International Journal of Advances in
Soft Computing and Intelligent Systems (IJASCIS)
2025, Vol 04, Issue 01, 14-28
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.

[e] Department of Computer Applications, Dayanand Sagar Business Academy, Karnataka, India.

[a]pushpa.mca@bmsce.ac.in, [b]umanthv.mca21@bmsce.ac.in,

[c]shreyas.mca22@bmsce.ac.in, [d]prajwal.mca22@bmsce.ac.in,

[e]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. This research focuses on the application of AR and OCR for recognizing text and rendering 3D models. We leverage different technologies to develop an application that enables users to aim their device at any text they have written on various materials. Using optical character recognition, the application detects the handwritten text and, if a 3D model is stored in the database, it will render and show that model.

KEYWORDS

Text Extraction, Optical Character Recognition, Augmented Reality.

1. INTRODUCTION

Augmented Reality (AR) is a technology that merges the real and virtual worlds, enhancing how users perceive and interact with their environment. By using Optical Character Recognition (OCR), text can be extracted from images or physical objects. The combination of these technologies enables the development of advanced applications that can recognize and respond to text in real-world settings. Google Firebase is a platform for building mobile and web applications, providing services like real-time databases, authentication, hosting, and cloud functionalities. The Firebase Realtime Database, a NoSQL database, allows developers to store and synchronize data in real time through cloud hosting. An Optical Character Recognition Augmented Reality application can effectively combine these technologies with the Firebase Database.

Integrating this application with the Firebase Database is crucial as it offers a scalable and real-time data storage solution. By utilizing the Firebase Realtime Database, stored text can be seamlessly synchronized among users and devices, enabling interaction with one piece of text at a time and fostering shared and collaborative experiences. The database's strong querying and indexing features make it easier to find and retrieve specific text content. Additionally, managing user authentication and access control through Firebase Authentication is essential. This functionality is important for creating services like text-based search and filtering, allowing for personalized experiences while ensuring the secure storage of user-specific data. Firebase Authentication provides a customized and secure user experience, while Firebase Database acts as the foundation, facilitating dynamic and collaborative engagement with text in real time [6].

In this paper, Section 2 provides an overview of the literature survey, Section 3 describes the methodologies used, Section 4 states the problem at hand, Section 5 lays out the proposed solution, Section 6 discusses the benefits of this approach, Section 7 highlights the technologies used, Section 8 showcases the results obtained, Section 9 concludes the work, and Section 11 looks ahead to potential future enhancements.

Pixel-based classifications are according to the results grey value of pixels, and only spectral information is used for classification. These are the smallest units that represent an image. This method makes use of reflectance statistics for specific pixels. It assembles pixels to represent land cover characteristics [4]. Supervised classification is a popular method for analysing sat images to identify and categorize land cover or land use classes Unsupervised organization is a technique used in sat image analysis and remote sensing to automatically categorize pixels into distinct classes without the need for pre-defined training data or class labels.

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.

[3] Multilevel fuzzy classifiers have been utilized greatly in character recognition. However, extracting membership functions and classification rules for only the 10 digits and 23 roman alphabet characters can be quite challenging. To overcome this challenge, we propose a semi-automatic method for fuzzy classifier designing that is mainly based on the automatic car plate recognition task in un-constrained environments, on letters and digits. Using the CART algorithm, the system is trained and produces fuzzified rules derived from original samples selected from the digital images of the particular characters. A fuzzy inference engine utilizes these established rules to classify the characters.

[4] It presents the approach for offline cursive handwriting recognition which is based on the series of image segmentation and recognition algorithms. Global parameters such as baselines and slant angles are first estimated. Afterward, character segmentation paths are generated using a method that combines grayscale and binary information. A hidden Markov model (HMM) for shape recognition is used by matching extracted codes with character candidates and ranking the candidates. Notice that the feature space and HMM model parameters are computed concurrently. Last but not least, a graph optimization strategy that combines lexical information with hidden Markov model (HMM) rankings

to improve recognition rates significantly compared to current state-of-the-art. This approach provides a complete framework for accurately interpreting cursive handwriting.

[5] The goal of this paper is building a semi-automatic fuzzy classifier system for stamping on vehicle license plate through harsh and disorganized condition. Despite this, the limitation still persists in other fuzzy classifiers that specifying membership functions and rules is hard; this is particularly the case for characters from the roman-alphabet and digits. To overcome this the work implement the CART approach in which fuzzy rules are generated from a set of fuzzified digital character images provided as training set. However, the second layer also usually uses syntactic corrections and improves the expressed results by combining a two-layer hierarchy classifier. Such a procedure improves significantly the accuracy and robustness of the character recognition systems.

[7] The paper, "Optical Character Recognition Using Automatically Generated Fuzzy Classifiers" proposes a new way of enhancing the character recognition system through the use of fuzzy logic. Their research focuses on the problem of determining the membership functions and the classification rules for optical character recognition, and specifically for the digits and letters on the car license plates in an unstructured environment. According to their findings, the method is efficient as it achieves very high accuracy in character recognition even with the decrease in cost of imaging and other environmental conditions. This contribution provides a real and effective way of developing fuzzy rules that can be applied to other forms of complicated classifications that would otherwise be difficult to solve by conventional techniques.

[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.

In the greater Godavari basin, Tmax ranges from 29.64 to 38.60 the average monthly, and an average annual Tmax of 32.45 [6]. An average annual rainfall of 1132 mm in the Godavari River basin has obtains, with the monsoon season accounting for nearly 84% of total rainfall [7]. Satellite Resourcesat-2 LISS-III imagery used in the study and acquired from free sphere of Bhuvan Indian Geo-Platform of ISRO. The image is ortho rectified [8].

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.

In this paper, Firebase Authentication is used to ensure that OCR is secure. This is done by conducting user acceptance testing to ensure quality. In order to provide a better user experience, a friendly interface with quality text display and augmented reality (AR) overlays are designed.

4. PROBLEM STATEMENT

The integration of augmented reality with optical character recognition technologies presents an opportunity to revolutionise how people interact with and extract information from the physical environment. However, there are a number of challenges that need to be addressed in order to fully realise the benefits of AR OCR applications. The problem of this project is how to identify and overcome the major barriers to creation and deployment of efficient AR applications for optical character recognition. These include concerns on the accuracy, real time processing, interface design and privacy issues. The purpose of this research is to identify the potential challenges that hinder the effectiveness, reliability and feasibility of AR OCR and to find ways on how to mitigate them in order to facilitate the adoption of this technology in the various fields of application.

5. PROPOSED SYSTEM

The addition of Firebase can greatly enhance the proposed system for the AR OCR app. The camera of the device will capture important photos or video frames which will be saved on Firebase Cloud Storage for further processing. Cloud Functions of Firebase would be used to apply OCR processing to the photos or frames uploaded to the app to identify text with the help of a library like Tesseract. All the recognized text along with other metadata will be saved in the Firebase Realtime Database. The text and other information like time and date and user details will be stored in this database. This information will be utilized in the AR framework for generating an AR display that can show real time camera feed with text detection. Thus, the outcome of this integration is a system that is very much flexible, secure and effective.

6. BENEFITS

Augmented Reality: Augmented Reality enhances the real world by overlaying a virtual layer, creating a more interactive and enjoyable experience for users. When Optical Character Recognition (OCR) is integrated into AR applications, users can extract text from real-world objects and engage with virtual content in real time, resulting in a more captivating and exciting experience.

Text Recognition: OCR can recognize and extract text from photos or live video feeds. By adding OCR to AR, users can point their device's camera at text containing objects and the app will extract and read the text in real time. This is super useful for extracting info from documents and giving real time subtitles during presentations or conferences.

7. TECHNOLOGIES USED

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

Flutter Framework: Google's Flutter is an open source UI SDK. From a single codebase, developers can build natively compiled apps for mobile, web and desktop.

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

Android Studio IDE: Android Studio is the official IDE for Android app development.

Java: Java is a programming language for Android app development, especially AR apps.

OCR: It's a method that allows users to convert text or documents found in images taken by an input device into an editable, searchable and reusable data format for further image processing. This technique allows a machine to recognize letters using an optical system.

8. RESULT

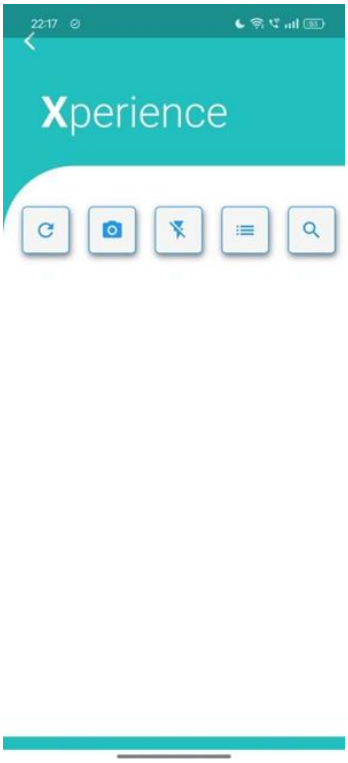


Fig.1. Home Screen.

Fig. 2. Capture Text by Camera.

Fig. 3. Result.

Fig. 4. Capture Text by Camera.

Fig. 5. Result.

Fig. 6. Searched List.

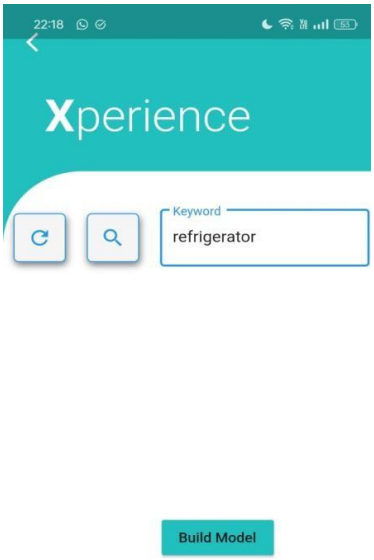


Fig. 7. Search by Text.

Fig. 8. Result.

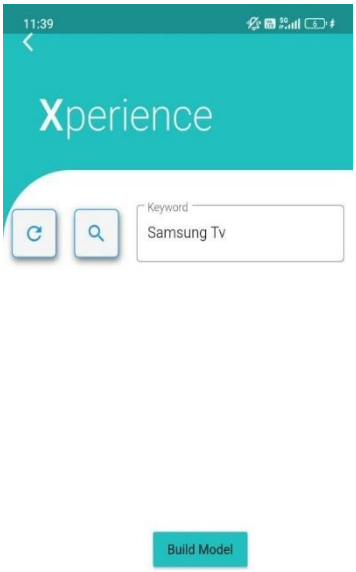


Fig. 9. Search by text.

Fig. 10. Result.

Fig. 11. Database.

Fig. 12. Firebase Storage.

9. CONCLUSION

AR and OCR together is huge. There have been many studies that have shown the feasibility and effectiveness of AR based OCR to recognize both printed and handwritten text while immersing the user in an AR environment. But AR based OCR has its downsides that needs more research. AR in OCR opens up new opportunities across many industries. AR devices allow students to scan textbooks and access additional info and 3D models related to the content. In retail, shoppers can scan product labels and get real time product details, prices and reviews. For construction workers, OCRAR integration allows them to overlay digital data onto physical blueprints and see plans, measurements and specs right at the construction site. This not only makes them more efficient but also more accurate and precise in their work. OCR-AR is a game changer across many industries.

10. FUTURE ENHANCEMENT

The future of OCR with AR has immense potential, with a whole new way of interacting with printed and handwritten text. As AR tech develops and becomes more mainstream the possibilities are endless. OCR with AR will improve translations and user experience. With this combo users just point their AR enabled devices at text and OCR algorithms will quickly recognise and present the content in real time. AR with OCR can also eliminate language barriers, especially in travel and tourism. It can translate menus, signs and printed materials in seconds, making conversation and navigating in foreign environments a breeze. This is a major step forward in information processing, communication and accessibility, a new era of interactive real-world text.

Authors Contributions (Compulsory):

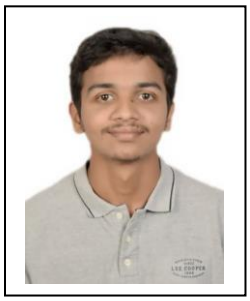
All authors have equally contributed.

REFERENCES

1. Caulfield, H.J., and W. T. Maloney, ‘Improved Discrimination in Optical Char- acter Recognition’, Appl. Opt., vol. 8, no. 11, p. 2354 (1969)
2. White, J.M., 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. Cash, G. L., 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. Impedovo, S., 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. Arica, N., 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. Fonseca, J.M., 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. Firdaus, M.B., 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. Memon, J., 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.

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.