additionally, using Firebase Authentication to manage user authentication and access control is crucial. The development of services like text-based search and filtering depends heavily on this functionality. While providing the secure preservation of user-specific data, it enables the production of personalized experiences. Firebase Authentication delivers a customized and secure user experience, while Firebase Database acts as the framework, enabling dynamic and collaborative engagement with text in real-time.

KEYWORDS

Text Extraction, Optical Character Recognition, Augmented Reality

1. 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. Using a method called OCR (Optical Character Recognition), text can be taken out of images or actual things. These two technologies can be combined to develop complex apps that can recognize and respond to text in real-world settings. A platform for creating mobile and online applications is Google Fire- base. A real time database, authentication, hosting, and cloud capabilities are few of the services it provides. Developers can store and sync data in real time using the cloud- hosted Firebase Realtime Database, a NoSQL database. Both technologies may be combined in an Optical Character Recognition Augmented Reality application that makes use of Firebase Database.

This application's integration with Firebase Database is crucial because it provides a scalable and real-time data storage option. Stored text offers smooth real-time synchronization between users and devices by utilizing the Firebase Realtime Database. This makes it easier for people to interact with one text at a time, fostering shared and collaborative experiences. Locating and obtaining particular

text material is made easier by the database's powerful querying and indexing capabilities. Additionally, using Firebase Authentication to manage user authentication and access control is crucial. The development of services like text- based search and filtering depends heavily on this functionality. While providing the secure preservation of user-specific data, it enables the production of personalized experiences. Firebase Authentication delivers a customized and secure user experience, while Firebase Database acts as the framework, enabling dynamic and collaborative engagement with text in real-time [6].

In this paper, Section 2 reviews the literature survey, Section 3 details the methodologies, Section 4 presents the problem statement, Section 5 outlines the proposed solution, Section 6 discusses the benefits, Section 7 covers the technologies used, Section 8 illustrates the results, Section 9 wraps up the work, and Section 11 explores future enhancements.

2. LITERATURE SURVEY

[1] The development of holography made complex optical linear filtering procedures possible. A traditional optical character recognition technique was proposed by Vander Lugt, which involved Fourier-transforming the input character (u1) and using a holographic filter matched to the test character (u2). In order to measure the direct cross correlation between u1 and u2, the filtered output is retransformed. Three difficulties arise when trying to tell between similar characters like "B," "E," and "F." First of all, they produce equivalent correlation values, which calls for exact thresholds for identification. Second, character distortion or rotation may throw off the decision-making process. Thirdly, it is difficult to prepare matched filters with the necessary accuracy. While the challenges of filter preparation are feasible, the fundamental limitations of strong cross correlations between identical features and limited distortion tolerance point to the necessity for a different strategy.

[2] Cost-effective thresholding algorithms are designed for extracting binary character images from printed documents, whether machine-generated or handwritten. By detecting whether a point in an analogue image belongs to a character stroke (resulting in a binary one) or not (resulting in a binary zero), they are essential to optical character recognition (OCR). This procedure is essential for OCR as well as other Character Image Extraction (CIE) jobs like handling bank checks or carbon copy forms with potentially problematic features like smudges or intricate backgrounds. A wide range of CIE applications can use the first algorithm, a lean, adaptable technique, which has little hardware requirements. The second strategy, which is more complicated, is designed for complex, high- volume applications.

[3] Fuzzy classifiers have demonstrated great potential in character recognition. However, defining membership functions and classification rules, even for just the 10 digits and 23 Roman alphabet characters, poses a significant challenge. This paper addresses this challenge by introducing a semi-automatic approach for designing a Fuzzy classifier tailored for letters and digits, specifically targeting the automatic recognition of car license plates in unstructured environments. The CART algorithm is used to train the system, which then automatically creates fuzzy rules using a training set of fuzzified samples taken from digital images of specific characters. A fuzzy inference engine then uses these created rules to classify characters. A hierarchical classifier with two layers is also presented, which incorporates a syntactical corrective phase by first distinguishing between letters and digits before performing classification. This two-layer classifier's performance is carefully evaluated and analysed, proving how effective it is in improving character recognition.

[4] The proposed approach for offline cursive handwriting recognition presents a distinctive methodology employing a sequential sequence of image segmentation and recognition algorithms. Global parameters like baselines and slant angles are initially estimated. After that, character segmentation routes are created using a segmentation technique that blends grayscale and binary data. Then, using the recovered codes to classify and rank potential character candidates, a hidden Markov model (HMM) is used for shape recognition. It is noted that the feature space and HMM model parameters are computed simultaneously. Last but not least, a graph optimization scheme that

combines lexical information with HMM ranks dramatically raises recognition rates and outperforms existing approaches. This ground-breaking technique presents a thorough framework for precisely interpreting cursive handwriting in offline contexts, demonstrating a multidimensional strategy that overcomes significant character identification issues.

[5] This study proposes a semi-automatic Fuzzy classifier architecture for recognizing alphanumeric characters on license plates in challenging, unstructured conditions. Even while fuzzy classifiers show promise for character recognition, defining membership functions and rules is still difficult, particularly for characters from the Roman alphabet and digits. The work uses the CART approach, which automatically creates fuzzy rules from a training set of fuzzified digital character pictures, to address this. A two-layer hierarchical classifier that is supplemented with syntactical correction methods improves performance. This method considerably improves character recognition tasks' accuracy and resilience.

[7] "Optical Character Recognition Using Automatically Generated Fuzzy Classifiers" presents a novel approach to improving the character recognition system by using fuzzy logic. Their research addresses the challenge of defining membership functions and designing classification rules for optical character recognition, in particular recognizing digits and letters on car license plates under unstructured conditions. Their results point to the effectiveness of the method, which shows high accuracy in character recognition despite the lowering effect of low-cost imaging and environmental factors. This contribution gives a practical and automated method of generating fuzzy rules that can be easily adapted for other complex classification tasks where traditional methods may fail.

[8] This study examines the use of augmented reality (AR) to display the outer design of college buildings. Authors stress that traditional architectural models often require precise scaling and considerable amount of time, thus are easily replaced by augmented technology. AR allows users to engage with the real world while virtual objects are overlaid, providing a more interactive, three-dimensional visualization of buildings. The application of AR provides prospective students the ability to examine the building from different perceptions through mobile applications, by using pamphlets as markers. This method not only adds value to printed materials but also expands feasibility of low cost and interaction in terms of architectural modelling, making it an excellent option for academic promotions.

[9] The authors provided a summary of the respective literature on the advancements of optical character recognition (OCR), establishing greater emphasis on handwritten documents than on typed documents. Reviews of this nature underscore how pervious advances have moved OCR advances from incredibly simple systems relying on extremely rudimentary feature extraction, to holistic deep learning methods. Previously, OCR behaved like previous advances constrained to methods like Support Vector Machine, k-Nearest Neighbors, and Decision Trees. Neural networks have greatly increased the precision of text recognition using convolutional neural networks and Long Short Term Memory networks. In general, recent papers have focused on applications of deep learning, computations with GPU use, as well as refinements to feature extraction capabilities for understanding handwritten text in different languages. The paper discussed forms of optical character recognition, concluding more work should be conducted on overlapping characters and variability in handwritten text.

[10] The article "Text Detection and Recognition Using Augmented Reality and Deep Learning" explains that the authors plan to combine OCR (optical character recognition) with augmented reality in order for users who are visually impaired to read text that appears in their surroundings. They put forward a concept of a mobile application which captures text with the camera of the smartphone while automatically detects and recognizes text using deep learning models, specifically the VGG-16 neural network. Once it identifies that text, the application renders the text for the user in easier to read 3D with AR. By doing so, the user can visualize and better interact with the text, which would be easier and more intuitive for the user to read and comprehend text in real life. The authors provide a strong promise that the application can enhance user usability and provide greater interactivity by advancing accessibility through a state-of-the art AR experience that benefits from modern OCR

systems. This research is a useful case to demonstrate the growing use of augmented reality in combination with OCR to create more immersive and engaged applications.

[11] The paper covers advanced image processing algorithms necessary for AR/VR applications, including real-time rendering, object recognition, feature detection, image registration, and depth estimation, all of which are relevant to generating an immersive virtual environment. This paper points out that feature detection can be done using Scale-Invariant Feature Transform (SIFT) and tracking/registration using Extended Kalman Filter (EKF) to enhance real-time performance and accuracy. It also touches on the development of machine learning, mainly CNNs, for the improvement of object detection and interaction in such environments. The authors further show that future studies into AI shall include integrations, optimizations, and the development of uniform frameworks that will increase scalability and performance.

3. METHODOLOGIES

A systematic strategy is necessary to create the envisioned AR-based OCR system with Firebase backend. It starts with project planning, setting goals, and determining the requirements of the target audience. It's important to choose the right tech stack, which includes appropriate programming languages and AR development kits.

AR functionality integration involves setting up the camera feed for real- world image capture, overlaying OCR capabilities, and processing images for text extraction. Firebase integration includes setting up a Firebase project, creating a Firebase Realtime Database, and incorporating Firebase SDK. The database schema is designed for efficient storage of recognized text data.

Using Firebase Authentication, OCR access is made secure. High quality is ensured by thorough testing, which includes user acceptance testing. For a better experience, a user-friendly interface with clear text displays and AR overlays is built.

4. PROBLEM STATEMENT

The combination of augmented reality (AR) with optical character recognition (OCR) technology has the potential to fundamentally alter how people interact and retrieve information in the physical world. However, a number of obstacles must be overcome in order to completely reap the rewards of AR OCR applications. Finding and removing major roadblocks to developing and implementing effective AR applications for optical character recognition is the main challenge of this project. These cover worries about precision, instantaneous processing, user interface layout, and privacy/security. Our research intends to improve the usefulness, dependability, and viability of AR OCR systems by addressing these problems head-on, paving the road for their wider adoption across various industries and fields.

5. PROPOSED SYSTEM

The incorporation of Firebase can significantly improve the proposed system for the AR OCR app. The device's camera will take important images or video frames, which will be stored on Firebase Cloud Storage to enable easy processing. Using Firebase Cloud Functions, OCR processing will be applied to uploaded photos or frames in order to extract text using a library like Tesseract. The Firebase Realtime Database will record the recognized text along with any relevant metadata. The text and related data, such as timestamps and user information, will be stored in this database. This information will be used by the AR framework (AR Core or ARKit) to provide an immersive AR display that overlays the camera feed with real-time, recognized text. The result of this integration is a system that is highly scalable, safe, and efficient. It also provides the AR OCR application with strong user authentication, effective cloud storage, dynamic real-time data- base capabilities, and optimized serverless processing.

6. BENEFITS

Augmented Reality: It adds a layer of virtual information to the actual environment, making it more immersive and engaging for users. Incorporating OCR into AR apps allows users to extract text from physical items and interact with virtual material in real-time, resulting in a more engaging and fascinating user experience.

Text Recognition: OCR technology can recognize and extract text from photos or live video feeds. By integrating OCR with AR, users may aim their device's camera towards text containing items, and the application will extract and interpret the text in real time. This is especially beneficial for activities like extracting information from documents, and giving real time subtitles during presentations or conferences.

7. TECHNOLOGIES USED

Firebase Database: A cloud-hosted NoSQL database offered by Google as part of the Firebase platform. It enables developers to store and sync data between clients and backend servers in real time.

Flutter Framework: Google's Flutter is an open-source UI software development kit (SDK). From a single codebase, developers may create natively built apps for mobile, web, and desktop platforms.

AR Framework: AR Core and ARKit may create the AR view, overlaying the camera feed with the recognized text acquired from Firebase Realtime Database.

Android Studio IDE: Android Studio is the official integrated development environment (IDE) for developing Android apps.

Java: Java is a popular programming language for Android application development, especially augmented reality (AR) applications.

OCR: It is a method that gives users the ability to convert text or documents found in images taken by an input device into an editable, searchable, and reusable data format for additional image processing. This technique enables a machine to recognize letters automatically using an optical system.

8. RESULT

Figure 1. Home Screen

Figure 2. Capture Text by Camera

Figure 3. Result

Figure 4. Capture Text by Camera

Figure 5. Result

Figure 6. Searched List

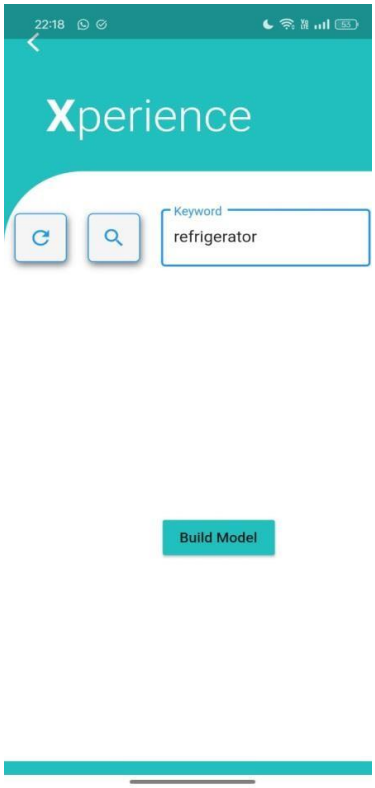

Figure 7. Search by Text

Figure 8. Result

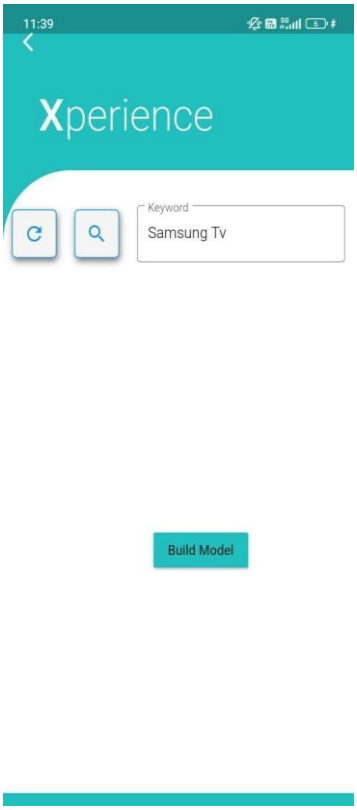

Figure 9. Search by text

Figure 10. Result

Figure 11. Database

Figure 12. Firebase Storage

9. CONCLUSION

The integration of augmented reality (AR) technology with optical character recognition (OCR) holds tremendous potential in revolutionizing the way we recognize and display printed or handwritten text. The feasibility and effectiveness of AR-based OCR systems have been demonstrated in various studies, showcasing their capability to accurately recognize both printed and handwritten text while immersing users in an augmented reality environment. Despite their potential, AR- based OCR systems have significant drawbacks that need for more research. The use of AR technology in OCR creates interesting new opportunities across a variety of industries.

AR devices empower students to scan textbooks, accessing additional information and 3D models tied to the content, fostering a more engaging and immersive learning experience. In retail, shoppers can effortlessly access real-time product details, prices, and reviews by scanning product labels. For construction workers, OCRAR integration facilitates the overlay of digital data onto physical blueprints, granting them a direct visualization of plans, measurements, and specifications right at the construction site. This innovation not only enhances efficiency but also elevates accuracy and precision in executing tasks, showcasing the transformative potential of OCR-AR applications across various industries.

10. FUTURE ENHANCEMENT

The future of Optical Character Recognition (OCR) integrated with Augmented Reality (AR) technology holds immense promise, heralding a paradigm shift in how we interact with printed and handwritten text. The potential for creativity is enormous as augmented reality technology develops and becomes more widely available. The combination of OCR with AR promises to improve translation efforts and enhance user experience. With this dynamic pair, users only need to point their AR-enabled devices towards text for OCR algorithms to quickly recognize and engage in real-time presentation of the content. Additionally, the fusion of AR with OCR has the potential to completely eliminate language barriers, especially in the travel and tourist industry. It is capable of rapidly

translating menus, signage, and other printed items, greatly facilitating conversation and navigating in foreign environments. This ground-breaking combination of OCR with AR technology marks an impressive advance in information processing, communication, and accessibility, paving the way for a new era of interactive, real- world text interaction.

REFERENCES

[1] H. J. Caulfield and W. T. Maloney, 'Improved Discrimination in Optical Char- acter Recognition', Appl. Opt., vol. 8, no. 11, p. 2354 (1969)

[2] J. M. White and G. D. Rohrer, 'Image Thresholding for Optical Character Recognition and Other Applications Requiring Character Image Extraction', IBM J. Res. & Dev., vol. 27, no. 4, pp. 400–411 (1983)

[3] G. L. Cash and M. Hatamian, 'Optical character recognition by the method of moments', Computer Vision, Graphics, and Image Processing, vol. 39, no. 3, pp. 291–310 (1987)

[4] S. Impedovo, L. Ottaviano, and S. Occhinegro, 'OPTICAL CHARACTER RECOGNITION — A SURVEY', Int. J. Patt. Recogn. Artif. Intell., vol. 05, no. 01n02, pp. 1–24 (1991)

[5] N. Arica and F. T. Yarman-Vural, 'Optical character recognition for cursive handwriting', IEEE Trans. Pattern Anal. Machine Intell., vol. 24, no. 6, pp. 801–813 (2002)

[6] Chetankumar, G. Shetty, Augmented Reality—Theory, Design and Development (1st ed.) McGraw Hill. (2020)

[7] J. M. Fonseca, N. M. Rodrigues, A. D. Mora, and R. A. Ribeiro, 'Optical character recognition using automatically generated Fuzzy classifiers', in 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Shanghai: IEEE, pp. 448–452 (2011)

[8] M. B. Firdaus, J. A. Widians, and J. Y. Padant, 'Augmented reality for interactive promotion media at Faculty of Computer Science and Information Technology Mulawarman University', J. Phys.: Conf. Ser., vol. 1341, no. 4, pp. 042- 017 (2019)

[9] J. Memon, M. Sami, R. A. Khan and M. Uddin, "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)," in *IEEE Access*, vol. 8, pp. 142642-142668, 2020, doi: 10.1109/ACCESS.2020.3012542.

[10] Ouali, Imene & Ben Halima, Mohamed & Wali, Ali. (2022). Text Detection and Recognition Using Augmented Reality and Deep Learning. 10.1007/978-3-030-99584-3_2.

[11] Divya Nimma. (2024). Image Processing in Augmented Reality (AR) and Virtual Reality (VR). International Journal on Recent and Innovation Trends in Computing and Communication, 12(2), 475–482.

Authors

Pushpa T S received her B.Sc degree from Bangalore University in 1992, MCA degree from Bangalore University in 1998 and M.Phil degree from Madurai Kamaraj University in 2008 and currently working as Assistant Professor in Department of Computer Applications, B.M.S. College of Engineering, Bangalore. At present she is pursuing her PhD under the guidance of Dr. K Vijaya Kumar, Department of Master of Computer Applications, B.M.S. College of Engineering, Bangalore from VTU, Belgaum. Her areas of interest are Data Mining, Data Analytics and Machine Learning.

Sumanth V received his degree in Computer Applications from BMS College of Engineering, Bangalore, in 2023 and his BCA degree from Dayananda Sagar Business Academy, Bangalore, in 2021. He is currently working on projects involving front-end and IoT technologies. His notable projects include an Amazon clone, an Instagram clone, and an automated irrigation system utilizing IoT and Augmented Reality. Sumanth's research interests focus on integrating emerging technologies in practical applications.

Shreyas C.S. is a Master of Computer Applications student at B.M.S. College of Engineering, Bengaluru. He holds a Bachelor of Computer Applications from Acharya Institute of Graduate Studies, Bengaluru. Shreyas has worked on several projects, including an image generation tool using the MERN stack, a library management app for his college with Java and Firebase, and a meal distribution system using PHP and MySQL. His technical expertise spans MEAN/MERN stack technologies, mobile app development, and various programming languages.

Prajwal R Bhat is a Master of Computer Applications student at B.M.S. College of Engineering, Bengaluru Prajwal R Bhat is a Master of Computer Applications student at B.M.S. College of Engineering, Bengaluru. He holds a Bachelor of Computer Applications from NIE College, Mysore. Prajwal has worked on several projects, including a medical app that tracks the popularity of drugs and the side effects reported by users, as well as a helmet detection system to track and control traffic violations. His technical expertise includes Java, Python, and HTML and various other technologies.

Nandan Kumar N received his BCA degree from Dayananda Sagar Business Academy in 2021 and his MCA degree from Dayananda Sagar Academy of Technology and Management in 2023. He is currently working as a Software Development Engineer at Likhitech Pvt Ltd, Bengaluru. His areas of expertise include full-stack development with experience in Angular, .NET Core, PHP Laravel, and data analytics with Python. His technical interests span across software engineering, web technologies, and database management.