

**SHOPRECOG: MOBILE APPLICATION FOR SHOP RECOGNITION WITH TEXT
TO SPEECH**

**BY
KHOO ZI YI**

**A REPORT
SUBMITTED TO
Universiti Tunku Abdul Rahman
in partial fulfillment of the requirements
for the degree of
BACHELOR OF COMPUTER SCIENCE (HONOURS)
Faculty of Information and Communication Technology
(Kampar Campus)**

JAN 2024

REPORT STATUS DECLARATION FORM

Title: ShopRecog: Mobile Application for Shop Recognition with Text to Speech

Academic Session: Jan 2024

I KHOO ZI YI
(CAPITAL LETTER)

declare that I allow this Final Year Project Report to be kept in
Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Verified by,



(Author's signature)



(Supervisor's signature)

Address:

1-01, Pangsapuri Emas,

Jalan Raja Uda

12300, Butterworth, Pinang

NG HUI FUANG

Supervisor's name

Date: 25/4/2024

Date: 26/4/2024

Universiti Tunku Abdul Rahman			
Form Title : Sample of Submission Sheet for FYP/Dissertation/Thesis			
Form Number: FM-IAD-004	Rev No.: 0	Effective Date: 21 JUNE 2011	Page No.: 1 of 1

FACULTY/INSTITUTE* OF INFORMATION AND COMMUNICATION TECHNOLOGY

UNIVERSITI TUNKU ABDUL RAHMAN

Date: 25/4/2024

SUBMISSION OF FINAL YEAR PROJECT /DISSERTATION/THESIS

It is hereby certified that *Khoo Zi Yi* (ID No: 20ACB03614)
has completed this final year project/ dissertation/ thesis* entitled “ ShopRecog: Mobile Application for Shop Recognition with Text to Speech ” under the supervision of Dr. Ng Hui Fuang
(Supervisor) from the Department of Computer Science, Faculty/Institute* of Information and Communication Technology

I understand that University will upload softcopy of my final year project / dissertation/ thesis* in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,




Khoo Zi Yi

*Delete whichever not applicable

DECLARATION OF ORIGINALITY

I declare that this report entitled “**SHOPRECOG: MOBILE APPLICATION FOR SHOP RECOGNITION WITH TEXT TO SPEECH**” is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature :  _____

Name : Khoo Zi Yi

Date : 25/4/2024

ACKNOWLEDGEMENTS

I would like to express my sincere thanks and appreciation to my supervisors, Dr Ng Hui Fuang who has given me this bright opportunity to engage in a development-based project and guide me to solve the challenges during the process. It was a hard time to develop a project myself but luckily Dr Ng Hui Fuang was always by my side to provide assistant to me. A million thanks to you.

To a very special person in my life, Miss Sow Siew Kee, for her patience, unconditional support, and love, and for supporting me during hard times. Finally, I must say thanks to my friends and coursemates for their love, support, and continuous encouragement throughout the course.

ABSTRACT

This report documents the development and implementation of the "ShopRecog" mobile application, designed to address the unique needs of visually impaired individuals by leveraging advanced technologies such as machine learning, text recognition, and API integrations. The application provides real-time shop recognition, interactive speech-to-text functionality, comprehensive shop information retrieval, AI-generated summaries, and navigation assistance, all tailored for a seamless and user-friendly experience. Through rigorous testing, performance evaluation, and user feedback, the application demonstrates its effectiveness in enhancing independence, information access, and navigation support for visually impaired users. The report concludes with recommendations for continuous improvement, localization, adherence to accessibility standards, user feedback integration, and collaboration for further impact and innovation in assistive technologies.

TABLE OF CONTENTS

TITLE PAGE	i
REPORT STATUS DECLARATION FORM	ii
FYP THESIS SUBMISSION FORM	iii
DECLARATION OF ORIGINALITY	iv
ACKNOWLEDGEMENTS	v
ABSTRACT	vi
TABLE OF CONTENTS	vii
LIST OF FIGURES	xi
LIST OF TABLES	xiii
LIST OF ABBREVIATIONS	xiv

CHAPTER 1	1
INTRODUCTION	1
1.1 Problem Statement and Motivation	1
1.2 Objectives	1
1.3 Project Scope and Direction	2
1.4 Contributions	3
1.5 Report Organization.....	4
CHAPTER 2	6
LITERATURE REVIEW	6
2.1 Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)	6
2.1.1 Strengths	10
2.1.2 Weaknesses	10
2.1.3 Recommendation	10
2.2 Character detection and recognition system for visually impaired people	12
2.2.1 Strengths	15
2.2.2 Weaknesses	16
2.2.3 Recommendation	16
2.3 An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition 17	
2.3.1 Strengths	19
2.3.2 Weaknesses	20
2.3.3 Recommendation	20
2.4 Scene Text Recognition in Mobile Applications by Character Descriptor and Structure Configuration	21
2.4.1 Strengths	26
2.4.2 Weaknesses	26
2.4.3 Recommendation	26
2.5 MORAN: A Multi-Object Rectified Attention Network for scene text recognition	1
2.5.1 Strengths	34

2.5.2 Weaknesses	34
2.5.3 Recommendation	34
2.6 Seeing AI	36
2.6.1 Strengths and Weaknesses	37
CHAPTER 3.....	39
SYSTEM METHODOLOGY/APPROACH.....	39
3.1 System Design Diagram/Equation.....	39
3.1.1 Use Case Diagram and Description	42
3.1.2 Activity Diagram	44
CHAPTER 4.....	53
SYSTEM DESIGN.....	53
4.1 System Block Diagram	53
4.2 System Components Specifications.....	55
CHAPTER 5.....	58
SYSTEM IMPLEMENTATION.....	58
5.1 Hardware setup	58
5.2 Software setup	59
5.3 Setting and Configuration.....	59
5.4 System Operation (with Screenshot)	61
5.5 Implementation Issues and Challenges.....	71
5.6 Concluding Remark	72
CHAPTER 6.....	73
SYSTEM EVALUATION AND DISCUSSION.....	73
6.1 System Testing and Performance Metrics	73
6.2 Testing Setup and Result	74
6.3 Project Challenges	76
6.4 Objectives Evaluation	77

6.5 Concluding Remark	79
CHAPTER 7.....	80
CONCLUSION AND RECOMMENDATION.....	80
7.1 Conclusion	80
7.2 Recommendation	81
REFERENCES	83
WEEKLY LOG	84
POSTER	91
PLAGIARISM CHECK RESULT	92
FYP2 CHECKLIST	96

LIST OF FIGURES

Figure Number	Title	Page
Figure 2.1	An architecture of Multilayer Perceptron (MLP)	7
Figure 2.2	An overview of template matching techniques	8
Figure 2.3	(a) Primitive and relations (b) Directed graph for capital letter R and E	9
Figure 2.4	General block diagram of system	12
Figure 2.5	Layout floor planning procedure.	13
Figure 2.6	Text detection and extraction	14
Figure 2.7	Text recognition system output	14
Figure 2.8	Block diagram of T2S converter	15
Figure 2.9	The network architecture	17
Figure 2.10	The flowchart of the Designed Scene Text Extraction Method	22
Figure 2.11	Flowchart of the Character Descriptor	23
Figure 2.12	Overall structure of the MORAN	27
Figure 2.13	Results of the MORN on challenging image text	28
Figure 2.14	Difference in α_t for training with and without fractional pickup	30
Figure 3.1	System Design Flowchart	39
Figure 3.2	System Use Case Diagram	42
Figure 3.3	Move Camera Activity Diagram	44
Figure 3.4	Single Tap Activity Diagram	46
Figure 3.5	Ask Question Activity Diagram	48
Figure 3.6	Swap Activity Diagram	50
Figure 3.7	Long Press Activity Diagram	51
Figure 3.8	Double Tap Activity Diagram	52
Figure 4.1	System Block Diagram	53
Figure 5.1	Fetching Nearby Shop	61
Figure 5.2	shopNameList and shopIDList	62

Figure 5.3	Launching Application	63
Figure 5.4	Recognized Text	64
Figure 5.5	Summarized Text	65
Figure 5.6	Asking Question	66
Figure 5.7	Speech-to-Text Conversion	67
Figure 5.8	Navigation using Google	68
Figure 5.9	Swapping	69
Figure 5.10	Double Tap	70

LIST OF TABLES

Table Number	Title	Page
Table 2.1	Architecture of ASRN	29
Table 2.2	Strengths and Weaknesses of Seeing AI	37
Table 5.1	Specifications of laptop	58
Table 5.2	Specifications of mobile device	58
Table 6.1	Test Case and Expected Output	75

LIST OF ABBREVIATIONS

<i>ANN</i>	Artificial Neural Network
<i>ASRN</i>	Attention-based Sequence Recognition Network
<i>AR</i>	Accuracy Rate
<i>BLSM</i>	Bidirectional Long Short-term Memory
<i>BOW</i>	Bag-of-Words
<i>CC</i>	Connected Component
<i>CNN</i>	Convolutional Neural Network
<i>CRF</i>	Conditional Random Field
<i>CRNN</i>	Convolutional Recurrent Neural Network
<i>DCE</i>	Discrete Contour Evolution
<i>DD</i>	Dense Detector
<i>DT</i>	Decision Tree
<i>GMM</i>	Gaussian Mixture Model
<i>HMM</i>	Hidden Markov Model
<i>KFDA</i>	Kernel Fisher Discriminant Analysis
<i>k-NN</i>	k-Nearest Neighbour
<i>KPCA</i>	Kernel Principal Component Analysis
<i>LR</i>	Logistic Regression
<i>LSTM</i>	Long Short-term Memory
<i>LDA</i>	Linear Discriminant Analysis
<i>MLP</i>	Multilayer Perceptron
<i>MORAN</i>	Multi-object Rectified Attention Network
<i>MORN</i>	Multi-object Rectification Network
<i>OCR</i>	Optical Character Recognition
<i>OMR</i>	Optical Music Recognition
<i>RAD</i>	Rapid Application Development
<i>RD</i>	Random Detector
<i>RNN</i>	Recurrent Neural Network
<i>SLR</i>	Systematic Literature Review
<i>SMF</i>	Standard Median Filter

<i>SVM</i>	Support Vector Machine
<i>SWT</i>	Stroke Width Transform
<i>T2S</i>	Text To Speech

Chapter 1

Introduction

In this chapter, we present the background and motivation of our research, our contributions to the field, and the outline of the thesis.

1.1 Problem Statement and Motivation

The visually impaired community encounters substantial obstacles in accessing information and navigating physical environments independently. One critical aspect of daily life is identifying shops, understanding their offerings, and accessing contact information or other relevant details. Existing solutions often lack real-time functionality, accuracy, or comprehensive support, leaving blind individuals dependent on assistance from others. This limitation significantly impacts their autonomy, freedom of movement, and overall quality of life.

The motivation for developing the "ShopRecog: Mobile Application for Shop Recognition with Text to Speech" stems from a deep-seated commitment to addressing these challenges. By harnessing the power of modern technology and innovative design, this application aims to bridge the accessibility gap, empower visually impaired users, and enhance their overall shopping and navigation experiences.

1.2 Objectives

The primary objectives of the project are multi-faceted and revolve around creating a feature-rich mobile application tailored to the unique needs of blind individuals. These objectives include:

- **Real-Time Shop Recognition:** Develop a robust mechanism using the camera input to identify nearby shops accurately. This involves leveraging the Places API nearby search to fetch a comprehensive list of shop names within the user's vicinity.
- **Text Recognition and Matching:** Integrate MLKit's text recognition capabilities to analyze the live camera preview and identify text corresponding to shop names. Employ advanced algorithms to match recognized text with entries in the shop name list, ensuring a threshold of 70% text similarity for accurate detection.
- **Text-to-Speech Functionality:** Implement text-to-speech technology to audibly announce the detected shop names in real time. This auditory feedback enables blind users to receive immediate information about their surroundings, promoting independent navigation and decision-making.
- **Shop Details Retrieval:** Upon user interaction (such as tapping the screen), retrieve detailed information about the recognized shops using the Places API. This includes essential details like address, phone number, business hours, and user ratings, enhancing the user's understanding of each establishment.
- **AI-Generated Summaries:** Utilize the Gemini AI API to generate summary paragraphs about each shop based on the retrieved details. Enable language translation options (Chinese, Korean, Japanese, and English) to cater to diverse user preferences and international travel scenarios.
- **Interactive Speech-to-Text:** Implement a user-friendly interface with a hold-to-open microphone button for speech-to-text interaction. Allow users to ask questions about specific shops (e.g., phone number inquiry), send these queries to the Gemini API for processing, and receive spoken answers via text-to-speech conversion.
- **Navigation Assistance:** Enhance user mobility by integrating navigation features with Google Maps. Enable users to set shop destinations and initiate walking mode navigation directly from the application, providing step-by-step guidance to their desired locations.

1.3 Project Scope and Direction

The project's scope encompasses the entire development lifecycle of the "ShopRecog" mobile application, focusing on creating a seamless and intuitive user experience for visually impaired individuals. Key aspects within the scope include:

- **Android Platform:** The application targets Android devices, leveraging the platform's accessibility features and compatibility with Google services (e.g., Google Maps, MLKit, Places API) for optimal performance and integration.
- **Permissions and APIs:** Implement functionality requiring user permissions for location access, camera usage, and microphone input, adhering to best practices for privacy and security. Integrate Google's Places API for nearby shop search and detailed information retrieval.
- **Machine Learning Integration:** Utilize MLKit's text recognition capabilities to analyze real-time camera input and identify text relevant to shop names. Employ machine learning algorithms for text matching and accuracy enhancement.
- **AI-powered Language Processing:** Integrate the Gemini AI API for generating shop summary paragraphs and supporting language translation. Enable seamless language switching (Chinese, Korean, Japanese, English) based on user preferences and regional context.
- **User Interface and Interaction:** Design an intuitive and accessible user interface, including features like single-tap shop freezing, speech-to-text interaction, language selection, and navigation controls. Prioritize user feedback and accessibility standards throughout the interface design process.
- **Testing and Optimization:** Conduct thorough testing phases to ensure application functionality, usability, and accessibility across different device configurations and user scenarios. Optimize performance, responsiveness, and reliability through iterative development and feedback-driven improvements.

1.4 Contributions

The project's contributions extend beyond technical implementation to encompass broader impacts on accessibility, inclusivity, and user empowerment:

- **Accessibility Advancements:** "ShopRecog" represents a significant leap in providing real-time shop recognition and detailed information access for visually impaired users, fostering greater independence and inclusivity in daily activities.
- **Innovative Technology Integration:** By leveraging cutting-edge technologies such as MLKit, Google Maps, and AI-powered language processing, the application sets a precedent for integrating diverse functionalities seamlessly into an assistive tool.

- **User-Centric Design:** The application's user interface and interaction design prioritize user experience, feedback mechanisms, and accessibility guidelines, ensuring a user-friendly and intuitive experience for blind individuals.
- **Empowering Independence:** Through features like real-time shop detection, interactive speech-to-text, and navigation assistance, "ShopRecog" empowers visually impaired users to navigate and engage with their surroundings autonomously.
- **Potential for Expansion:** The modular architecture and scalable design of the application pave the way for future enhancements, feature additions, and integration with emerging technologies, further improving its utility and impact.

In summary, "ShopRecog" is not merely a mobile application but a transformative tool that empowers visually impaired individuals, promotes inclusivity, and demonstrates the potential of technology to enhance accessibility and quality of life.

1.5 Report Organization

This report is structured into seven cohesive chapters, each contributing to a comprehensive understanding of the project's development and outcomes.

In Chapter 1, the introduction sets the stage by presenting the problem statement, project background, and motivation that drove the development of the application. It outlines the scope of the project, defines clear objectives, highlights the unique contributions of the project, summarizes key achievements, and provides an overview of how the report is organized.

Chapter 2 delves into a thorough literature review, analyzing existing IoT applications available in the market. This review aims to evaluate the strengths and weaknesses of each product, providing valuable insights into industry trends, technological advancements, and user expectations.

The methodology and approach used in the development of the system are detailed in Chapter 3. This chapter discusses the overall strategy, methodologies, tools, and frameworks employed to design and implement the application. It provides a roadmap of how the project was conceptualized and executed.

CHAPTER 2

Chapter 4 focuses on the intricate system design of the application. It includes detailed diagrams, architectural plans, and technical specifications that illustrate the structural components, interactions, and workflows within the system. This chapter lays the foundation for understanding the application's functionality and design principles.

In Chapter 5, the actual implementation of the system is discussed. It covers the development process, coding practices, integration of APIs, testing methodologies, and any challenges faced during implementation. This chapter provides insights into the technical aspects of transforming design concepts into a functional application.

Chapter 6 is dedicated to evaluating the system's performance, functionality, and user experience. It includes comprehensive testing procedures, performance metrics, analysis of results, and discussions on system strengths, weaknesses, and improvements. This chapter also addresses any challenges encountered during development and how they were mitigated.

Finally, Chapter 7 serves as the conclusion of the report. It summarizes the key findings, achievements, contributions, and implications of the project. This chapter also discusses future recommendations, potential areas for further research, and concludes with reflections on the overall success and impact of the project.

Chapter 2

Literature Review

2.1 Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)

The research paper [2] offers a comprehensive overview of studies in character recognition of handwritten documents, aiming to guide future research in this area. The paper conducts a Systematic Literature Review (SLR) focusing on research articles published from 2000 to 2019 which are related to handwritten OCR. The search for relevant articles was using keywords, forward reference searching and backward reference searching to ensure comprehensive coverage. The review [2] effectively presents the latest advancements and techniques in OCR as well as highlighting the existing research gaps. There are many sections in this paper however this discussion will focus only on sections related to the Final Year Project (FYP) title.

The section V of this paper [2] provides an overview of the prevalent classification methods employed in handwritten OCR from 2000 to 2019. The first method is **Artificial Neural Networks (ANN)**. Artificial Neural Networks (ANN), inspired by biological neurons, have gained significant attention due to their capacity to model complex relationships. Multi-Layer Perceptrons (MLPs), a type of feedforward network, have demonstrated remarkable results in character recognition tasks. The architecture of MLPs involves input, hidden, and output layers as show in Figure 2.1. By adjusting weights associated with neurons through supervised learning, MLPs can effectively classify characters. The evolution of neural architectures, including Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), has established neural networks as a leading classification technique in OCR. These architectures have been extensively used across various languages and recognition tasks, ensuring their relevance and effectiveness in OCR research.

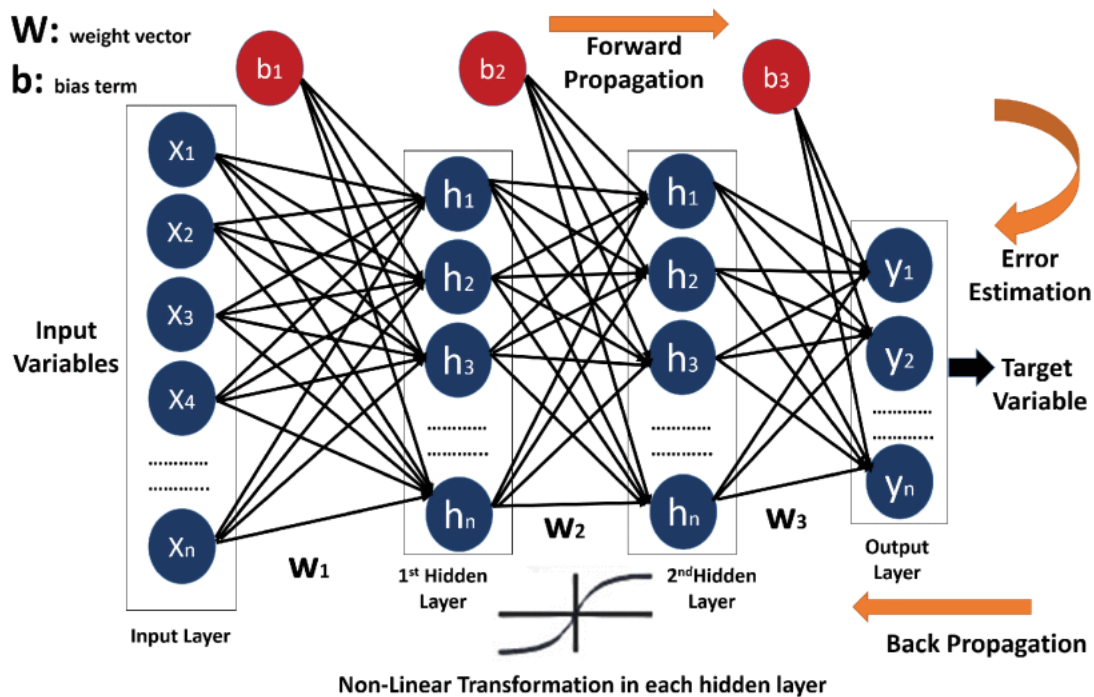


Figure 2.1: An architecture of Multilayer Perceptron (MLP)

Secondly, **kernel methods** such as Support Vector Machines (SVMs), Kernel Fisher Discriminant Analysis (KFDA) and Kernel Principal Component Analysis (KPCA), offer robust solutions to classification problems. SVMs employ kernel functions to map input data into higher-dimensional spaces, effectively separating classes with a maximum margin hyperplane. SVMs have been recognized for their efficiency in handwritten digit recognition, image and text classification, as well as object and face detection. KFDA and KPCA are also instrumental in offline handwritten character recognition, with SVMs often preferred due to their classification accuracy.

Third, **statistical classifiers** encompass both parametric and non-parametric approaches. Parametric classifiers, like Logistic Regression (LR) and Linear Discriminant Analysis (LDA), assume a fixed number of parameters, while non-parametric methods, such as k Nearest Neighbor (k NN) and Decision Trees (DT), offer flexibility in learning concepts but grow in complexity with input data size. The k-NN algorithm, a non-parametric statistical model, has been widely used in OCR due to its simplicity and good performance. Parametric methods, exemplified by Hidden Markov Models (HMMs), were prevalent in the early 2000s.

HMMs which is capable of capturing sequential dependencies has contributed to speech and optical character recognition particularly when lexicon availability was limited [36].

Fourth, **template matching techniques**. Template matching techniques involve matching predefined templates to small image portions using sliding window approaches. Taxonomy of template matching techniques is show in Figure 2.2. Deformable template matching which is a sub-class of template matching, is particularly useful for character recognition as it can accommodate variations caused by different writers. In rigid template matching, shape deformations are not considered but it works with feature extraction.

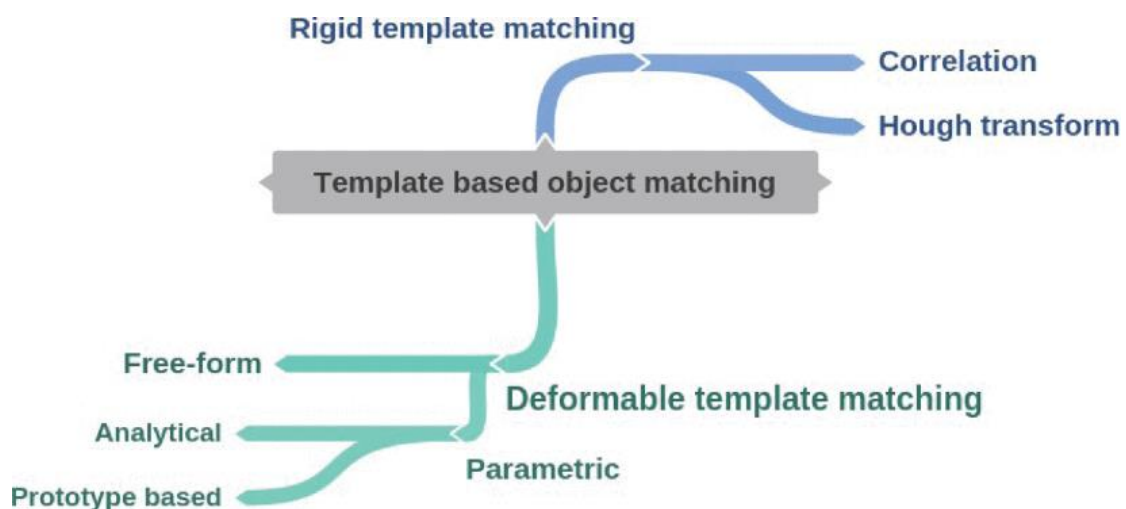


Figure 2.2: An overview of template matching techniques

Fifth, **structural pattern recognition methods** focus on the relationship between pattern structures, utilizing pattern primitives such as edges and contours. Graphical methods involve representing characters using nodes and edges with similarity measures determining classification. Grammar-based methods conduct syntax analysis to find similarities in structural graph primitives using the concept of grammar. Strings and trees are used to represent models based on grammar, aiding in robust character recognition. Some examples are shown in Figure 2.3.

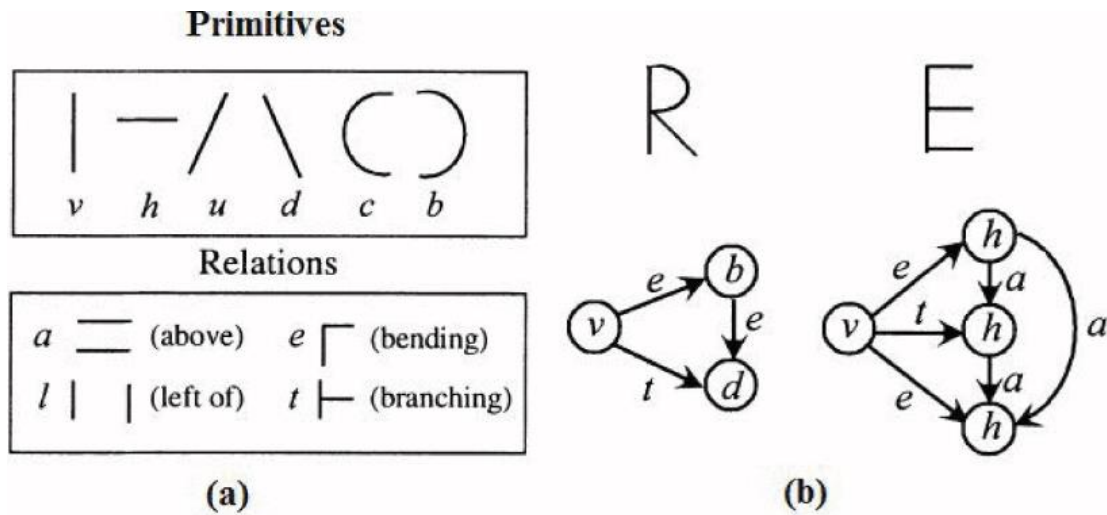


Figure 2.3: (a) Primitive and relations (b) Directed graph for capital letter R and E

In conclusion, this review [2] systematically analyzes research publications across six widely spoken languages and observes variations in the performance of techniques across different scripts. For example, the multilayer perceptron classifier demonstrated better accuracy for Devanagari and Bangla numerals but yielded average results for other languages, potentially due to modeling differences and dataset quality.

Furthermore, the review notes that many research studies focus on specific languages or subsets, often lacking diversity in writing styles, distorted strokes, variable character thickness, and illumination, which are common in real-life scenarios. Additionally, there is a growing trend in the use of Convolutional Neural Networks (CNNs) for character recognition, initially designed for object recognition tasks in images.

The paper identifies several promising directions for future OCR research. One crucial area is the exploration of languages beyond the widely spoken ones, including regional and endangered languages. This not only preserves cultural heritage but also contributes to global collaboration.

Another important challenge is the development of systems capable of recognizing on-screen characters and text in various daily life scenarios, such as captions, news tickers, signboards, and billboards. This "text in the wild" domain presents complex challenges like background clutter, variable lighting conditions, different camera angles, distorted characters, and variable writing styles.

CHAPTER 2

To tackle these challenges effectively, researchers should create comprehensive datasets that encompass all possible character variations. Initiatives like the "ICDAR 2019 Robustreading challenge on multilingual scene text detection and recognition" are encouraging advancements in this direction, with winning methods based on deep learning architectures like CNNs and RNNs.

However, the increased use of complex deep learning architectures, while improving classification accuracy, also escalates computational complexity, posing challenges for real-time and robust character recognition systems.

Lastly, the review emphasizes the need for commercializing OCR research to develop cost-effective, real-life systems that can convert vast amounts of invaluable information into searchable digital data, expanding the practical applications of OCR technology.

2.1.1 Strengths

This paper provides a detailed analysis of prevalent classification methods used in handwritten OCR such as Artificial Neural Networks (ANNs), kernel methods, statistical classifiers, template matching techniques and structural recognition methods. This analysis offers valuable insights into the various techniques used in OCR. Additionally, this article identifies promising directions for future OCR research such as exploring regional and endangered languages and addressing the challenges of recognizing text in real-life scenarios.

2.1.2 Weaknesses

Only research articles up to 2019 were studied in this paper [2]. Thus, it may not capture the latest developments in the field. Another weakness is it does not provide a comparative evaluation for the discussed classification methods. Readers do not know when to use one method over another and how the results are using each method.

2.1.3 Recommendation

Since new technology is invented from time to time, recent research findings and advancements in handwritten OCR beyond 2019 should also be included and studied. Next,

CHAPTER 2

conducting a comparative analysis of the classification methods discussed enables the readers to make appropriate choices when implementing OCR solutions.

2.2 Character detection and recognition system for visually impaired people

The research paper [3] under consideration presents a system comprising three principal stages: Acquisition, Processing, and Text to Speech conversion (T2S). The general block diagram of the system is shown in Figure 2.4 below.

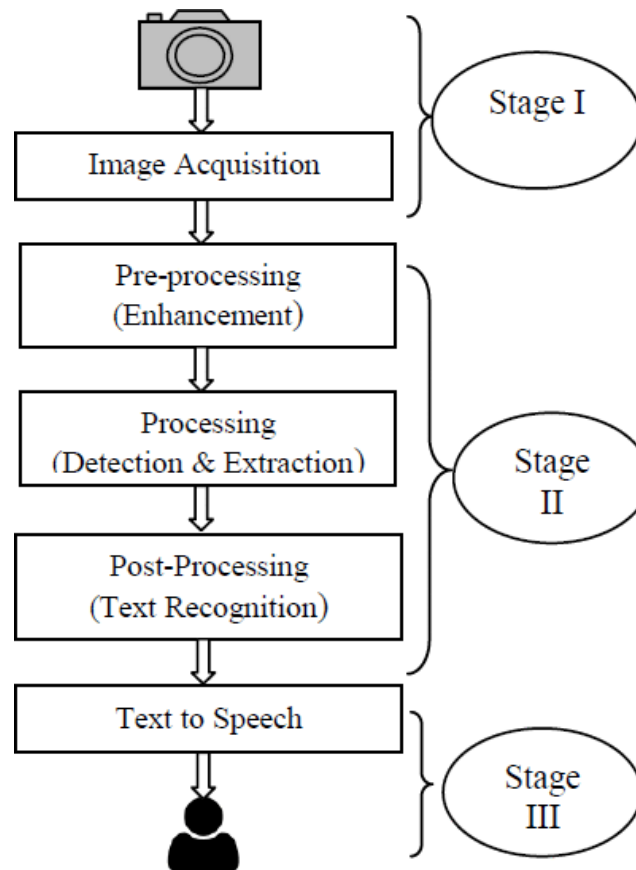


Figure 2.4: General block diagram of system

During the acquisition phase, the system receives a high-resolution video input from a camera. This video is subsequently divided into distinct frames, each functioning as an independent image. Notable challenges encountered within this stage include issues with camera positioning, blurring arising from user motion, and perspective distortion.

The image processing stage encompasses three sub-stages: pre-processing, processing, and post-processing. In the initial step, the acquired colour image is transformed into grayscale. Subsequently, enhancement techniques are applied to mitigate noise, uneven lighting, and

blurring artifacts. To address noise, a Standard Median Filter (SMF), known for its denoising efficacy and computational efficiency, is recommended. For contrast enhancement, the Histogram Equalization method is employed. To combat blurring, the paper suggests employing de-blurring methods such as the Lucy Richardson algorithm, Blind de-convolution algorithm, and Wiener de-blurring techniques (as show in Figure 2.5).

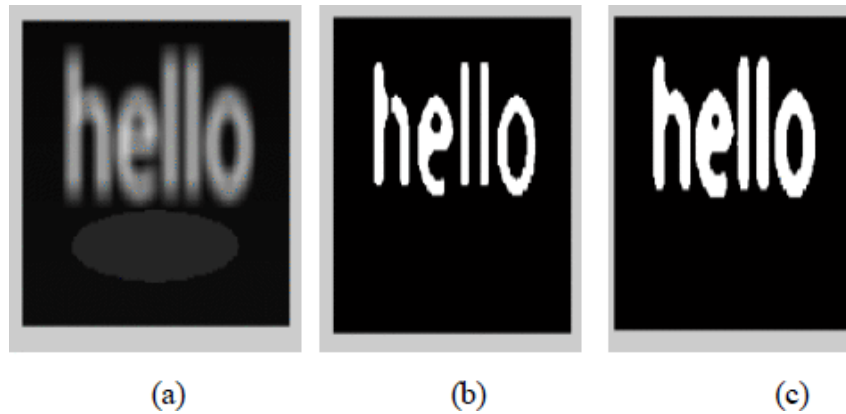


Figure 2.5: De-blurring of an image (a) Blurred image; (b) Binarized image without filtering; (c) Binarization after De-blurring

In the processing sub-stage, the image is binarized through adaptive thresholding. A combination of connected component (CC) analysis and a region-based approach is applied to the binarized image. Areas exhibiting text-like patterns, characterized by white pixels against a dark background, are identified using CC analysis implemented with MATLAB software. These identified areas are subsequently extracted into separate windows using a feature extraction algorithm, as illustrated in Figure 2.6.

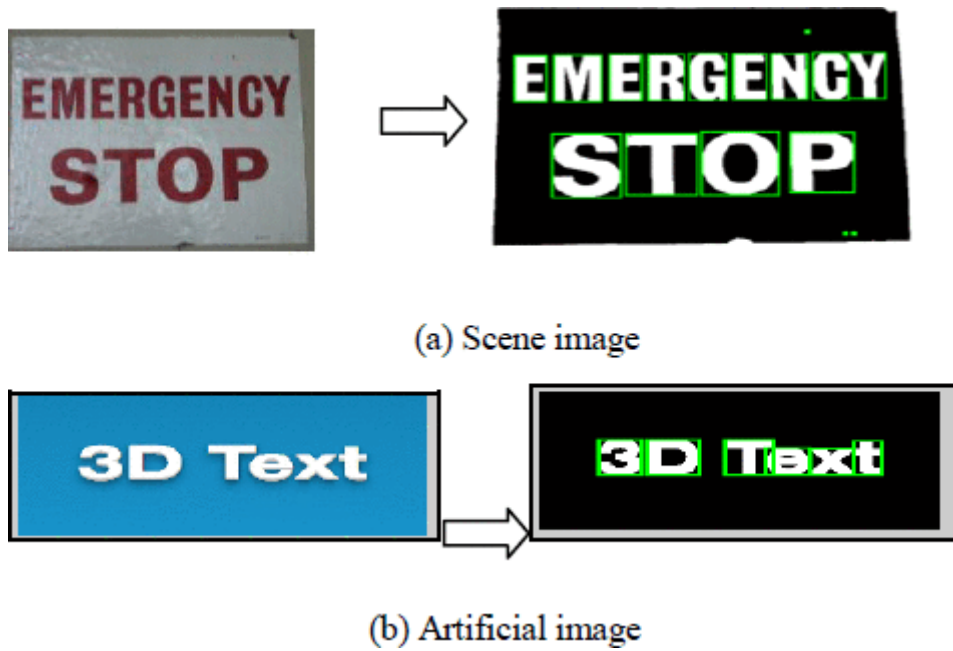


Figure 2.6: Text detection and extraction

The post-processing stage involves text recognition based on the outcomes of the preceding stages. Approaches such as Linear Discriminate Analysis (LDA), Support Vector Machine (SVM), Conditional Random Field (CRF), and Stroke Width Transform (SWT) have been developed for image text recognition. The paper's author employed Feature Learning to attain superior results in comparison to other recognition models. The output is shown in Figure 2.8.

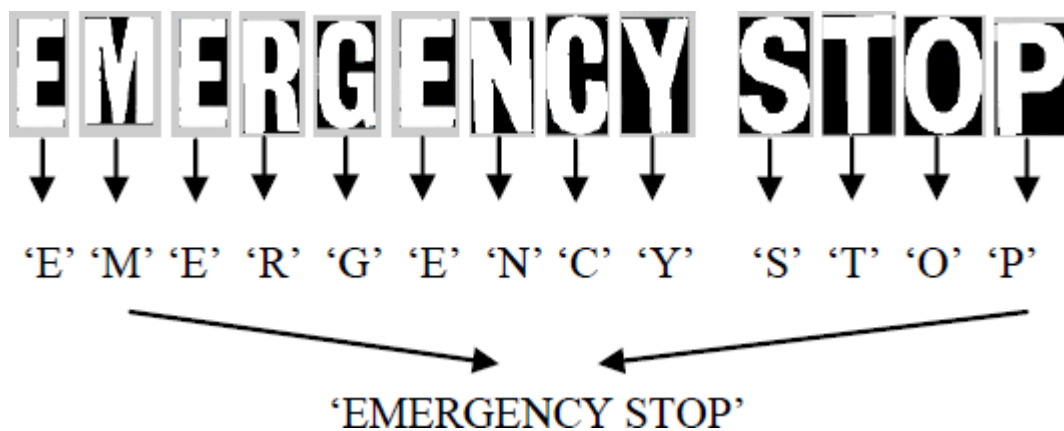


Figure 2.7: Text recognition system output

Text-to-Speech Conversion (T2S) serves the purpose of converting the recognized text into voice, thereby enabling individuals with visual impairments to perceive scene text through auditory means. To accomplish this, the LabVIEW software employs an in-built speech synthesizer function. The utilization of Speech Synthesizer Nodes, including Property and Invoke, derived from .NET objects within LabVIEW, facilitates the T2S process.

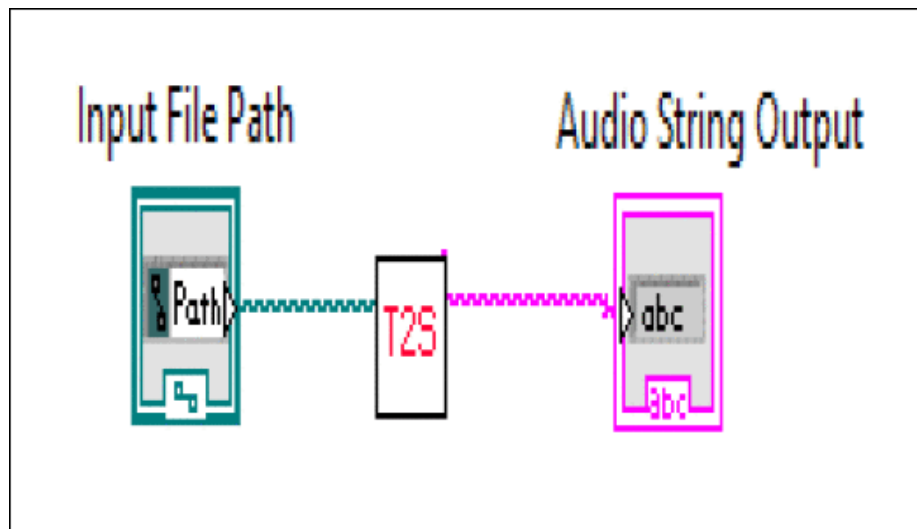


Figure 2.8: Block diagram of T2S converter

In conclusion, this paper [3] integrated several techniques for text detection and extraction to achieve better result than using single techniques for overall system. Text detection followed by recognition using supervised pattern recognition algorithm increases the speed of the system as well as improving accuracy. The text is converted into audio output after successful recognition.

2.2.1 Strengths

The strength of this paper is that it acknowledges and addresses real-world challenges encountered in the acquisition phase such as camera positioning, motion blur and perspective distortion. This demonstrates a practical understanding of the issues faced in implementing such a system. Various techniques are discussed and implemented in detail during the

acquisition, processing and text to speech conversion stage and the effort to improve accuracy is highly appreciated.

2.2.2 Weaknesses

The text-to-speech conversion process is mentioned briefly. A more in-depth explanation of this critical component would have been beneficial. While this paper mentions the use of Feature Learning for text recognition, it does not provide a detailed comparative evaluation of this approach against other recognition models. A comparative analysis could have strengthened the argument for its superiority.

2.2.3 Recommendation

This paper can elaborate more on the text-to-speech conversion process to enhance reader's understanding on implementing this crucial component. A comprehensive comparative analysis of different text recognition techniques should be conducted to provide a clear understanding of the advantages and disadvantages of each approach.

2.3 An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition

The article [4] initiates by introducing the CRNN (Convolutional Recurrent Neural Network), an innovative neural network architecture designed to tackle image-based sequence recognition tasks. It emphasizes the significance of CRNN in the context of both text and music recognition which are the famous domains known for their inherent complexity and challenges in computational analysis.

The authors delve into the core components of the CRNN (Convolutional Recurrent Neural Network) architecture as shown in Figure 2.9 below. This neural network design represents a novel approach to image-based sequence recognition, bringing together Convolutional Layers, Recurrent Layers and a Transcription Layer to tackle complex tasks in a unified framework.

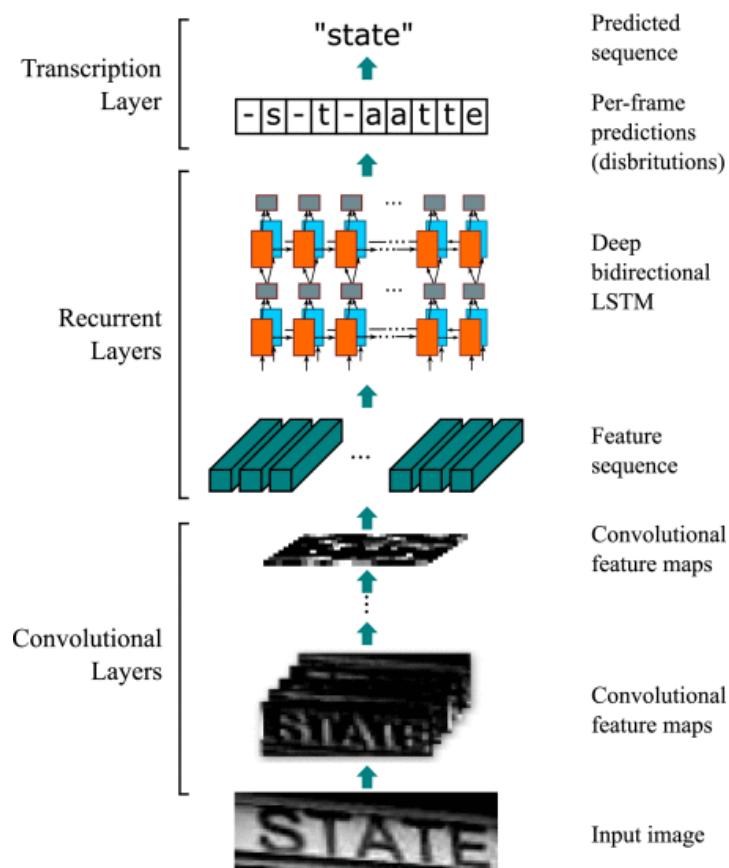


Figure 2.9: The network architecture

CHAPTER 2

The Convolutional Layers, constituting the initial segment of CRNN, play a pivotal role in enabling the network to effectively process and comprehend visual information from input images. These layers act as filters to extract relevant features and patterns from the images, thus facilitating a deeper understanding of the content under analysis. In essence, Convolutional Layers are the foundation upon which CRNN builds its understanding of the visual world.

Recurrent layers are the layers complementing the Convolutional Layers. They represent the brainpower behind the sequence prediction capabilities of CRNN. This article [4] implement a specialized type of neural network called LSTM (Long Short-Term Memory) units within these layers. LSTMs are well-suited for sequential data as they possess the unique ability to retain and utilize information from previous steps in the sequence. This memory component is instrumental in predicting subsequent elements within a sequence, such as characters in a word or musical notes in a score. In essence, the Recurrent Layers provide CRNN the contextual awareness crucial for accurate predictions.

The Transcription Layer, serving as the final piece of the puzzle, takes the network's predictions which are normally in the form of sequences then transforms them into the ultimate output. These predictions typically comprise sequences of labels such as words in text recognition or musical notes in score recognition. The choice of employing LSTM units within this layer imparts a crucial advantage – the network can make informed and contextually relevant predictions that take into account the broader context of the sequence. This contextual understanding is a critical factor in CRNN's impressive performance.

Some comprehensive experiments were conducted to evaluate the performance and versatility of CRNN. Two challenging tasks are chosen: scene text recognition and musical score recognition. These tasks serve as robust benchmarks to assess CRNN's capabilities.

For scene text recognition, the authors employ the Synth dataset, a synthetic dataset with 8 million training images and corresponding ground truth words. CRNN demonstrates remarkable adaptability by working well on real-world test datasets without fine-tuning on their training data. Four established benchmarks, including ICDAR 2003, ICDAR 2013, IIIT 5k-word, and Street View Text, are used to evaluate CRNN's performance.

In the realm of musical score recognition, the Optical Music Recognition (OMR) problem is tackled. The authors emphasize that OMR has historically involved multiple

preprocessing steps, including binarization, staff line detection, and individual note recognition. CRNN takes a different approach by framing OMR as a sequence recognition problem, directly predicting sequences of musical notes from images. Although there is a lack of public datasets specifically for pitch recognition, manually labelling ground truth sequences for a substantial number of images can address this issue. They then evaluate CRNN's performance against two commercial OMR engines, Capella Scan and PhotoScore.

In the concluding section, the authors summarize the key findings and contributions of their work. They emphasize that CRNN represents a novel and versatile framework for image-based sequence recognition, capitalizing on the combined strengths of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). The capacity of CRNN to handle inputs of varying dimensions and produce variable-length predictions stands out as a major advantage.

The experiments conducted on scene text recognition benchmarks demonstrate CRNN's superior performance compared to conventional methods and other deep learning approaches. CRNN's ability to perform exceptionally well even when trained purely on synthetic data is a noteworthy achievement.

In the domain of musical score recognition, CRNN outperforms two commercial OMR systems by a significant margin. Its robustness to noise and contextual understanding of musical scores are highlighted as key factors contributing to its success in this task.

The authors also emphasize the general applicability of CRNN to other domains and tasks that involve sequence prediction in images, such as Chinese character recognition. They express a commitment to further refining CRNN to enhance its speed and practicality for real-world applications.

2.3.1 Strengths

This paper [4] introduces the CRNN (Convolutional Recurrent Neural Network) architecture which is a novel approach to image-based sequence recognition. The integration of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) in CRNN allows the system to handle inputs of varying dimensions and produce variable-length predictions. CRNN demonstrates versatility and adaptability by successfully handling two challenging tasks: scene text recognition and musical score recognition.

2.3.2 Weaknesses

The paper [4] provides a high-level overview of the CRNN architecture and its applications but lacks in-depth technical implementation details. Readers interested in replicating or understanding the architecture at a deeper level may feel insufficient. While the article mentions the successful performance of CRNN, it does not delve deeply into the specific challenges faced during implementation or potential limitations of the architecture. A more thorough discussion of challenges and limitations would provide a more balanced perspective.

2.3.3 Recommendation

The authors can provide a more detailed technical documentation of the CRNN architecture, including network configurations, hyperparameters, and training methodologies. This would enable readers to better understand and implement CRNN in their own projects. It is also pivotal to discuss in greater detail the challenges encountered during the implementation of CRNN and the potential limitations of this architecture. Understanding these aspects is important for readers to determine whether to deploy this architecture.

2.4 Scene Text Recognition in Mobile Applications by Character Descriptor and Structure Configuration

Scene text extraction is divided into two processes: text detection and text recognition. Text detection aims to locate text regions in images and remove non-text elements. Text recognition transforms pixel-based text into readable code. This paper [5] focuses on text recognition, dealing with 62 identity categories of text characters, including digits and English letters in both upper and lower case.

The proposed approach combines text detection and recognition algorithms. In the text detection process, pixel-based layout analysis is used to extract text regions based on color uniformity and horizontal alignment. In text recognition, two schemes are introduced: character recognition and binary character classification for text understanding and retrieval.

The paper's [5] main contributions are twofold. First, a character descriptor is introduced to extract features from character patches using feature detectors like Harris-Corner, MSER, dense sampling, and HOG descriptors. Second, a novel concept called stroke configuration is proposed to model character structure for binary classification in text retrieval.

The method is illustrated in a flowchart, demonstrating how text understanding and retrieval can be facilitated using character recognition and classification. The proposed feature representation combines low-level descriptors with stroke configuration to model text character structure. The concepts of text understanding and retrieval are introduced and evaluated through experiments.

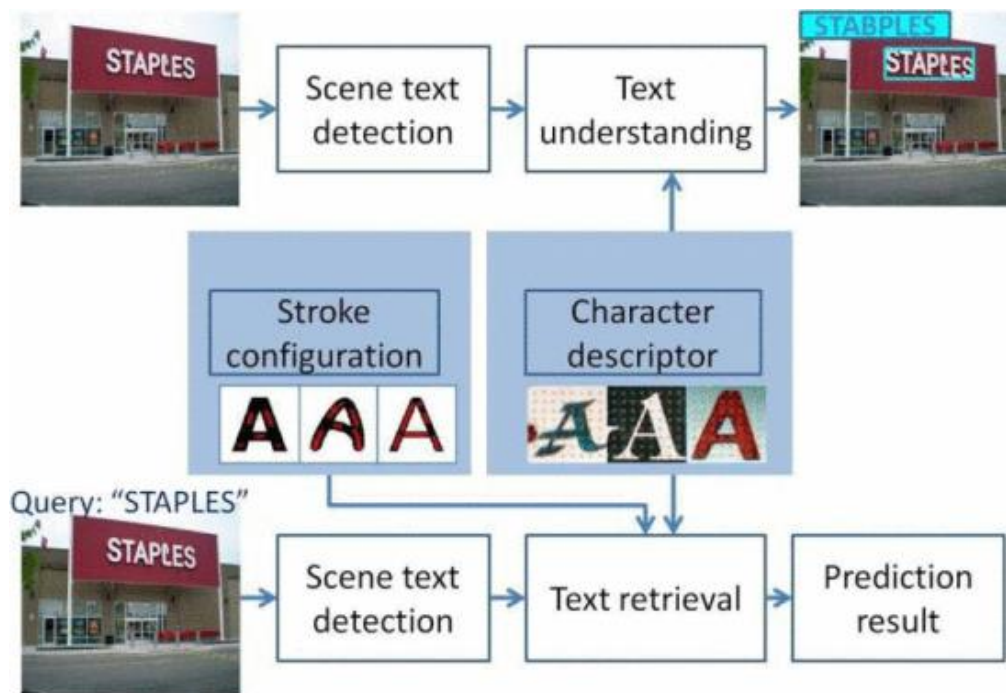


Figure 2.10: The flowchart of the Designed Scene Text Extraction Method

The next section focuses on layout-based scene text detection, which is essential for extracting text regions from scene images. It describes the improvements made to the text detection process to make it compatible with mobile applications.

First, the paper [5] talks about Layout Analysis of Color Decomposition by introducing a boundary clustering algorithm based on bigram color uniformity. This algorithm decomposes a scene image into color-based layers, effectively separating text from background outliers with different colors. Color difference is used to identify character boundaries.

Second, Layout Analysis of Horizontal Alignment. The paper [5] proposes an adjacent character grouping algorithm that identifies image regions containing text strings. It uses bounding boxes to group consecutive neighboring bounding boxes of similar size and horizontal alignment. This process effectively identifies text string fragments. The algorithm is also adaptive to slightly non-horizontal text orientations and can handle font variations.

Some technical adjustments were made to the scene text detection algorithm to ensure compatibility with a blind-assistant demo system. These adjustments include down-sampling input images, adopting edge pixels based on specific geometrical constraints, and fine-tuning parameters related to horizontal similarity and alignment.

Bachelor of Computer Science (Honours)

Faculty of Information and Communication Technology (Kampar Campus), UTAR

Next, the paper [5] delves into the character recognition aspect of scene text extraction, with a focus on character descriptor design and two distinct recognition schemes. The scene text characters mentioned in this paper [5] include 10 digits and 26 English letters in both upper and lower case, totaling 62-character classes. These classes are used for training character recognition. This paper [5] outlines two-character recognition schemes: text understanding and text retrieval. In text understanding scheme, character recognition is treated as a multi-class classification problem. A character recognizer is trained to classify characters into one of the 62 classes. This scheme is designed for text understanding applications. In the text retrieval scheme, character recognition is framed as a binary classification problem. For each of the 62-character classes, a binary classifier is trained. These classifiers are used to distinguish whether a character patch belongs to a specific character class or not. This scheme is suitable for text retrieval applications.

The paper [5] also introduces a novel character descriptor as show in Figure 2.11 below to extract structural features from character patches effectively. The descriptor is designed to be robust and discriminative. It combines information from four keypoint detectors: Harris detector, MSER detector, Dense detector, and Random detector. HOG features are extracted at these keypoints. The descriptor also incorporates the Bag-of-Words (BOW) model and Gaussian Mixture Model (GMM) to aggregate extracted features.

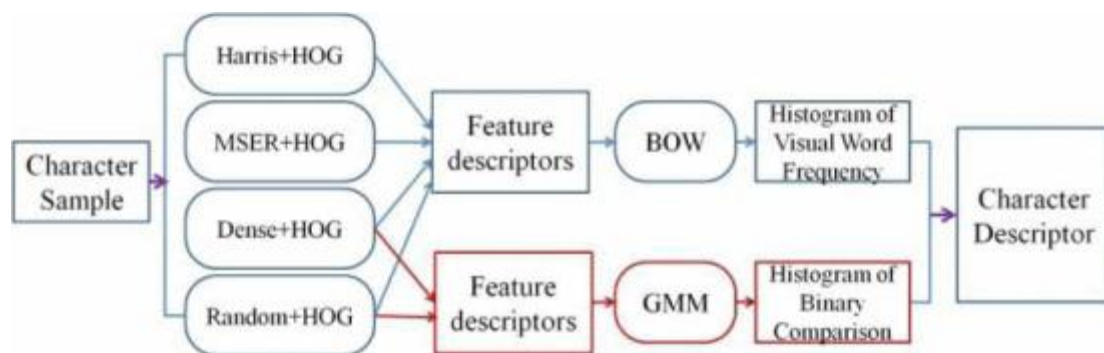


Figure 2.11: Flowchart of the Character Descriptor

BOW is applied to keypoints from all four detectors. It represents character patches as histograms of visual words, which are derived from clustering HOG features. The BOW

representation is efficient and resists intra-class variations. The paper [5] mentions that it employs soft-assignment coding and average pooling for BOW-based feature representation.

GMM is applied to keypoints from Dense detector (DD) and Random detector (RD). These keypoints are generated uniformly or randomly in character patches. GMM is used to describe the local feature distributions. Each GMM contains 8 Gaussian distributions. The paper [5] describes the process of building GMM using an EM algorithm.

For each character patch, both BOW-based and GMM-based feature representations are generated. These two feature representations are then concatenated into a character descriptor, creating a feature vector for the character patch.

The next section discusses character stroke configuration, which is crucial for character recognition in text retrieval applications. Stroke configuration represents the basic structural elements of characters, such as their orientation and width. In this paper [5], stroke configurations are estimated using synthesized characters generated by computer software rather than real-world scene images.

A character's boundary and skeleton are extracted using discrete contour evolution (DCE) and skeleton pruning. This process simplifies characters into polygons and defines the polygon as the character's boundary. The skeleton is then pruned for refinement. Then, a set of evenly sampled points is chosen along the character's boundary, including the polygon vertices. Stroke width and orientation are estimated at each sampled point. Stroke width is determined by probing length along the normal vector at each point until another boundary point is encountered. Stroke configurations are derived from the consistency of stroke width and orientation at sampled points. Sample points that satisfy the stroke-related features create the stroke sections of the character boundary, while others form junction sections. Skeleton points are extracted from stroke sections to construct the stroke configuration.

To handle variations in fonts, styles, and sizes, stroke configurations are aligned to calculate a mean value of stroke configuration for each character class. This alignment is achieved using an objective function that minimizes the distance between stroke configurations while considering transformations like translation and scaling. Stroke configurations are divided into partitions, each containing neighboring stroke components. These partitions create a stroke configuration map for each character class, providing a comprehensive view of character structure. For character recognition in text retrieval, character patches are partitioned

into sub-patches based on the stroke configuration map. Structural features are extracted from each sub-patch, and these features are combined into a feature vector for the character patch. In text retrieval, where queried character classes are sought, Adaboost learning with a cascade is employed to handle imbalanced data. Character classifiers are trained for each character class to confirm the presence of queried characters in text retrieval app.

Quantitative experimental analyses were conducted to the their character recognition system on three public datasets: Chars74K EnglishImg Dataset, Sign Dataset, and ICDAR-2003 Robust Reading Dataset.

On the Chars74K Dataset, the proposed character descriptor, which combines BOW-based and GMM-based feature representations, outperformed previous algorithms in text understanding. Specifically, it achieved a high accuracy rate (AR) in recognizing text characters, demonstrating its effectiveness in this dataset.

In the Sign Dataset, which primarily consists of regular fonts and styles, the character recognition system faced challenges due to imbalanced character samples. This resulted in some categories having an AR of 0. While the system may not perform uniformly across all categories, it provides insights into the dataset's limitations.

The ICDAR-2003 Dataset posed challenges with non-text background outliers and low-resolution character samples. Despite these difficulties, the character recognition system exhibited varying ARs, with categories containing more samples generally achieving higher ARs. This highlights the system's adaptability to different types of data and its ability to provide valuable insights even in challenging scenarios.

Future work aims to enhance text detection accuracy and extend the system to word-level recognition with lexicon analysis. Improving text structure modeling by designing more representative features is also planned. A specific scene text word database will be collected for a stronger training set. Additionally, the integration of scene text extraction with other techniques, such as content-based image retrieval, is envisioned to create a more versatile vision-based assistant system.

2.4.1 Strengths

The introduction of a character descriptor that combines features from multiple keypoint detectors and uses Bag-of-Words (BOW) and Gaussian Mixture Model (GMM) for feature aggregation is a notable contribution. This descriptor enhances the effectiveness of character recognition. The concept of stroke configuration to model character structure is valuable for text retrieval applications. It provides insights into the structural elements of characters, such as orientation and width. Quantitative experiments conducted on multiple public datasets demonstrate the effectiveness of the proposed character descriptor and recognition schemes. It provides clear performance metrics and insights into the system's adaptability to different datasets.

2.4.2 Weaknesses

While this article discusses character recognition, stroke configuration, and text detection extensively, it lacks real-world examples or case studies demonstrating the practical application of the proposed methods in mobile applications. Although adjustments for compatibility with a blind-assistant demo system are mentioned, it could provide more information on the specific challenges and solutions related to deploying the system on mobile devices.

2.4.3 Recommendation

Include real-world examples or case studies that showcase the practical application of the proposed scene text extraction method in mobile applications will help to demonstrate its utility in real scenarios and strengthen the article's relevance. It is also important to provide more details on the specific challenges faced in making the system compatible with mobile applications.

2.5 MORAN: A Multi-Object Rectified Attention Network for scene text recognition

In this research paper [6], the authors address the challenging problem of recognizing irregular scene text, which includes text with various shapes, perspectives, and distortions. They propose a framework called MORAN (Multi-Object Rectified Attention Network) to tackle this issue effectively. MORAN consists of two key components: the Multi-Object Rectification Network (MORN) for image rectification and an Attention-Based Sequence Recognition Network (ASRN) for text recognition. Figure 2.12 below shows the overall structure of the MORAN.

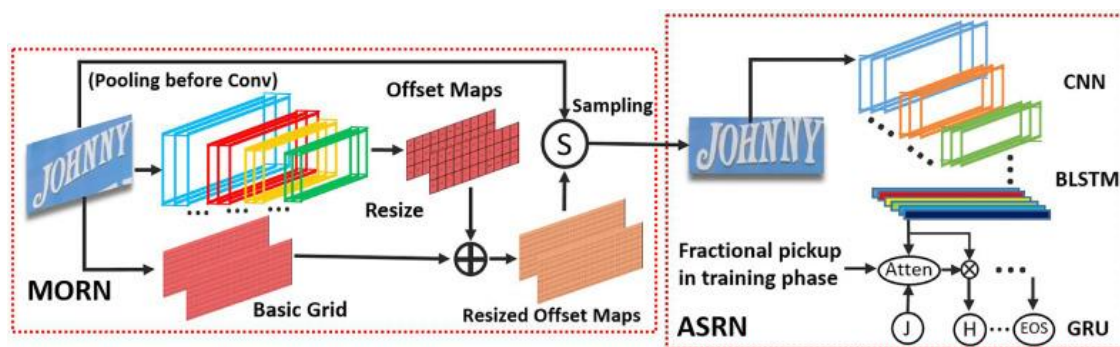


Figure 2.12: Overall structure of the MORAN

The MORN is developed to rectify images containing irregular text without geometric constraints. Traditional methods like affine transformation fail to handle complex deformations. The MORN predicts offsets for different parts of the image, and these offsets are used to rectify the image, making it easier to recognize. Notably, the MORN focuses on position offsets rather than character categories, reducing noise and calculation complexity. It utilizes a unique grid and bilinear interpolation to achieve smooth rectification. Importantly, the MORN can be trained with weak supervision, using text labels without requiring pixel-level deformation information.



Figure 2.13: Results of the MORN on challenging image text

The advantages of the MORN include its ability to make text more readable, its flexibility in handling various deformations, its scalability with different image sizes, and its suitability for weak supervision training. Examples demonstrate the effectiveness of MORN in rectifying slanted, perspective, and curved text, eliminating noise and improving readability.

The ASRN is primarily composed of a CNN (Convolutional Neural Network) followed by a BLSTM (Bidirectional Long Short-Term Memory) framework. An essential component is the one-dimensional attention mechanism placed atop the CRNN (Convolutional Recurrent Neural Network). This attention-based decoder plays a vital role in accurately aligning the target text and the predicted labels.

The attention mechanism is crucial for determining the alignment between the target text and the image features. It utilizes attention weights $\alpha_{t,i}$, which are calculated based on

learnable parameters, to decide which parts of the feature maps are relevant for prediction. This mechanism enables the decoder to focus on specific regions of the image.

The decoder can produce predicted words in a lexicon-free manner, meaning it generates the word without constraints. However, if lexicons are available, the ASRN evaluates probability distributions for all possible words and selects the word with the highest probability as the final result.

The architecture of the ASRN is described in Table 2.1, outlining the configuration of layers, including convolutional layers with batch normalization and ReLU activation functions.

Table 2.1: Architecture of ASRN

Type	Configurations	Size
Input	–	$1 \times 32 \times 100$
Convolution	maps:64, k3x3, s1x1, p1x1	$64 \times 32 \times 100$
MaxPooling	k2x2, s2x2	$64 \times 16 \times 50$
Convolution	maps:128, k3x3, s1x1, p1x1	$128 \times 16 \times 50$
MaxPooling	k2x2, s2x2	$128 \times 8 \times 25$
Convolution	maps:256, k3x3, s1x1, p1x1	$256 \times 8 \times 25$
Convolution	maps:256, k3x3, s1x1, p1x1	$256 \times 8 \times 25$
MaxPooling	k2x2, s2x1, p0x1	$256 \times 4 \times 26$
Convolution	maps:512, k3x3, s1x1, p1x1	$512 \times 4 \times 26$
Convolution	maps:512, k3x3, s1x1, p1x1	$512 \times 4 \times 26$
MaxPooling	k2x2, s2x1, p0x1	$512 \times 2 \times 27$
Convolution	maps:512, k2x2, s1x1	$512 \times 1 \times 26$
BLSTM	hidden unit:256	$256 \times 1 \times 26$
BLSTM	hidden unit:256	$256 \times 1 \times 26$
GRU	hidden unit:256	$256 \times 1 \times 26$

Here, k, s, p are kernel, stride and padding sizes, respectively. "BN" and "BLSTM" stand for batch normalization and bidirectional-LSTM respectively. "GRU" is in attention-based decoder.

The authors introduce a training method called "fractional pickup" for enhancing the performance and robustness of the Attention-Based Sequence Recognition Network (ASRN)

within the MORAN framework. The primary goal of fractional pickup is to improve the ASRN's ability to focus on the relevant regions of an image, especially in the presence of various types of noise and challenging scenarios.

The motivation for fractional pickup arises from the fact that scene text recognition often involves complex backgrounds and the risk of the decoder focusing on the wrong regions. This can lead to failed predictions. To address these challenges, fractional pickup is proposed as a means to enable the ASRN to perceive adjacent characters and widen its field of attention. Some challenging samples for recognition are presented in Figure 2.13.

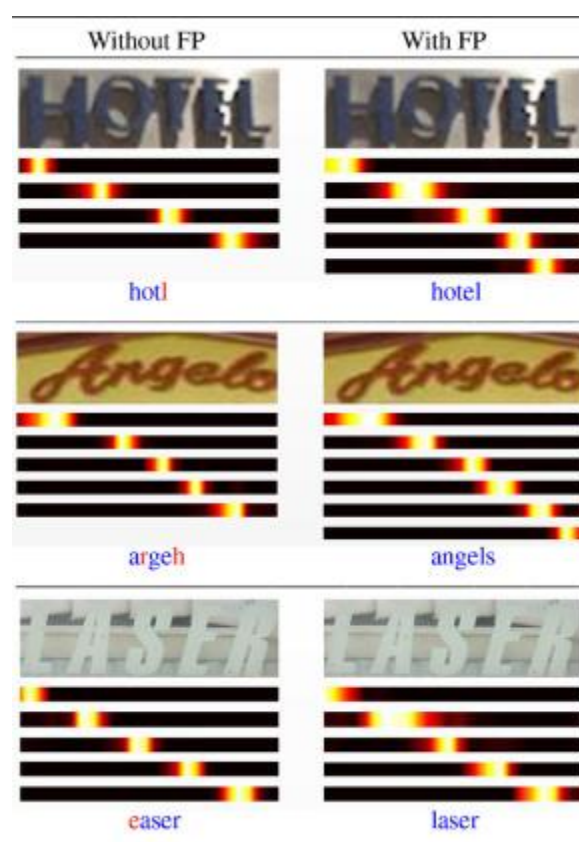


Figure 2.14: Difference in α_t for training with and without fractional pickup

Fractional pickup operates at each time step of the decoder and involves selecting and modifying a pair of attention weights, namely $\alpha_{t,k}$ and β , in a fractional manner. These values are generated randomly and independently at each time step, adding a degree of randomness to the decoder's behavior during training. This randomness leads to variations in the distribution of attention weights α_t , contributing to the robustness of the decoder.

One key aspect of fractional pickup is the creation of shortcuts in the training process. These shortcuts connect the current time step to a previous one (step k). This connection retains some features from the previous step and introduces interference to the forget gate in the bidirectional-LSTM, thereby providing additional information about the previous step and increasing the robustness of the ASRN.

The broader visual field achieved through fractional pickup helps improve the ASRN's performance. By disturbing the decoder's behavior through the local variation of $\alpha_{t,k}$ and β , the training process allows back-propagated gradients to optimize the decoder over a broader range of neighboring regions. This enhances the ASRN's ability to correctly predict target characters in challenging scenarios.

A curriculum learning strategy was proposed to efficiently train its two key components: the Multi-Object Rectification Network (MORN) and the Attention-Based Sequence Recognition Network (ASRN). The rationale behind this strategy is that end-to-end training can be time-consuming and inefficient due to the potential hindrance of one network by the other during training.

The curriculum learning strategy comprises three stages. In the first stage, the ASRN is trained using regular samples, which have tightly bounded annotations, providing a strong initial foundation. Then, the training set is expanded to include irregular samples obtained by cropping text using minimum circumscribed horizontal rectangles. This diversifies the training data and further enhances the ASRN's accuracy, especially in recognizing irregular text.

In the second stage, the ASRN trained on regular samples is employed to guide the training of the MORN. By fixing the ASRN's parameters and stacking it after the MORN, the ASRN can assess whether the transformations applied by the MORN indeed reduce the difficulty of recognition. If they do, meaningful gradients are provided to the MORN for optimization.

Finally, in the third stage, both the MORN and ASRN are optimized in an end-to-end fashion. This joint training allows the MORAN framework to complete its end-to-end optimization, ultimately surpassing state-of-the-art methods in text recognition tasks.

Extensive evaluations of the MORAN framework are presented across various benchmark datasets, with a focus on word accuracy as the performance metric. The datasets

used for evaluation include both regular and irregular text, showcasing the framework's versatility.

Datasets: The MORAN framework is tested on a range of datasets, such as IIIT5K-Words, Street View Text (SVT), ICDAR 2003 (IC03), ICDAR 2013 (IC13), SVT-Perspective (SVT-P), CUTE80, and ICDAR 2015 (IC15). These datasets vary in terms of text quality, resolution, and lexicon sizes, providing a comprehensive evaluation environment.

Implementation Details: The MORAN framework utilizes a combination of the Multi-Object Rectification Network (MORN) and the Attention-Based Sequence Recognition Network (ASRN). Specific network details and hyperparameters are provided, including the number of hidden units in the decoder, output classes in ASRN, and training strategies. The curriculum learning strategy is outlined, which involves training stages for ASRN, MORN, and end-to-end optimization.

Performance of the MORAN: Table 3 demonstrate the effectiveness of the MORAN framework. A comparison of different pooling layers shows that a max-pooling layer with a kernel size of 2 and a stride of 1 yields the highest accuracy. The MORAN is evaluated at various stages of development, showcasing its performance gains with curriculum learning. Total edit distance results for ICDAR OCR tasks are also provided.

Table 3. Comparison of pooling layers in lexicon-free mode. “No”, “AP” and “MP” respectively indicate no pooling layer, an average-pooling layer and a max-pooling layer at the top of the MORN. The kernel size is 2. “s” represents the stride.

	s	IIIT5K	SVT	IC03	IC13	SVT-P	CUTE80	IC15
No	–	85.7	87.9	92.9	91.5	75.8	65.9	59.4
AP	2	89.2	87.4	94.8	91.1	75.9	71.1	64.6
AP	1	89.3	87.9	94.7	91.6	75.9	72.9	64.9
MP	2	90.4	88.2	94.5	91.8	76.1	76.4	68.4
MP	1	91.2	88.3	95.0	92.4	76.1	77.4	68.8

Table 4. Performance of the MORAN.

Method	IIIT5K	SVT	IC03	IC13	SVT-P	CUTE80	IC15
End-to-end training	89.9	84.1	92.5	90.0	76.1	77.1	68.8
Only ASRN	84.2	82.2	91.0	90.1	71.0	64.6	65.6
MORAN without FP	89.7	87.3	94.5	91.5	75.5	77.1	68.6
MORAN with FP	91.2	88.3	95.0	92.4	76.1	77.4	68.8

Table 5. Performance of the MORAN (total edit distance).

Method	IC03	IC13	IC15
End-to-end training	29.1	57.7	368.8
Only ASRN	33.8	69.1	376.8
MORAN without FP	22.7	45.3	345.2
MORAN with FP	19.8	42.0	334.0

Comparisons with Rectification Methods: The MORAN is compared with other text rectification methods, such as affine transformation and RARE, on benchmark datasets. The MORAN's flexibility in handling irregular text is highlighted, allowing it to rectify every character in an image. It is also capable of handling text of infinite length. However, the MORAN's training is acknowledged to be more challenging than some other methods.

Results on General Benchmarks: The MORAN outperforms all current state-of-the-art methods in lexicon-free mode on general benchmarks. Notably, it excels in recognizing irregular text, even outperforming methods specifically designed for perspective text.

Results on Irregular Text: The MORAN's performance on irregular text datasets, including perspective and curved text, is impressive. It matches or surpasses the state-of-the-art results, demonstrating its robustness in handling various forms of irregular text.

Despite its strengths, the MORAN has limitations. It may struggle with text that has a very large curve angle due to its training on horizontal synthetic text. Additionally, it is not

designed for vertical text, which limits its applicability in scenarios with predominantly vertical text. Moreover, the MORAN is designed for cropped text recognition and does not include text detection, making it not an end-to-end scene text recognition system. The challenges of multi-oriented text detection and recognition in complex backgrounds are highlighted.

Future directions include extending the MORAN for arbitrary-oriented text recognition and finding effective ways to combine it with scene text detection for more comprehensive scene text recognition solutions.

2.5.1 Strengths

The MORAN framework addresses the challenging problem of recognizing irregular scene text, including text with various shapes, perspectives, and distortions. It provides an effective solution for handling complex deformations, making text more readable. The article extensively evaluates MORAN on various benchmark datasets, including regular and irregular text. This comprehensive evaluation demonstrates the framework's versatility and robustness in handling different types of text.

2.5.2 Weaknesses

MORAN is not designed for vertical text, which restricts its usability in scenarios where predominantly vertical text is present. The framework may struggle with text that has a very large curve angle due to its training on horizontal synthetic text. This limitation can affect its performance on highly curved text. MORAN is designed for cropped text recognition and does not include text detection. This makes it unsuitable as a complete end-to-end scene text recognition system, and users would need to integrate it with a text detection module separately.

2.5.3 Recommendation

To enhance the framework's applicability, researchers could work on extending MORAN to recognize vertical text effectively. This would make it more versatile in handling different text orientations. Future research could focus on improving the framework's performance on text with very large curve angles. This could involve training on synthetic text with more extreme deformations. Considering the importance of end-to-end scene text recognition

CHAPTER 2

systems, efforts could be made to integrate MORAN with a text detection module. This would provide a more comprehensive solution for text recognition in complex scenes.

2.6 Seeing AI

Seeing AI, developed by Microsoft, is a helpful app designed to assist visually impaired individuals in their daily activities. The app relies on artificial intelligence and utilizes the phone camera to identify various things. Here's a breakdown of its features:

1. Short Text Channel:

- This channel automatically reads out text detected by the camera.
- It restarts if a clearer image is captured, ensuring accurate readings.

2. Document Channel:

- Seeing AI guides users in placing the camera to capture all edges of a document.
- After recognition, users can add more pages or have the text read aloud.
- Users can ask questions about the document, receiving answers from the app.

3. Product Channel:

- Recognizes products using codes on packaging, including barcodes.
- Beep sounds indicate proximity to a code, with faster beeps indicating closeness.
- When a barcode/QR is detected, Seeing AI announces the product name.

4. Person Channel:

- Users can teach the app to recognize specific individuals, announcing their names.
- Provides information on the position, facial characteristics, and expressions of detected faces.

5. Currency Channel:

- Allows users to recognize the value of currency, with the option to change the country for currency recognition.

6. Scene Channel:

- Users can capture a scene with the camera and receive descriptions in both text and speech.
- The app identifies objects within the photo.

7. Colour Channel:

- Voices out the perceived colour of objects captured by the camera.

8. Handwriting Channel:

- Recognizes handwritten text.

9. Light Channel:

- Detects the amount of light in the surroundings.
- The pitch of the tone reflects the intensity of light.

Seeing AI provides a comprehensive set of tools catering to different needs, from reading texts and recognizing products to identifying people and understanding the surroundings. It stands out for its user-friendly features, making it a valuable companion for visually impaired individuals in their daily lives.

2.6.1 Strengths and Weaknesses

Table 2.2: Strengths and Weaknesses of Seeing AI

	Strengths	Weaknesses
Short Text	Fast response time to read out the text.	Seeing AI will restart the speech if the camera keeps moving. It is hard for blind people to focus the camera on the text they want.
Document	The edge helps to detect the document	Blind people are unable to focus on the documents and paragraphs they want
Product	A beeping sound indicates the proximity of the product.	Does not inform the expiration date of the product
Person	Detect the distance and the identity of person	If people hide their faces, they will not be recognized as people.
Currency	Able to detect value of money from different countries.	Blind person will face difficulty choosing the country because

CHAPTER 2

		they cannot change the country using their voice.
Scene	Describe the scene using short and concise sentences.	The details are not included. Only the rough ideas are described.
Color	Voice out the color in fast response time	Blind people do not know what the camera is viewing. The colour accuracy is not so high.
Handwriting	Recognize handwriting and convert it into text.	Some noise will also be converted.
Light	High pitch indicates high light	Blind people do not need this

Chapter 3

System Methodology/Approach

3.1 System Design Diagram/Equation



Figure 3.1: System Design Flowchart

Figure 3.1 shows the system design flowchart. Upon installation and initial launch, the application prompts the user to grant permissions, such as accessing location data, using the camera, and utilizing the microphone. These permissions are crucial for the application's functionality. The app checks for these permissions each time it is launched, ensuring that the necessary access is granted. In case a user revokes any permission, the application requests access again before proceeding. After acquiring permissions, the application initiates a background thread to fetch nearby shop information, while the main thread waits for the thread to finish its operation.

The application utilizes an HTTP request to interact with the Places API for a nearby search. This search retrieves information about shops in the vicinity, with the potential to gather data for up to 60 shops. Subsequently, the app extracts and stores the shop names from the retrieved data into a list for further processing.

After compiling the list of nearby shop names, the application activates the device's camera. The camera is configured with both a preview use case, enabling live camera preview, and an analysis use case integrated with ML Kit's text recognizer. Users have the option to choose from various text recognition modes, including Chinese, Latin, Korean, Japanese, or Devanagari. The text recognizer continuously scans and interprets scene text detected by the camera.

While the camera identifies scene text, the application concurrently calculates the text similarity between the recognized text and the stored shop name list. If any scene text matches a shop name with a similarity score exceeding 0.7, the application employs text-to-speech functionality to audibly announce the shop name. This feature aids visually impaired users by alerting them when a shop name is detected. Interested users can tap the screen for more information.

A single tap on the screen freezes the camera, allowing the text recognizer to identify the current scene text and retrieve all shop names. The application then fetches detailed shop information using the Place details feature of the Places API, utilizing the stored placeID associated with each shop. This data is passed to AI for generating comprehensive summaries.

CHAPTER 2

The application seamlessly integrates with the Gemini API to summarize shop details obtained from the Places API. The summarized information, available in multiple languages such as English, Chinese, Korean, and Japanese based on user preference, is then converted into speech using text-to-speech functionality, providing users with a narrated overview of the shop.

When the user long-presses the screen, the application opens Google Maps and sets the destination to the shop's longitude and latitude in walking mode. This enables users to navigate to the shop efficiently through Google Maps' navigation system.

If multiple shops are detected in the camera scene, the application fetches and stores detailed information about each shop in a list. Swiping left allows the user to move to the next shop summary, while swiping right takes them back to the previous shop summary. When swiping to the last shop summary, further left swipes will maintain the current shop summary. Similarly, when viewing the first shop summary, swiping right will retain the same shop summary.

Users can hold the microphone button to ask questions about the shop. For example, they can inquire about the shop's name, and the Gemini AI will provide a response. This feature allows users to ask specific questions and obtain detailed shop information effortlessly.

Double-tapping activates the camera's live preview mode, allowing users to capture another shop name or scene of interest.

3.1.1 Use Case Diagram and Description

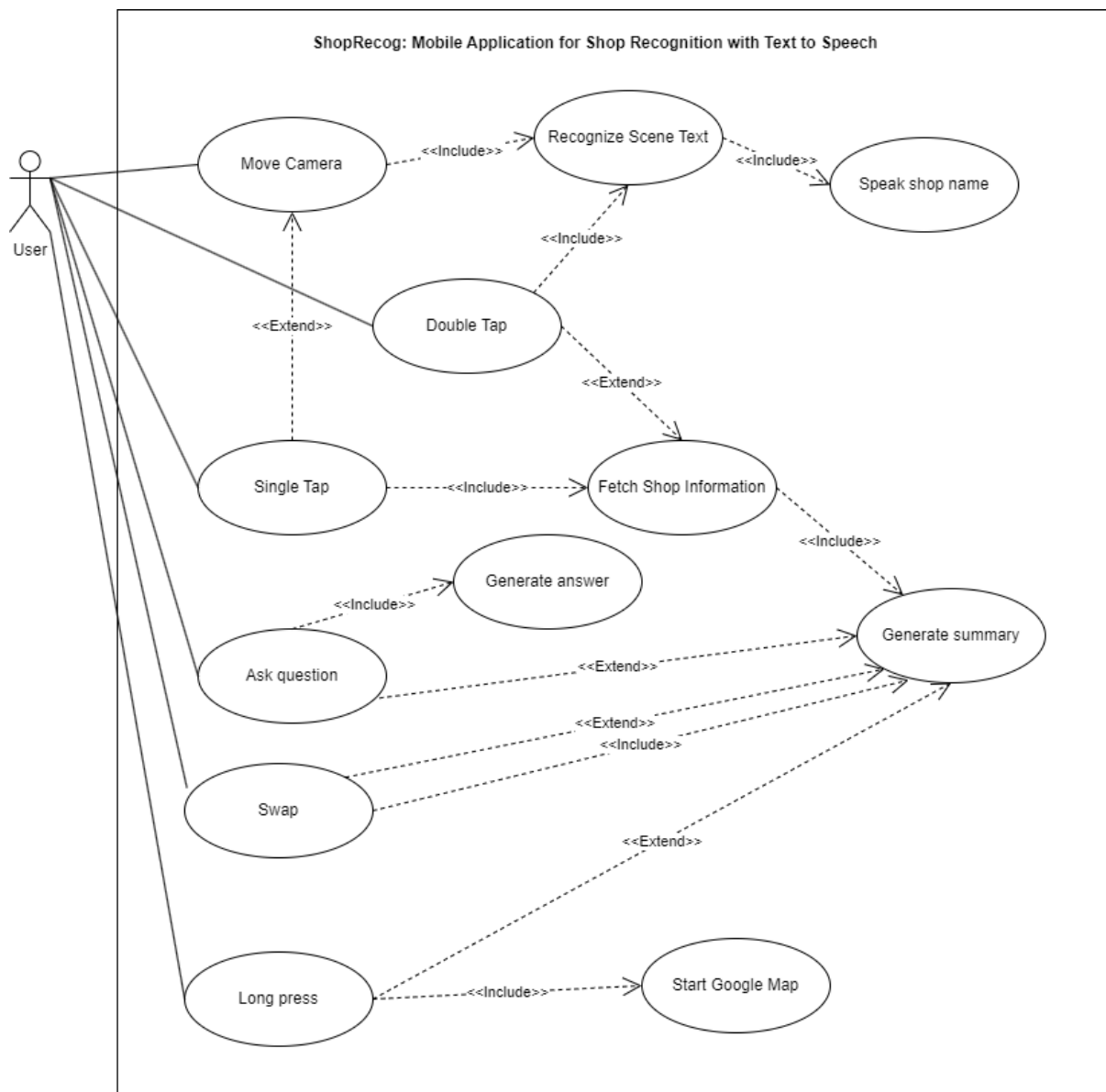


Figure 3.2: System Use Case Diagram

Figure 3.2 shows the System Use Case Diagram. As a user, there are various actions you can perform while using this application. However, some actions can only be executed under specific conditions. Firstly, you can move the camera to recognize scene text. When the recognized scene text has a text similarity of more than 0.7 compared to the shop name list, the application will speak the shop name. Then, you have the option to single tap to freeze the camera or not.

CHAPTER 2

If you single tap the screen, the detected shop name will be used to fetch shop information. This shop information is then passed to Gemini AI to generate a summary paragraph, which can be translated into different languages. You can choose whether to ask questions and get answers from Gemini AI or swap to get the next shop information summary. Additionally, you can long-press the screen to initiate Google Maps, setting the destination as the shop location in walking mode. This allows you to navigate to the shop effortlessly.

If you want to retrieve information about another shop, you can double tap to activate the camera's live view and recognize new scene text.

3.1.2 Activity Diagram

Move Camera

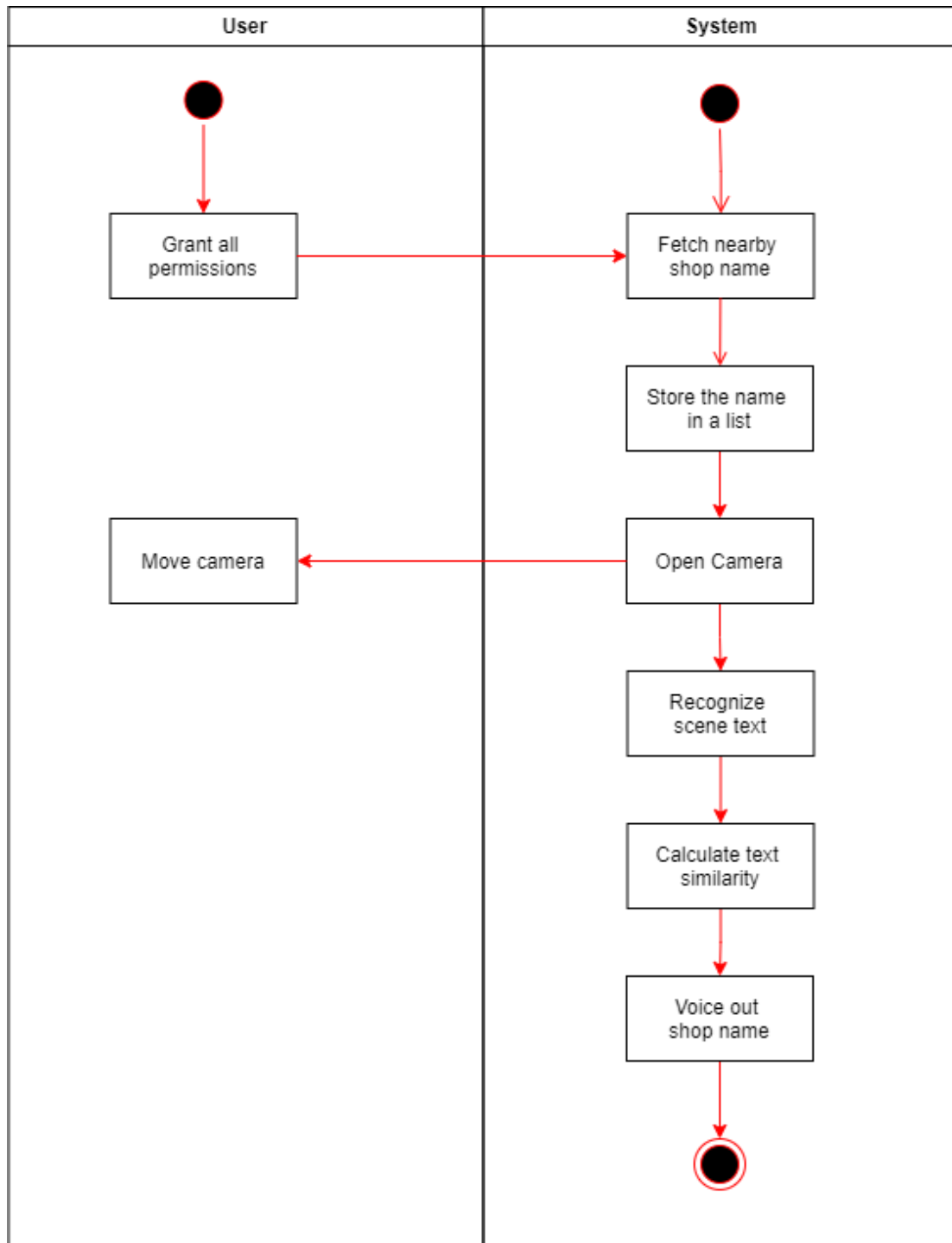


Figure 3.3: Move Camera Activity Diagram

CHAPTER 2

Upon installation and initial launch, the app requests users to grant essential permissions, such as accessing location data, using the camera, and utilizing the microphone. These permissions are vital for the app's functionality. Every time the app is launched, it checks for these permissions to ensure that necessary access is available. If a user revokes any permission, the app prompts for access again before proceeding. Once permissions are granted, the app initiates a background task to fetch nearby shop information, while the main process awaits the completion of this task.

The app employs an HTTP request to communicate with the Places API for a nearby search. This search retrieves details about shops in the vicinity, capable of gathering data for up to 60 shops. Afterwards, the app extracts and stores shop names from the retrieved data into a list for further handling.

Upon gathering the list of nearby shop names, the app activates the device's camera. The camera is set up with both a preview mode for live camera feed and an analysis mode integrated with ML Kit's text recognizer. Users can select from various text recognition options, such as Chinese, Latin, Korean, Japanese, or Devanagari. The text recognizer continually scans and interprets scene text captured by the camera.

While the camera processes scene text, the app simultaneously computes the text similarity between recognized text and the stored shop name list. If any scene text matches a shop name with a similarity score above 0.7, the app utilizes text-to-speech functionality to audibly announce the shop name. This feature assists visually impaired users by alerting them when a shop name is detected. Users can tap the screen for further details if interested.

Single Tap

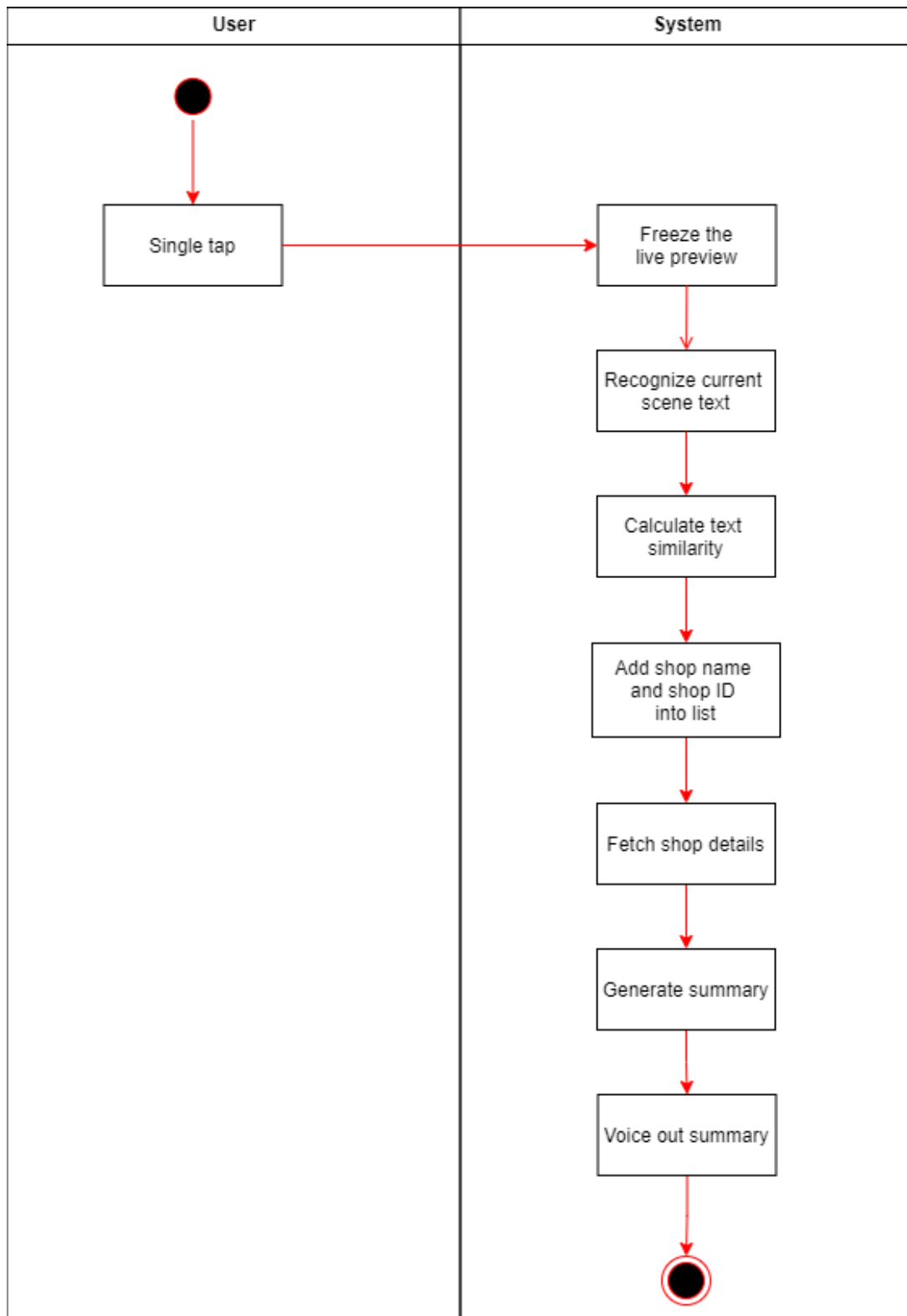


Figure 3.4: Single Tap Activity Diagram

CHAPTER 2

A single tap on the screen freezes the camera's live preview, enabling the text recognizer to analyze the current scene text and retrieve all shop names for text similarity assessment. The shop names and placeIDs of those shops that exhibit a text similarity score above 0.7 are added to specific lists. Subsequently, the app leverages the Place details feature of the Places API to fetch comprehensive shop information, utilizing the stored placeIDs linked to each shop. This detailed shop data is then utilized by AI to generate thorough summaries. The app smoothly integrates with the Gemini API to summarize shop details sourced from the Places API. The summary, available in multiple languages based on user preference, such as English, Chinese, Korean, and Japanese, is then converted into spoken content using text-to-speech functionality. This process offers users a narrated overview of the shop's information details.

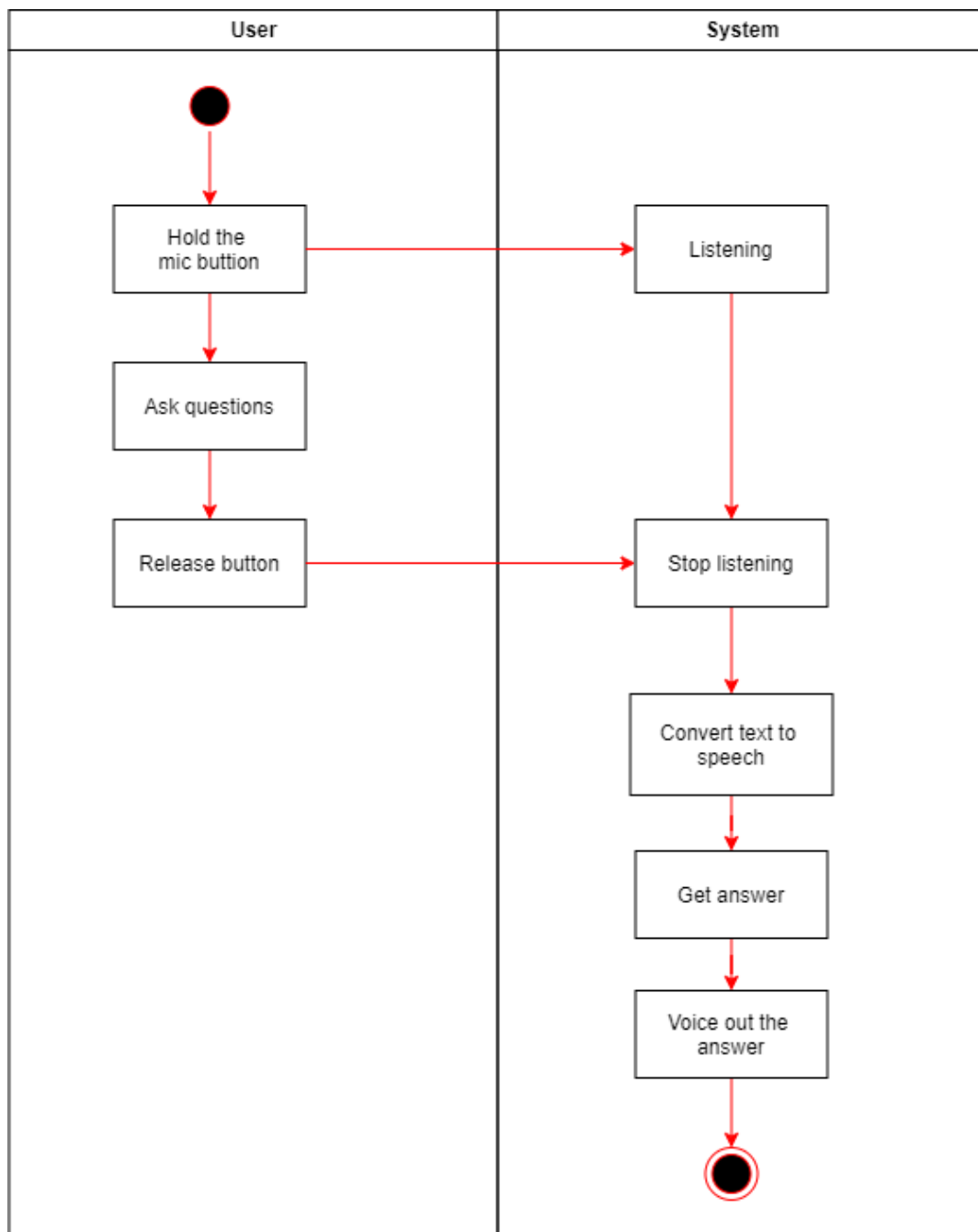
Ask question

Figure 3.5: Ask Question Activity Diagram

When the user presses and holds the microphone button, the application activates the speech-to-text recognizer, which begins listening to the user's speech input. Once the user releases the button, the speech-to-text recognizer ceases listening and proceeds to process the captured speech. It converts the spoken words into text format, which is then transmitted to the Gemini AI for analysis and response generation.

CHAPTER 2

The Gemini AI processes the text input, interprets the user's query or request, and formulates an appropriate response. This response is then passed back to the application, which utilizes text-to-speech functionality to audibly articulate the generated answer. This seamless integration between speech recognition, AI processing, and text-to-speech synthesis ensures that users can effortlessly interact with the application using spoken commands and receive spoken responses in return, enhancing the overall user experience.

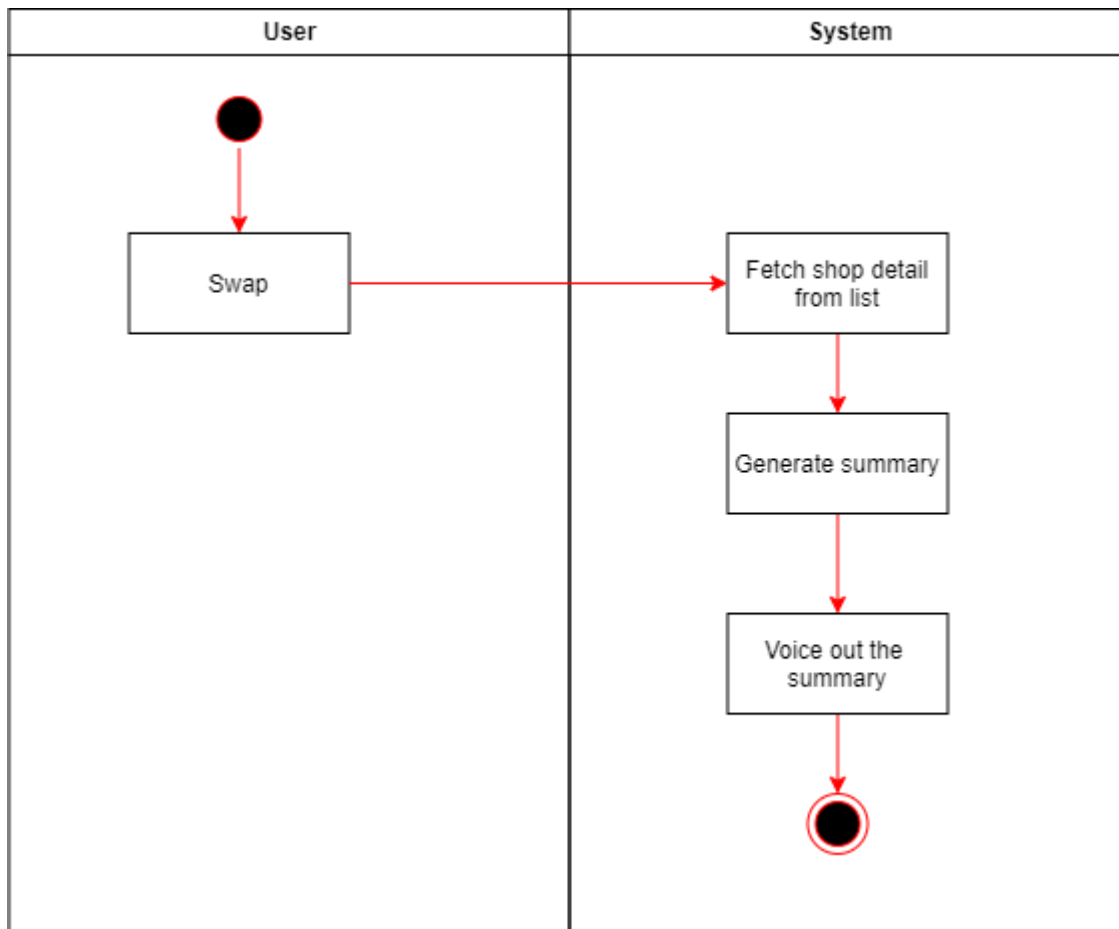
Swap

Figure 3.6: Swap Activity Diagram

When the user performs a swipe gesture on the screen, the application fetches the next shop's detailed information from a predefined list. This information, which includes specifics about the shop such as its name, location, and address, is then passed on to the Gemini AI for processing. The Gemini AI analyzes the gathered data and generates a concise summary of the shop, highlighting key details that are relevant to the user.

If the user swipes to the right, the application retrieves the information about the next shop in the list, allowing users to seamlessly explore different shops in the vicinity. Conversely, swiping to the left retrieves the details of the previous shop, enabling users to revisit previously viewed shop information if needed. This intuitive swipe-based navigation enhances user interaction, making it easier for users to access and digest information about various shops through the application.

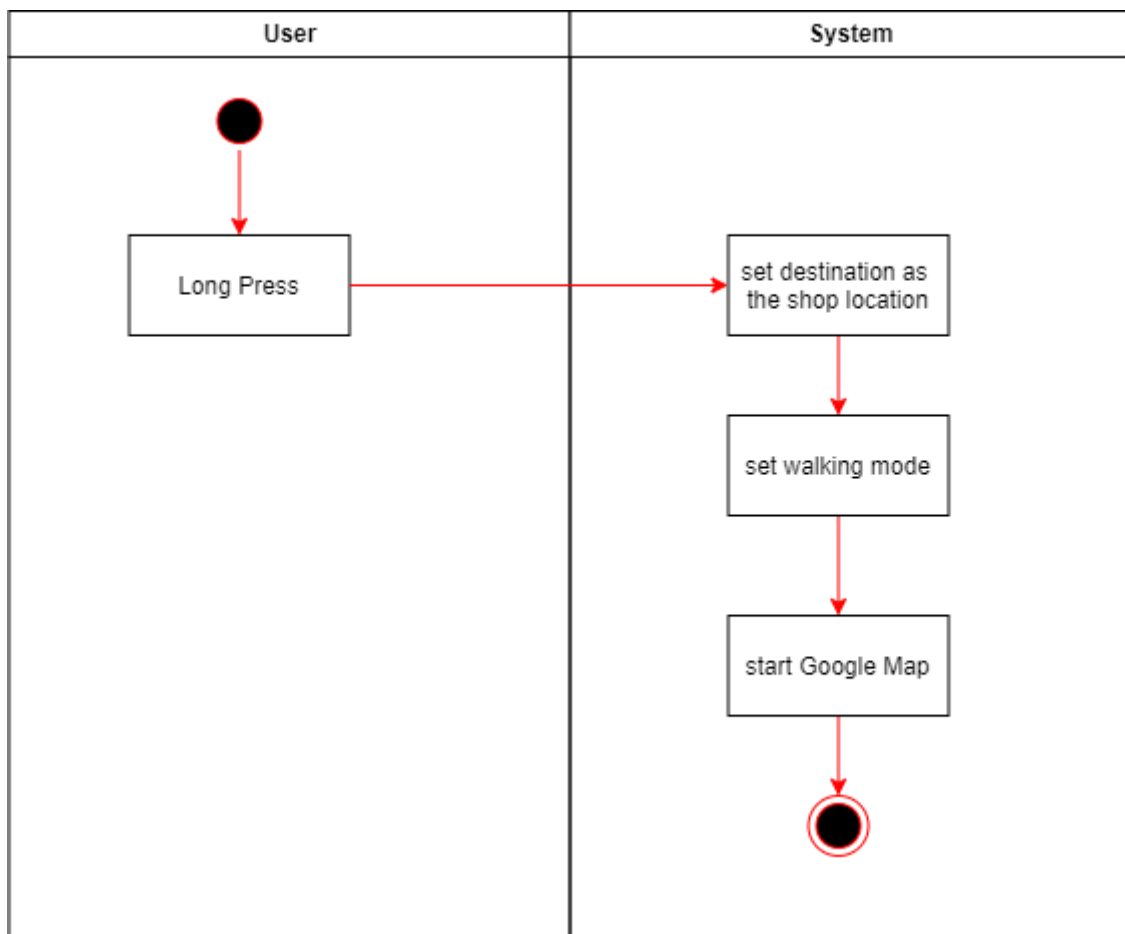
Long Press

Figure 3.7: Long Press Activity Diagram

Upon a long press of the screen, the application seamlessly integrates with Google Maps, initiating the navigation process with the shop's location set as the destination. Specifically, the navigation mode is set to walking mode, ensuring users can navigate to the shop efficiently and effectively on foot. This feature enhances user convenience by providing a straightforward method for reaching their desired shop location, making the navigation experience smooth and hassle-free.

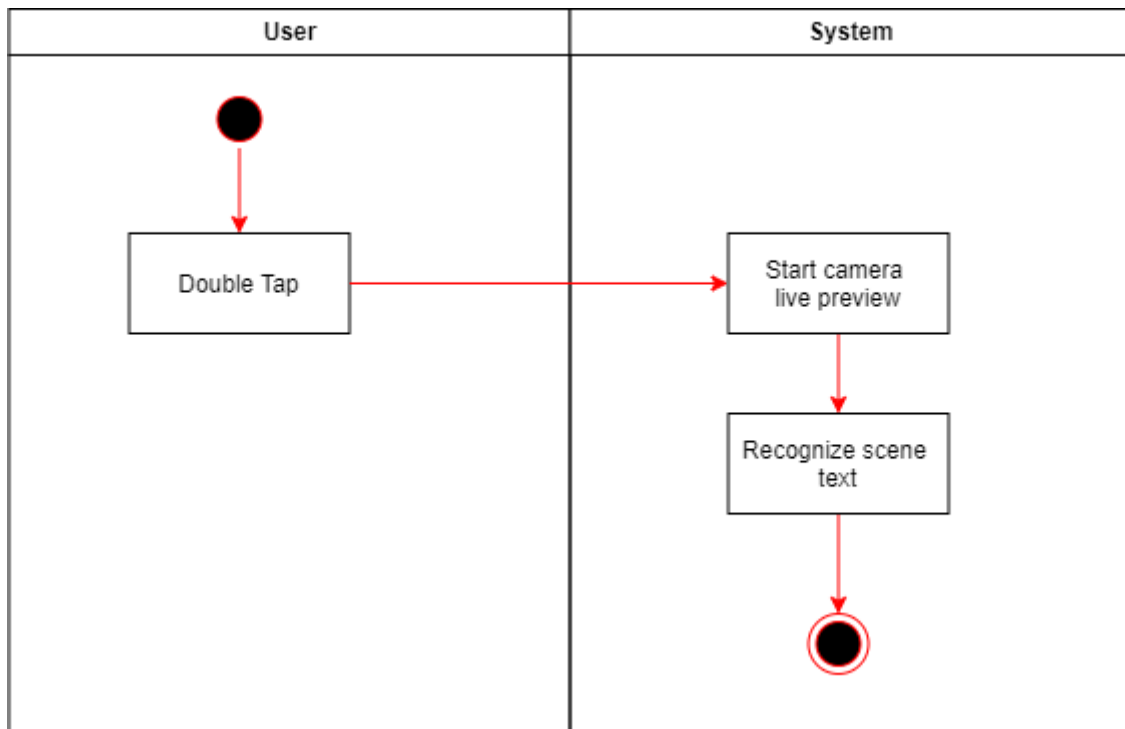
Double Tap

Figure 3.8: Double Tap Activity Diagram

A double tap on the screen activates the camera's live preview mode within the application. This feature enables continuous recognition of scene text, empowering users to identify and capture new shop names using the camera. Once a new shop name is detected and the user single taps the screen, the application automatically retrieves the corresponding shop information, providing users with access to the desired details about the shop.

Chapter 4

System Design

4.1 System Block Diagram

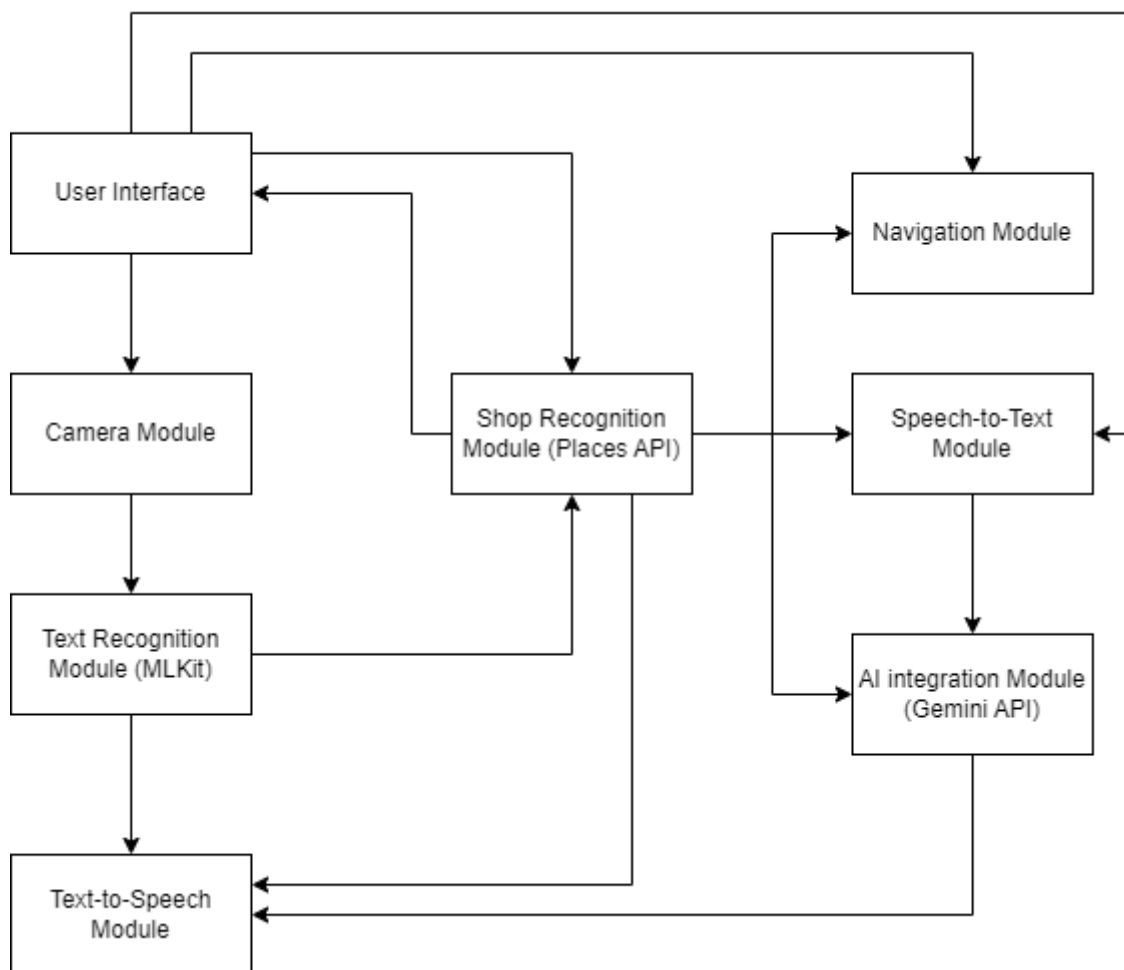


Figure 4.1 System Block Diagram

Figure 4.1 shows the system block diagram. At the core of the system is the User Interface (UI) component, providing a tactile and auditory interface for user interaction. Users can initiate actions through touch gestures and voice commands, facilitating intuitive navigation and engagement with the application's functionalities.

CHAPTER 2

Upon launching the application, the user interface allows for a single tap on the screen, which triggers the freezing of the camera feed. This action is pivotal as it enables the text recognizer to analyze and interact with the captured scene text effectively.

The frozen camera feed is bound to the Analysis Use Case, which initiates the creation of a text recognizer. This component, powered by machine learning algorithms such as those found in MLKit, processes the visual data to extract and recognize text within the camera's field of view.

The recognized text undergoes multiple processes simultaneously. Firstly, it is converted into speech output through the Text-to-Speech Module, providing users with auditory feedback regarding the identified text, such as shop names within the camera's view.

Simultaneously, the recognized text serves as input for the Shop Recognition Module, a critical component responsible for identifying and matching shop names or keywords within the recognized text. This module interfaces with external APIs, such as the Google Places API, to fetch detailed information about recognized shops, including addresses, contact numbers, operating hours, and user ratings.

The retrieved shop detail information is then channeled to the Gemini API Integration Module. This module leverages AI-powered algorithms to generate concise summaries about each recognized shop, encapsulating essential details in a user-friendly format. The generated summaries are voiced out through the text-to-speech functionality, providing users with comprehensive and accessible information about nearby shops.

The user interface offers additional functionality through touch gestures. A long press on the screen triggers the Navigation Mode, seamlessly integrating with Google Maps to provide step-by-step navigation guidance to selected shops. This feature enhances user mobility and independence, empowering them to navigate unfamiliar environments confidently.

Moreover, users can swipe horizontally on the screen to navigate between different shop information summaries. This functionality enables seamless exploration of nearby shops, allowing users to access comprehensive details about multiple establishments with ease.

CHAPTER 2

The Shop Recognition Module acts as a crucial intermediary within the system, producing outputs that serve as inputs for various modules such as navigation, speech-to-text interaction, and AI-powered summary generation. This modular and interconnected design ensures efficient data flow and enhances the overall user experience within the application.

4.2 System Components Specifications

User Interface

The user interface (UI) module is designed for simplicity and ease of use, catering specifically to visually impaired users. At the top of the screen, a navigation bar features two spinners, allowing users to select the language for both text recognition and speech output within the application. This language customization ensures a personalized and accessible experience for users from different linguistic backgrounds.

Additionally, a bottom navigation bar contains a microphone button, visible after the shop summary is generated. This button enables users to interact with the application through speech commands, enhancing user engagement and accessibility. The UI supports various touch gestures, including single tap, double tap, long press, and swipe, providing intuitive control over the application's functionalities.

The central screen area is dedicated to the live camera feed, crucial for detecting and recognizing text in real time using ML Kit's text recognition capabilities.

Shop Recognition Module

Upon application launch, the Shop Recognition Module utilizes the Google Places API to fetch nearby shop names and their corresponding place IDs, storing them in a list for reference. This module plays a pivotal role in matching recognized text from the camera feed with the shop name list, ensuring a text similarity threshold of higher than 0.7 for accurate shop name detection.

When a match is found, the shop name is voiced out using the Text-to-Speech Module. Subsequently, upon user interaction such as a single tap on the screen, this module fetches detailed shop information. This information is then utilized by other modules, including the

CHAPTER 2

Navigation Module, Speech-to-Text Module, and AI Integration Module, enhancing the overall functionality and utility of the application.

Camera Module

The Camera Module is responsible for initializing and managing the device's camera functionality within the application. It is bound to two key use cases:

- **Preview Use Case:** Enables live preview of the camera feed, essential for real-time text recognition and analysis.
- **Analysis Use Case:** Builds and configures the text recognizer, facilitating the continuous recognition of scene text within the camera's field of view.

The camera module works in conjunction with the Text Recognition Module to capture and process scene text, subsequently identifying and voicing out recognized shop names.

Text Recognition Module

Implemented using ML Kit's textRecognizer class, the Text Recognition Module offers users the flexibility to choose the text recognition language from options such as Latin, Chinese, Japanese, Korean, or Devanagari. This module's primary function is to accurately recognize text from the camera feed and pass the results to the Shop Recognition Module or the Text-to-Speech Module based on the identified language and context.

Text-to-Speech Module

The Text-to-Speech Module converts recognized text into speech output, ensuring that the spoken language aligns with the user's selected preferences, which may include Chinese, English, Japanese, or Korean. This customization enhances user comprehension and engagement with the auditory feedback provided by the application.

Navigation Module

Utilizing the longitude and latitude information obtained from the Shop Recognition Module, the Navigation Module integrates with Google Maps API to initiate navigation to the selected shop's location in walking mode. This seamless integration enhances user mobility and independence by providing step-by-step guidance to the desired destination.

Speech-to-Text Module

CHAPTER 2

When activated by the user holding the mic button, the Speech-to-Text Module utilizes speech-to-text recognition to transcribe spoken queries or commands. It stops listening when the user releases the button, passing the transcribed questions to the AI Integration Module for processing and response generation.

AI Integration Module

The AI Integration Module collaborates with the Shop Recognition Module to receive detailed shop information, facilitating the generation of summary paragraphs using Gemini API. Additionally, queries and commands transcribed by the Speech-to-Text Module are processed by the AI Integration Module, leveraging Gemini API to provide accurate and informative spoken responses, thereby enhancing the interactive and informative capabilities of the application.

Chapter 5

System Implementation

5.1 Hardware setup

The hardware involved in this project is a computer and an android mobile device. A computer is used to do the coding and a mobile device is used for testing and deploying the application.

Table 5.1 Specifications of laptop

Description	Specifications
Model	Acer Swift 3
Processor	AMD Ryzen 7 4700U
Operating System	Windows 11
Graphic	AMD Radeon Graphics
Memory	16GB RAM
Storage	475GB

Table 5.2 Specifications of mobile device

Description	Specifications
Model	POCO F4 GT
Processor	Snapdragon 8 Gen 1
Operating System	Android 12
Memory	16GB RAM
Storage	256GB

5.2 Software setup

Diagram Tools:

1. Visual Paradigm:

- Visual Paradigm is a versatile diagramming and modeling tool. It will create visual representations of the project's architecture, system design, and data flow.

2. Draw.io:

- Draw.io is a user-friendly diagramming tool that is particularly useful for creating flowcharts, process diagrams, and system diagrams.

Programming Tools:

3. Android Studio

- Android Studio is the official Integrated Development Environment for Android app development. It is used for developing the Android application's frontend. It provides tools for designing the user interface, coding the app's functionality, and testing it on Android devices.

Machine Learning Framework:

4. ML Kit

- Google's ML Kit is a machine learning (ML) platform designed to make it easy for mobile app developers to integrate machine learning capabilities into their applications. It is a part of Firebase, Google's mobile development platform. ML Kit provides a range of pre-trained models and APIs that developers can use to add machine learning features to their Android and iOS apps without requiring extensive expertise in machine learning.

5.3 Setting and Configuration

CHAPTER 2

Before initiating the project development, it is essential to install and download Android Studio on the laptop. Subsequently, the user has the option to create a virtual device within Android Studio or utilize their personal phone for debugging the application. The step-by-step setup instructions are provided in Android Studio.

Google's ML Kit offers the Text Recognition v2 API, requiring Android API level 21 (Android 5.0) or above. Additionally, the device must be connected to the internet to access Google APIs and Gemini API. Finally, the application requires granted access to the camera, microphone and location for proper functionality.

5.4 System Operation (with Screenshot)

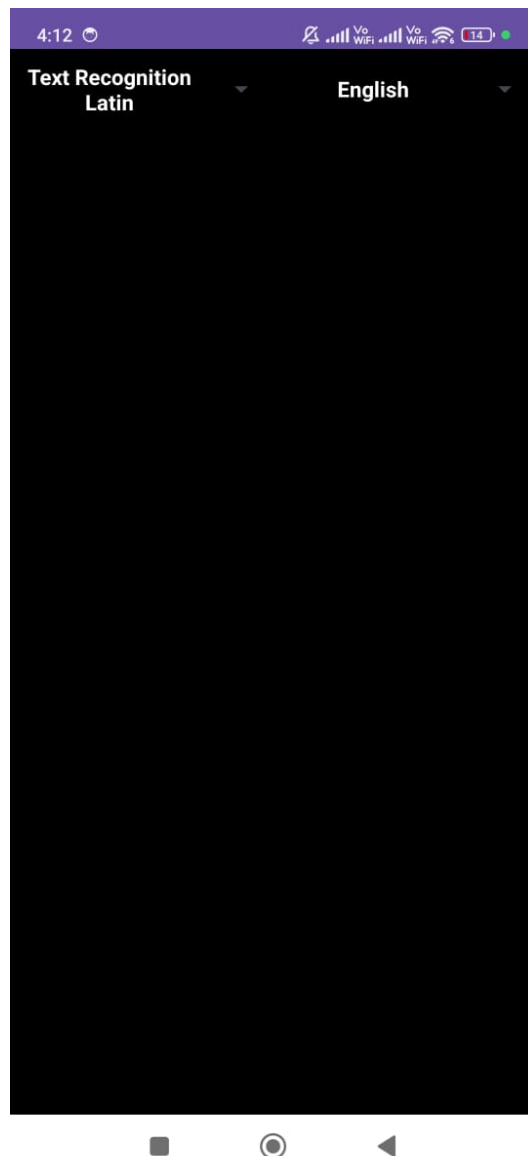
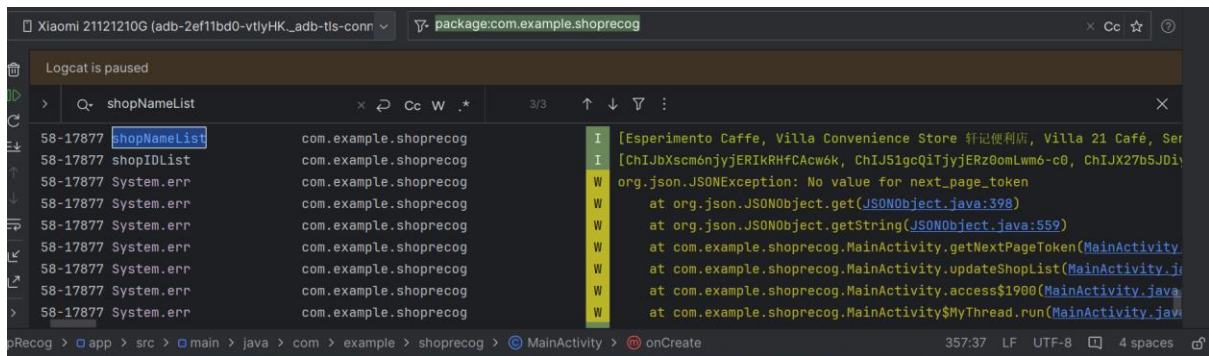


Figure 5.1: Fetching Nearby Shop

When the application is launched, it will undergo black screen for a few seconds because Places API is fetching nearby shop information to store the shop name and shop ID into a list.

CHAPTER 2



```
Logcat is paused
shopNameList
58-17877 shopNameList com.example.shoprecog I [Esperimento Caffè, Villa Convenience Store 轩记便利店, Villa 21 Café, Ser
58-17877 shopIDList com.example.shoprecog I [ChIJbXscm6njyJERIKRhfAcw6k, ChIJ51gcQiTjyjERz0omLwm6-c0, ChIJX27b5JD1
58-17877 System.err com.example.shoprecog W org.json.JSONException: No value for next_page_token
58-17877 System.err com.example.shoprecog W at org.json.JSONObject.get(JSONObject.java:398)
58-17877 System.err com.example.shoprecog W at org.json.JSONObject.getString(JSONObject.java:559)
58-17877 System.err com.example.shoprecog W at com.example.shoprecog.MainActivity.getNextPageToken(MainActivity.j
58-17877 System.err com.example.shoprecog W at com.example.shoprecog.MainActivity.updateShopList(MainActivity.j
58-17877 System.err com.example.shoprecog W at com.example.shoprecog.MainActivity.access$1900(MainActivity.java
58-17877 System.err com.example.shoprecog W at com.example.shoprecog.MainActivity$MyThread.run(MainActivity.java
```

Figure 5.2: shopNameList and shopIDList

In Figure 5.2, shopNameList is used to store the shopName and shopIDList is used to store the placeID. Places API can fetch up to 60 nearby shop information.



Figure 5.3: Launching Application

After all the shop names are stored in the list, the application will open device camera for live preview and text recognition. The user can move the camera. If the shop name is detected, the application will voice out the shop name.

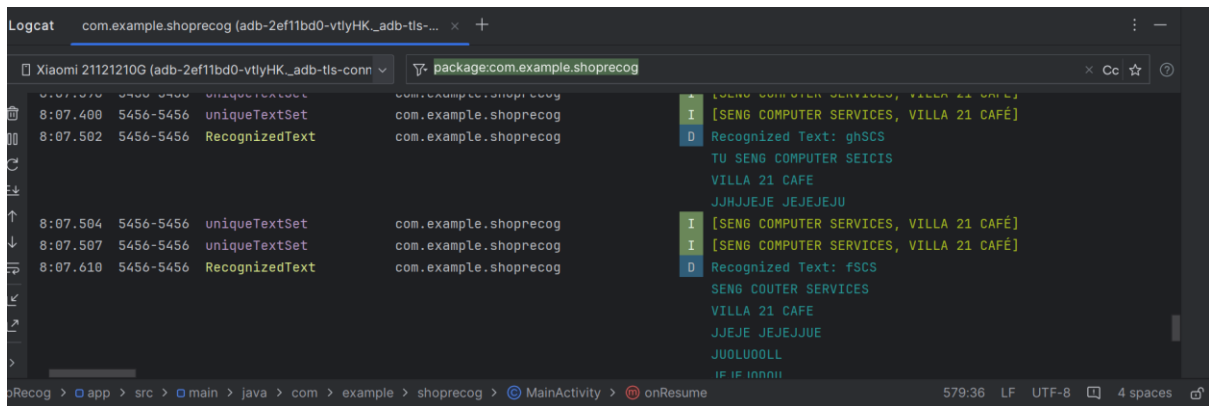


Figure 5.4: Recognized Text

Since the recognized text has more than 0.7 text similarity compared to the shop name, the shop names are added into `uniqueTextSet`. The shop names inside the `uniqueTextSet` will be voiced out.

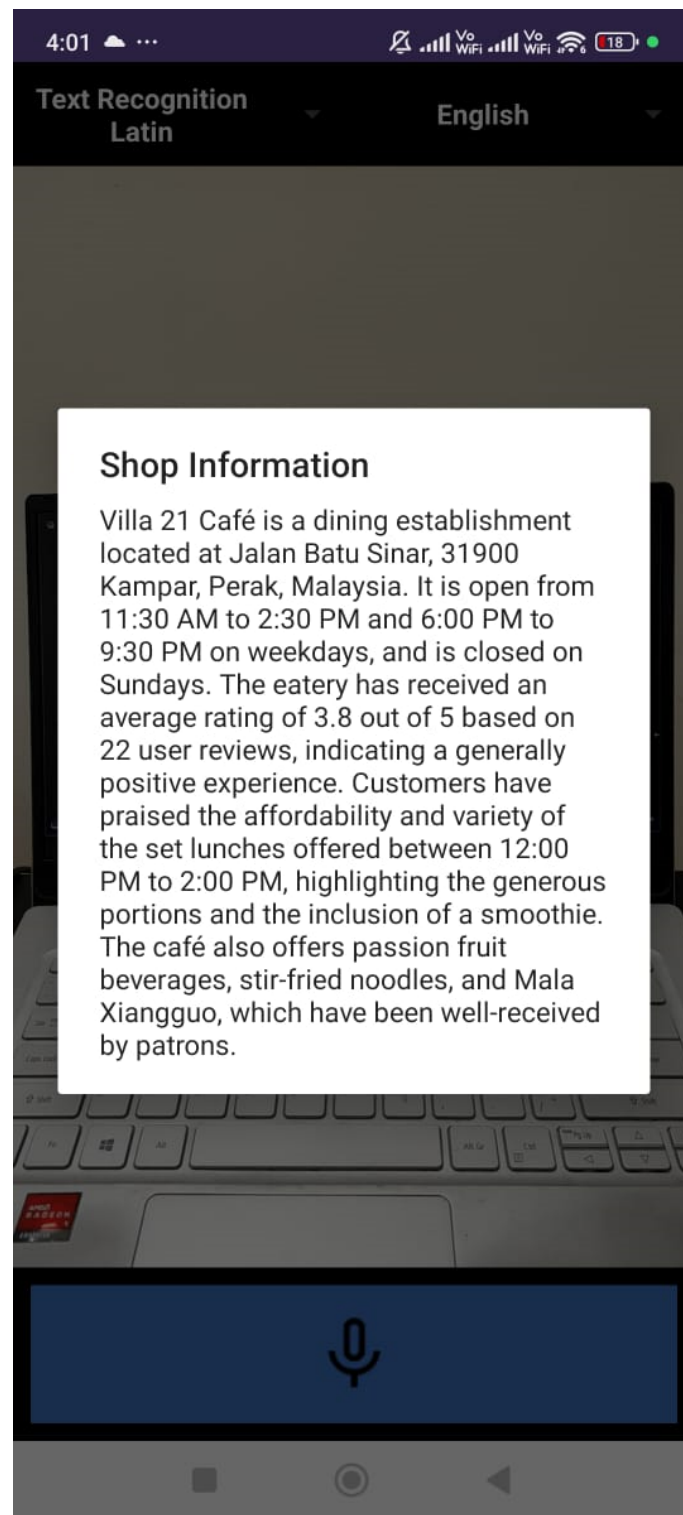


Figure 5.5: Summarized Text

After the user single tap the screens, both Villa 21 café and Seng Computer services place IDs are used to fetch the shop detail information. The shop information is then passed to Gemini AI to generate a shop summary.



Figure 5.6: Asking Question

If the user holds the button, the button will become green colour and the application will say “Please speak.”. This indicates that the speech-to-text recognizer is listening to the speech.



Figure 5.7: Speech-to-text Conversion

Figure 5.7 shows that the speech is successfully converted into text. The text will then be passed to Gemini AI to get the answer. The answer will be voiced out by the application.

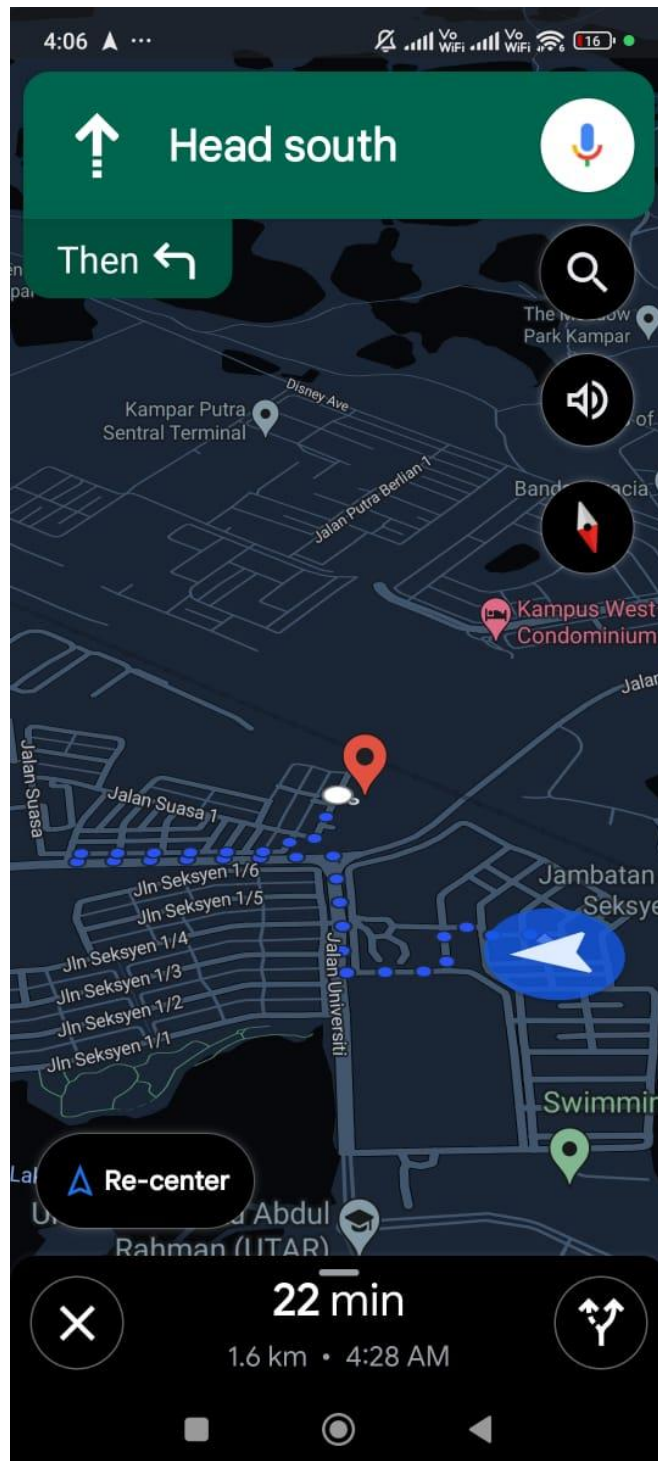


Figure 5.8: Navigation using Google Map

If the user long presses the screen, the application will call Google Map and set the destination as the shop location. In this case, Villa 21 café location is set as the destination, Google Map will be in charge of navigating the user to the destination.

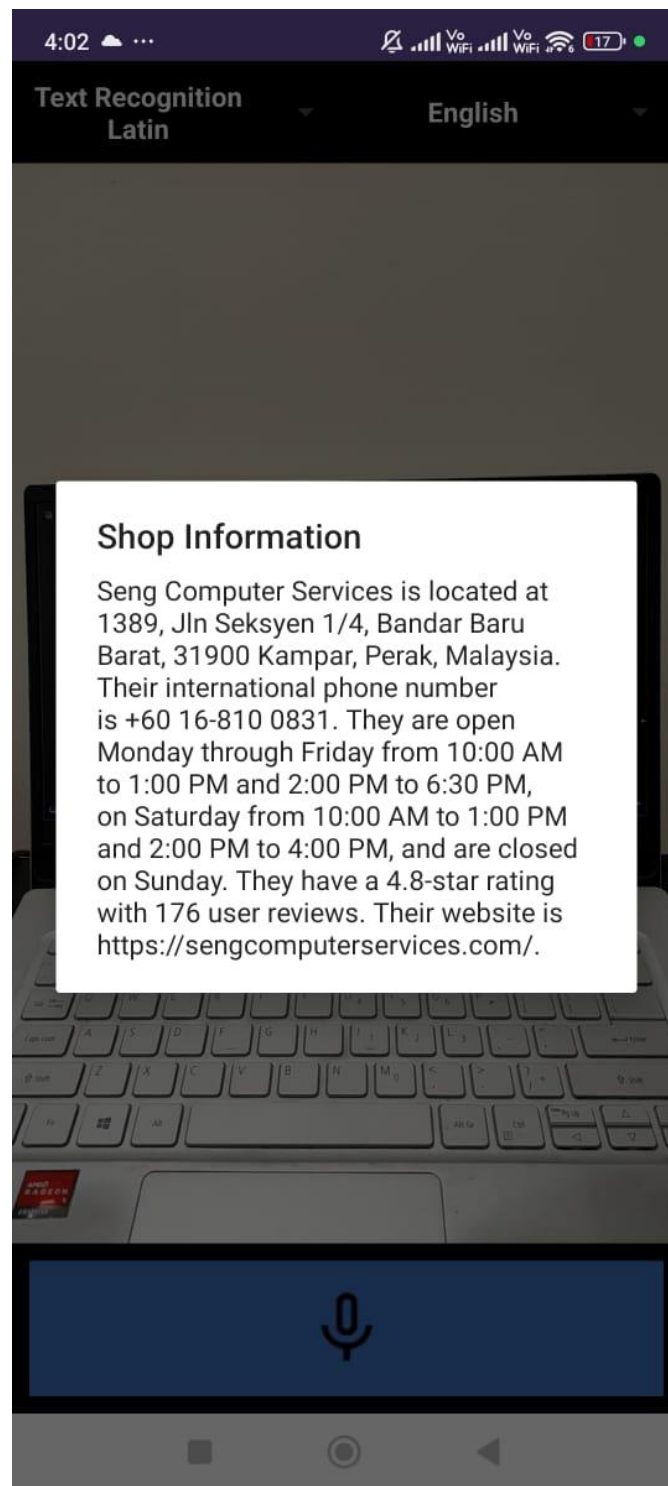


Figure 5.9: Swapping

If the user swipes to left, next shop information is passed to Gemini AI to generate a summary text and the result is as shown as Figure 5.9.



Figure 5.10: Double Tap

After the user double tap, the camera will return to its live preview mode and the mic button will disappear. The user can then capture another shop name to get the shop information.

5.5 Implementation Issues and Challenges

One of the primary challenges encountered during implementation was the efficient management of background threads. The application extensively utilizes background threads for tasks such as calling the Places API to fetch nearby shop information. However, coordinating these threads to ensure that the main thread waits for essential results before proceeding to the next step posed a significant challenge. This requires careful design and implementation of threading mechanisms to ensure smooth execution and prevent race conditions or data inconsistency issues.

The integration of ML Kit classes and APIs presented another hurdle during development. Understanding the functionalities provided by each ML Kit class, as well as the nuances of sending requests and receiving responses from APIs, required in-depth study of documentation and iterative testing. Issues such as timing discrepancies, where rapid requests could overwhelm the API and lead to unexpected errors or delays, needed to be carefully addressed through proper rate limiting and error handling strategies.

Designing a user interface tailored for blind users posed its own set of challenges. The goal was to simplify the interface by focusing on auditory feedback through text-to-speech functionality while implementing intuitive gestures for navigation and interaction. However, implementing gesture controls smoothly proved challenging, with issues such as override conflicts and unresponsive buttons arising. Overcoming these challenges involved iterative testing, refining gesture recognition algorithms, and ensuring that the interface remained accessible and responsive for users with visual impairments.

Overall, addressing these implementation challenges required a combination of technical expertise, thorough testing, and a deep understanding of accessibility principles to ensure that the application delivers a seamless and user-friendly experience for visually impaired individuals.

5.6 Concluding Remark

In conclusion, the development journey of the "ShopRecog" application has been both challenging and rewarding. The hardware and software setups provided a robust foundation for coding, testing, and deploying the application on a laptop and an Android mobile device. Utilizing tools such as Visual Paradigm, Draw.io, Android Studio, and Google's ML Kit enabled the creation of a feature-rich and accessible application tailored for visually impaired users.

Despite facing implementation issues and challenges such as managing background threads efficiently, integrating complex ML Kit functionalities and APIs, and designing an intuitive user interface for blind users, the development team persevered through careful planning, thorough testing, and iterative refinement.

The system operation screenshots provided a visual representation of the application's functionality, showcasing key features such as fetching nearby shop information, text recognition, shop summary generation, speech-to-text conversion, navigation using Google Maps, and seamless interaction through gestures.

Addressing the implementation challenges required technical expertise, attention to detail, and a deep understanding of accessibility principles. By overcoming these challenges, the application delivers a seamless and user-friendly experience, empowering visually impaired individuals to navigate and interact with their surroundings independently.

Overall, the project's successful implementation underscores the importance of accessibility in technology, highlighting how innovative solutions can enhance the quality of life and independence for individuals with disabilities.

Chapter 6

System Evaluation And Discussion

6.1 System Testing and Performance Metrics

The system underwent comprehensive testing to ensure its functionality met the specified requirements and performance standards. Various test cases were designed to evaluate different aspects of the application, including user interaction, data processing, API integrations, and overall system responsiveness.

During testing, the following performance metrics were considered:

- **Accuracy of Text Recognition:** The system's ability to accurately recognize and match shop names from the camera feed.
- **Speed of Shop Information Retrieval:** The time taken to fetch and display detailed shop information after the user triggers the shop summary generation.
- **Reliability of API Integrations:** Testing the reliability and responsiveness of external APIs such as Places API and Gemini API for fetching shop details and generating summaries.
- **User Interface Responsiveness:** Evaluating the responsiveness of the user interface to various touch gestures and user interactions.
- **Voice Output Quality:** Assessing the clarity and accuracy of the text-to-speech conversion for voiced shop names and summaries.

The system testing aimed to validate that the application functions as expected under different scenarios and user inputs, providing an intuitive and seamless experience for visually impaired users.

6.2 Testing Setup and Result

The testing setup involved executing a series of test cases designed to validate key functionalities and interactions within the application. Each test case outlined the expected result based on user actions and system responses. The table below summarizes the test cases along with the actual results obtained during testing:

Test Case	Expected Result	Result (Pass/Fail)
When the user launches the application	The application should ask permission for camera, microphone, and location access. Then, the application will open the phone camera.	Pass
When the user moves the camera	If shop name is detected, the name will be voiced out.	Pass
When the user taps the screen one time	The camera will be frozen. The application will fetch the shop detail information and generate summary automatically.	Pass
When there is more than one shop name detected in a photo	The application will fetch all shop information detail and generate first shop summary automatically.	Pass
When the user screen still displays the taken photo and the user long presses the screen	The application will open Google Map where the destination is the selected shop.	Pass
When the user screen still displays the taken photo and the user hold the mic button.	The speech-to-text recognizer will listen to what the user says and convert the speech to text	Pass

CHAPTER 2

When the user screen still displays the taken photo and the user double tap	The camera will resume live preview and the user can take picture again	Pass
When the user screen still displays the taken photo and the user swap to left	Next shop summary is voiced out.	Pass
When the user screen still displays the taken photo and the user swap to right	Previous shop summary is voiced out.	Pass
When the user screen still displays the taken photo and the user swaps to right when it is the first shop	First shop summary is voiced out.	Pass
When the user screen still displays the taken photo and the user swaps to left when it is the last shop	Last shop summary is voiced out.	Pass
When the user changes the application language	The application will change language in text and in voice	Pass
When the user changes the language for text recognition model	The application will build the text recognition model using that language	Pass
The user return to the application after pausing it for some time	The application will open live preview camera and allow user to take photo	Pass
All process can produce an output within 5 seconds.	All process can produce an output within 5 seconds.	Pass

Table 6.1: Test Case and Expected Output

CHAPTER 2

The testing results indicate that the application successfully passed all test cases, meeting the expected outcomes for various user interactions and functionalities.

6.3 Project Challenges

Throughout the development and implementation of the application, several challenges were encountered and overcome.

First is the integration of Background Threads: Managing background threads efficiently to ensure proper synchronization and handling of asynchronous tasks, especially when interacting with external APIs like Places API and Gemini API, posed a challenge. Careful threading mechanisms and error handling strategies were implemented to address this challenge.

Integrating and leveraging ML Kit functionalities, including text recognition and language processing, required a deep understanding of ML Kit classes and APIs. Overcoming issues such as rapid request handling and API response management involved thorough testing and refinement.

Designing a user interface that prioritizes accessibility for visually impaired users while maintaining simplicity and intuitiveness presented a significant challenge. Overcoming override conflicts, ensuring gesture recognition accuracy, and optimizing button responsiveness were key aspects addressed during interface design and testing.

Dependence on external APIs such as Google's Places API and Gemini API for shop information retrieval and summary generation necessitated robust error handling and fallback mechanisms to ensure consistent performance and reliable data retrieval.

Despite these challenges, the project team successfully navigated through iterative development, testing, and refinement phases to deliver a functional, accessible, and user-friendly application for visually impaired users.

6.4 Objectives Evaluation

The project's objectives, as outlined in Chapter 1, were designed to address the unique needs of visually impaired individuals by creating a comprehensive mobile application. Let's evaluate each objective to determine the extent to which they have been achieved.

Real-Time Shop Recognition: The objective to develop a robust mechanism for real-time shop recognition using camera input has been successfully accomplished. The application effectively leverages the Places API nearby search to fetch a comprehensive list of shop names, providing accurate and timely information about nearby establishments to the user.

Text Recognition and Matching: Integration of ML Kit's text recognition capabilities has enabled the application to analyze live camera previews and identify text corresponding to shop names. Advanced algorithms ensure accurate matching of recognized text with entries in the shop name list, meeting the targeted 70% text similarity threshold for precise detection.

Text-to-Speech Functionality: The implementation of text-to-speech technology has significantly enhanced the user experience. The application audibly announces detected shop names in real time, providing immediate auditory feedback to blind users and empowering them with essential information for independent navigation and decision-making.

Shop Details Retrieval: Upon user interaction, such as tapping the screen, the application successfully retrieves detailed information about recognized shops using the Places API. This includes vital details like addresses, phone numbers, business hours, and user ratings, enriching the user's understanding of each shop.

AI-Generated Summaries: Integration with the Gemini AI API has facilitated the generation of summary paragraphs for each shop based on retrieved details. The application offers language translation options (Chinese, Korean, Japanese, and English), catering to diverse user preferences and international travel scenarios effectively.

Interactive Speech-to-Text: The implementation of a user-friendly interface with a hold-to-open microphone button for speech-to-text interaction has been achieved. Users can ask

CHAPTER 2

questions about specific shops, send queries to the Gemini API for processing, and receive spoken answers via text-to-speech conversion seamlessly.

Navigation Assistance: The integration of navigation features with Google Maps has significantly enhanced user mobility. Users can set shop destinations and initiate walking mode navigation directly from the application, receiving step-by-step guidance to their desired locations with ease.

In conclusion, the application has successfully achieved and even exceeded its primary objectives. The comprehensive features implemented, including real-time shop recognition, text-to-speech functionality, shop details retrieval, AI-generated summaries, interactive speech-to-text, and navigation assistance, collectively contribute to a highly accessible, informative, and user-friendly experience for visually impaired individuals. The project's successful outcome demonstrates effective integration of cutting-edge technologies to address real-world accessibility challenges.

6.5 Concluding Remark

In conclusion, the System Evaluation and Discussion chapter provides a comprehensive overview of the rigorous testing, performance metrics, challenges faced, and the evaluation of project objectives. The thorough testing and successful validation of key functionalities ensure that the application meets the specified requirements and performance standards, delivering an intuitive and seamless experience for visually impaired users.

Despite encountering challenges such as thread management, API integrations, UI design complexities, and external dependencies, the project team effectively navigated through these hurdles to achieve a highly functional and accessible mobile application. The project's successful outcome reflects the dedication, technical expertise, and innovative problem-solving approach employed throughout the development process.

The project's objectives were not only met but exceeded, as evidenced by the application's robust real-time shop recognition, accurate text-to-speech functionality, comprehensive shop details retrieval, AI-generated summaries, interactive speech-to-text capabilities, and seamless navigation assistance.

Overall, the application stands as a testament to the effective integration of cutting-edge technologies to address real-world accessibility challenges, ultimately empowering visually impaired individuals with enhanced independence, information access, and navigation support.

Chapter 7

Conclusion and Recommendation

7.1 Conclusion

The development of the application represents a significant milestone in leveraging technology to enhance the accessibility and independence of visually impaired individuals. Through a comprehensive system design, meticulous implementation, and rigorous testing, the application has successfully addressed key challenges faced by visually impaired users in navigating and accessing information about their surroundings.

The core features of the application, including real-time shop recognition, text-to-speech functionality, shop details retrieval, AI-generated summaries, interactive speech-to-text capabilities, and navigation assistance, have been meticulously implemented and tested. These features collectively provide a seamless and intuitive user experience, empowering visually impaired users to make informed decisions, navigate unfamiliar environments confidently, and access essential information about nearby shops efficiently.

The successful achievement of project objectives, as outlined in Chapter 1, underscores the dedication, technical expertise, and innovative problem-solving approach of the development team. The application's ability to accurately recognize shop names, fetch detailed shop information, generate informative summaries, and facilitate interactive user engagement demonstrates its effectiveness in addressing the unique needs of visually impaired individuals. Moreover, the project's adherence to accessibility principles, including intuitive gesture controls, language customization options, and seamless integration with assistive technologies, ensures that the application is truly inclusive and user-friendly.

Overall, the "ShopRecog" application stands as a testament to the positive impact of technology in enhancing accessibility and independence for individuals with visual impairments. Its successful implementation serves as a model for future endeavors aimed at leveraging technology to create inclusive solutions for diverse user needs.

7.2 Recommendation

While the "ShopRecog" application has achieved considerable success in meeting its objectives and delivering a user-friendly experience, there are several recommendations for further enhancement and future development:

Continuous Improvement: Maintain a process of continuous improvement by gathering feedback from users, incorporating suggestions for feature enhancements, addressing any usability issues, and staying updated with emerging technologies and accessibility standards.

Expand Language Support: Consider expanding language support to include additional languages, dialects, and localization options, catering to a broader user base and enhancing the application's global usability.

Enhanced AI Integration: Explore opportunities to further enhance AI integration by leveraging advanced natural language processing (NLP) techniques, sentiment analysis, and personalized recommendations, providing users with more tailored and contextually relevant information.

Collaboration with Accessibility Organizations: Foster collaborations with accessibility organizations, advocacy groups, and community stakeholders to gain insights into evolving accessibility needs, promote awareness about the application, and facilitate outreach efforts to reach a wider audience.

Integration with Wearable Technologies: Explore possibilities for integration with wearable technologies such as smart glasses or wearable haptic devices, enhancing the application's accessibility and usability in diverse environments and scenarios.

User Training and Support: Provide comprehensive user training materials, tutorials, and support resources to ensure that visually impaired users can fully leverage the application's features and functionalities, fostering user empowerment and independence.

Security and Privacy Measures: Implement robust security and privacy measures to safeguard user data, ensure compliance with data protection regulations, and build trust among users regarding the confidentiality and integrity of their information.

By incorporating these recommendations and embracing a user-centric approach to development, the "ShopRecog" application can continue to evolve as a leading solution for enhancing accessibility, fostering independence, and improving the quality of life for visually impaired individuals.

REFERENCES

- [1] P. Tripathi, "A journey through history: The evolution of OCR technology," Docsumo.com, <https://www.docsumo.com/blog/optical-character-recognition-history> (accessed Dec. 7, 2023).
- [2] 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.
- [3] A. A. Panchal, S. Varde and M. S. Panse, "Character detection and recognition system for visually impaired people," 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 2016, pp. 1492-1496, doi: 10.1109/RTEICT.2016.7808080.
- [4] B. Shi, X. Bai and C. Yao, "An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 11, pp. 2298-2304, 1 Nov. 2017, doi: 10.1109/TPAMI.2016.2646371.
- [5] C. Yi and Y. Tian, "Scene Text Recognition in Mobile Applications by Character Descriptor and Structure Configuration," in *IEEE Transactions on Image Processing*, vol. 23, no. 7, pp. 2972-2982, July 2014, doi: 10.1109/TIP.2014.2317980.
- [6] C. Luo, L. Jin, and Z. Sun, "Moran: A multi-object rectified attention network for scene text recognition," *Pattern Recognition*, vol. 90, pp. 109–118, 2019. doi:10.1016/j.patcog.2019.01.020

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	Study week no.: 1
Student Name & ID: Khoo Zi Yi (20ACB03614)	
Supervisor: Dr Ng Hui Fuang	
Project Title: Mobile application for Shop Recognition with Text to Speech	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Camera module is done.

2. WORK TO BE DONE

Shop recognition module.

3. PROBLEMS ENCOUNTERED

No

4. SELF EVALUATION OF THE PROGRESS

Keep improving



Supervisor's signature



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	Study week no.: 3
Student Name & ID: Khoo Zi Yi (20ACB03614)	
Supervisor: Dr Ng Hui Fuang	
Project Title: Mobile application for Shop Recognition with Text to Speech	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Shop recognition is done.

2. WORK TO BE DONE

Text-to-speech module

3. PROBLEMS ENCOUNTERED

No

4. SELF EVALUATION OF THE PROGRESS

Keep improving



Supervisor's signature



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	Study week no.: 5
Student Name & ID: Khoo Zi Yi (20ACB03614)	
Supervisor: Dr Ng Hui Fuang	
Project Title: Mobile application for Shop Recognition with Text to Speech	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Text-to-speech is done.

2. WORK TO BE DONE

Shop recognition module.

3. PROBLEMS ENCOUNTERED

No

4. SELF EVALUATION OF THE PROGRESS

Keep improving



Supervisor's signature



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	Study week no.: 7
Student Name & ID: Khoo Zi Yi (20ACB03614)	
Supervisor: Dr Ng Hui Fuang	
Project Title: Mobile application for Shop Recognition with Text to Speech	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Shop recognition is done.

2. WORK TO BE DONE

AI integration module.

3. PROBLEMS ENCOUNTERED

No

4. SELF EVALUATION OF THE PROGRESS

Keep improving



Supervisor's signature



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	Study week no.: 9
Student Name & ID: Khoo Zi Yi (20ACB03614)	
Supervisor: Dr Ng Hui Fuang	
Project Title: Mobile application for Shop Recognition with Text to Speech	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

AI integration module is done.

2. WORK TO BE DONE

Speech-to-Text module.

3. PROBLEMS ENCOUNTERED

No

4. SELF EVALUATION OF THE PROGRESS

Keep improving



Supervisor's signature



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	Study week no.: 11
Student Name & ID: Khoo Zi Yi (20ACB03614)	
Supervisor: Dr Ng Hui Fuang	
Project Title: Mobile application for Shop Recognition with Text to Speech	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Speech-to-Text module is done.

2. WORK TO BE DONE

Language module.

3. PROBLEMS ENCOUNTERED

No

4. SELF EVALUATION OF THE PROGRESS

Keep improving



Supervisor's signature



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	Study week no.: 13
Student Name & ID: Khoo Zi Yi (20ACB03614)	
Supervisor: Dr Ng Hui Fuang	
Project Title: Mobile application for Shop Recognition with Text to Speech	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Language module is done.

2. WORK TO BE DONE

Design the user interface.

3. PROBLEMS ENCOUNTERED

No

4. SELF EVALUATION OF THE PROGRESS

Keep improving



Supervisor's signature



Student's signature

POSTER



Faculty of Information and communication Technology

Bachelor of Computer Science (Hons)



SHOPRECOG: MOBILE APPLICATION FOR SHOP RECOGNITION WITH TEXT TO SPEECH

Prepared by: Khoo Zi Yi
Supervised by: Dr. Ng Hui Fuang

INTRODUCTION

This project addresses the challenges faced by blind individuals when identifying and accessing nearby shops. Leveraging Google's ML Kit and scene text recognition, the mobile application assists blind users in recognizing shop names through the phone camera. The system not only voices out shop names but also fetches detailed information about them, enhancing the independence of blind individuals. Key objectives include developing a user-friendly interface, providing directional guidance, and estimating distances.



Shop Information

Seng Computer Services is located at 1389, Jln Seksyen 1/4, Bandar Baru Barat, 31900 Kampar, Perak, Malaysia. Their international phone number is +60 16-810 0831. They are open Monday through Friday from 10:00 AM to 1:00 PM and 2:00 PM to 6:30 PM, on Saturday from 10:00 AM to 1:00 PM and 2:00 PM to 4:00 PM, and are closed on Sunday. They have a 4.8-star rating with 176 user reviews. Their website is <https://sengcomputerservices.com/>.

PROBLEM STATEMENT

- Blind individuals encounter difficulty in identifying the shops around them.
- Blind individuals face challenges in estimating the distance between themselves and the shops.
- Blind people require directional guidance to reach shops.

RESEARCH OBJECTIVES

- Develop a user-friendly mobile application for blind people.
- Assist blind people in recognizing nearby shops
- Provide shop information to blind individuals through speech
- Offer directional guidance for blind individuals reaching their destinations
- Estimate the distance between blind individuals and the shop

PLAGIARISM CHECK RESULT

Turnitin Originality Report

Processed on: 26-Apr-2024 07:48 +08
 ID: 2362005192
 Word Count: 13690
 Submitted: 1

FYP2 Report By Khoo Zi Yi

Similarity Index	Similarity by Source	
13%	Internet Sources:	10%
	Publications:	9%
	Student Papers:	N/A

1% match (Internet from 04-Mar-2023) https://deepai.org/publication/a-multi-object-rectified-attention-network-for-scene-text-recognition
1% match (Chucai Yi, Yingli Tian. "Scene Text Recognition in Mobile Applications by Character Descriptor and Structure Configuration", IEEE Transactions on Image Processing, 2014) Chucai Yi, Yingli Tian. "Scene Text Recognition in Mobile Applications by Character Descriptor and Structure Configuration", IEEE Transactions on Image Processing, 2014
1% match (Internet from 15-Dec-2022) https://www.ijert.org/research/scene-text-recognition-in-mobile-applications-by-character-descriptor-and-structure-configuration-IJERTCONV3IS04012.pdf
1% match () Memon, Jamshed, Sami, Maira, Khan, Rizwan Ahmed. "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)", 2019
1% match (Internet from 22-Dec-2023) https://www.coursehero.com/file/p596trc/Chapter-1-Chapter-1-Chapter-1-Chapter-1-Chapter-1-Chapter-1-Chapter-1-Chapter-1/
1% match (Canjie Luo, Lianwen Jin, Zenghui Sun. "MORAN: A Multi-Object Rectified Attention Network for Scene Text Recognition", Pattern Recognition, 2019) Canjie Luo, Lianwen Jin, Zenghui Sun. "MORAN: A Multi-Object Rectified Attention Network for Scene Text Recognition", Pattern Recognition, 2019
1% match (Internet from 02-Jul-2015) http://journal.selvamtech.com/files/volu03/is01/vixe.pdf
< 1% match (Internet from 10-Oct-2023) http://eprints.utar.edu.my/5563/1/fyp_IA_2023_ORFY.pdf
< 1% match (Internet from 16-Sep-2023) http://eprints.utar.edu.my/5526/1/fyp_IA_2023_ACD.pdf
< 1% match (Internet from 20-Jan-2024) http://eprints.utar.edu.my/5504/1/fyp_CT_2023_HJH.pdf
< 1% match (Internet from 30-Mar-2023) http://eprints.utar.edu.my/4715/1/fyp_IA_2022_CWS.pdf
< 1% match (Internet from 13-Mar-2024) http://eprints.utar.edu.my/6034/1/fyp_CS_2023_HIZQ.pdf
< 1% match (Internet from 20-Jan-2024) http://eprints.utar.edu.my/6033/1/fyp_CS_2023_CJV.pdf
< 1% match (Internet from 15-Dec-2022) http://eprints.utar.edu.my/4674/1/fyp_CS_2022_TCH.pdf
< 1% match (Internet from 15-Dec-2022) http://eprints.utar.edu.my/4640/1/fyp_CS_2022_CGQ.pdf

CHAPTER 2

- < 1% match (Internet from 30-Mar-2023)
http://eprints.utar.edu.my/4741/1/fyp_IB_2022_NWB.pdf
- < 1% match (Internet from 20-Oct-2023)
http://eprints.utar.edu.my/5519/1/fyp_CS_2023_KSZH.pdf
- < 1% match (Internet from 08-Oct-2022)
<http://eprints.utar.edu.my/1952/1/CN%2D2016%2D1102867%2D1.pdf>
- < 1% match (Internet from 15-Dec-2022)
http://eprints.utar.edu.my/4653/1/fyp_CS_2022_LCS.pdf
- < 1% match (Internet from 31-Aug-2022)
<https://deepai.org/publication/handwritten-optical-character-recognition-ocr-a-comprehensive-systematic-literature-review-sl/>
- < 1% match (Internet from 09-Aug-2020)
<https://www.ijert.org/artificial-neural-networks-applied-to-obtain-saturation-curves-of-a-three-phase-induction-motor>
- < 1% match (Internet from 29-Jul-2023)
https://cloud.tencent.com/developer/article/1008764?areaSource=106000.2&from=article_detail.1665394&traceId=Z1_R2XHGFjggVBSq3pRxd
- < 1% match (Internet from 02-Oct-2022)
<https://pnrsolution.org/Datacenter/Vol4/Issue2/Vol4%20Issue2.pdf>
- < 1% match (Internet from 10-Nov-2021)
https://fict.utar.edu.my/documents/FYP/FYP2_template/FYP2_Report_Template_CN.docx
- < 1% match (Internet from 05-Jan-2023)
<https://www.ijeter.everscience.org/Manuscripts/Volume-6/Issue-3/Vol-6-issue-3-M-25.pdf>
- < 1% match (Internet from 13-Feb-2023)
https://www.researchgate.net/publication/322060419_Towards_End-to-End_Text_Spotting_with_Convolutional_Recurrent_Neural_Networks
- < 1% match ("Computer Vision", Springer Science and Business Media LLC, 2017)
["Computer Vision", Springer Science and Business Media LLC, 2017](#)
- < 1% match (Akhilesh A. Panchal, Shrugal Varde, M. S. Panse. "Character detection and recognition system for visually impaired people", 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 2016)
[Akhilesh A. Panchal, Shrugal Varde, M. S. Panse. "Character detection and recognition system for visually impaired people", 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology \(RTEICT\), 2016](#)
- < 1% match (Internet from 06-Apr-2024)
<https://jchr.org/index.php/JCHR/article/download/673/676/1335>
- < 1% match ("Pattern Recognition", Springer Science and Business Media LLC, 2020)
["Pattern Recognition", Springer Science and Business Media LLC, 2020](#)
- < 1% match (Internet from 03-Jan-2024)
<https://fastercapital.com/keyword/resource-constrained-devices.html>
- < 1% match (Internet from 10-Apr-2024)
https://www.acadlore.com/journals/HF/2023_HF_SI001
- < 1% match (Toshiaki Nishio, Yuichiro Yoshikawa, Kohei Ogawa, Hiroshi Ishiguro. "Development of an Effective Information Media Using Two Android Robots", Applied Sciences, 2019)
[Toshiaki Nishio, Yuichiro Yoshikawa, Kohei Ogawa, Hiroshi Ishiguro. "Development of an Effective Information Media Using Two Android Robots", Applied Sciences, 2019](#)
- < 1% match (Internet from 20-Aug-2022)
http://olarik.it.msu.ac.th/wp-content/uploads/2021/11/Thesis_Thanadol.pdf
- < 1% match (Internet from 24-Feb-2023)
<https://sciendo.com/downloadpdf/journals/amcs/23/4/article-p887.pdf>
- < 1% match (Internet from 27-Sep-2022)
<http://www.ijarcsms.com/docs/paper/volume3/issue4/V3I4-0014.pdf>
- < 1% match (Baoguang Shi, Xiang Bai, Cong Yao. "An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017)
[Baoguang Shi, Xiang Bai, Cong Yao. "An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017](#)
- < 1% match (Internet from 14-Jan-2024)
<https://www.thepharmajournal.com/archives/2019/vol8issue2S/PartA/S-12-12-407-955.pdf>
- < 1% match ("Assistive Text Reading from Natural Scene for Blind Persons", Mobile Cloud Visual Media Computing, 2015.)
["Assistive Text Reading from Natural Scene for Blind Persons", Mobile Cloud Visual Media Computing, 2015.](#)
- < 1% match (Internet from 08-Feb-2023)
<https://aem65-origin.sprint.com/en/shop/cell-phones/samsung-s22-5g.html>
- < 1% match (Internet from 17-Nov-2022)
https://eprints.ucm.es/id/eprint/68286/1/VARELA_LORENZO_Glucose_classification_andprediction_system_withNeural_Networks_-_Alejandro_Varela_-_Alvaro_Delgado_4398577_1519069772.pdf
- < 1% match (Internet from 15-Dec-2022)
<https://vdocument.in/fpga-based-maximum-power-point-tracking-of-project-report-submitted-in-partial.html>

CHAPTER 2

< 1% match (Tong Shan, Jun Li, Xiao Hou, Peijin Huang, Xiaojun Guo. "Live Demonstration: Efficient Organic Photodetector based Active Matrix Imager for Real-time Optical Character Recognition", 2023 IEEE 5th International Conference on Artificial Intelligence Circuits and Systems (AICAS), 2023)

[Tong Shan, Jun Li, Xiao Hou, Peijin Huang, Xiaojun Guo. "Live Demonstration: Efficient Organic Photodetector based Active Matrix Imager for Real-time Optical Character Recognition", 2023 IEEE 5th International Conference on Artificial Intelligence Circuits and Systems \(AICAS\), 2023](#)

< 1% match (Internet from 28-Oct-2023)

https://www.techscience.com/files/cmc/2023/TSP_CMC-76-2/TSP_CMC_41191/TSP_CMC_41191.epub

< 1% match (Internet from 30-Mar-2024)

https://i1login.easychair.org/publications/preprint_download/RRq3

< 1% match (Cong Wang, Cheng-Lin Liu. "Multi-Branch Guided Attention Network for Irregular Text Recognition", Neurocomputing, 2020)

[Cong Wang, Cheng-Lin Liu. "Multi-Branch Guided Attention Network for Irregular Text Recognition", Neurocomputing, 2020](#)

< 1% match (Internet from 26-Sep-2023)

<https://ijecs.in/index.php/ijecs/article/download/4751/4085/9477>

< 1% match (Internet from 15-Jul-2023)

<https://pubmed.ncbi.nlm.nih.gov/30535020/>

< 1% match (Internet from 23-Jan-2023)

<http://www.cs.rpi.edu/~zaki/PaperDir/TWEB20.pdf>

< 1% match (Internet from 09-Nov-2022)

<https://www.ijserd.com/articles/IJSRDV5I10542.pdf>

< 1% match (Indhumathi Chandrasekeran, A Dharmaraj, Ashima Juyal, M. Shravan, Rajesh Deb Barman, Melanie Lourens. "Cryptocurrency and Data Privacy in Human Resource Management", 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2023)

[Indhumathi Chandrasekeran, A Dharmaraj, Ashima Juyal, M. Shravan, Rajesh Deb Barman, Melanie Lourens. "Cryptocurrency and Data Privacy in Human Resource Management", 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering \(ICACITE\), 2023](#)

< 1% match (Pasqual Martí Gimeno. "Towards Sustainable and Efficient Road Transportation: Development of Artificial Intelligence Solutions for Urban and Interurban Mobility", Universitat Politècnica de Valencia, 2024)

[Pasqual Martí Gimeno. "Towards Sustainable and Efficient Road Transportation: Development of Artificial Intelligence Solutions for Urban and Interurban Mobility", Universitat Politècnica de Valencia, 2024](#)

< 1% match (Xavier Righetti, Sylvain Cardin, Daniel Thalmann. "Chapter 13 WAPA: A Wearable Framework for Aerobic Pilot Aid", Springer Science and Business Media LLC, 2009)

[Xavier Righetti, Sylvain Cardin, Daniel Thalmann. "Chapter 13 WAPA: A Wearable Framework for Aerobic Pilot Aid". Springer Science and Business Media LLC, 2009](#)

< 1% match (Xiang Bai, Cong Yao, Wenyu Liu. "Strokelets: A Learned Multi-Scale Mid-Level Representation for Scene Text Recognition", IEEE Transactions on Image Processing, 2016)

[Xiang Bai, Cong Yao, Wenyu Liu. "Strokelets: A Learned Multi-Scale Mid-Level Representation for Scene Text Recognition", IEEE Transactions on Image Processing, 2016](#)

< 1% match (Internet from 17-May-2018)

<http://www.rroij.com/open-access/text-extraction-from-natural-sceneimages-and-conversion-to-audio-in-smartphone-applications-.php?aid=44224>

Universiti Tunku Abdul Rahman			
Form Title : Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)			
Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 01/10/2013	Page No.: 1 of 1



**FACULTY OF INFORMATION AND COMMUNICATION
TECHNOLOGY**

Full Name(s) of Candidate(s)	Khoo Zi Yi
ID Number(s)	20ACB03614
Programme / Course	FICT / CS
Title of Final Year Project	ShopRecog: Mobile application for Shop Recognition with Text to Speech

Similarity	Supervisor's Comments (Compulsory if parameters of originality exceeds the limits approved by UTAR)
Overall similarity index: <u>13</u> % Similarity by source Internet Sources: <u>10</u> % Publications: <u>9</u> % Student Papers: <u>N/A</u> %	
Number of individual sources listed of more than 3% similarity: <u>0</u>	
Parameters of originality required and limits approved by UTAR are as Follows: (i) Overall similarity index is 20% and below, and (ii) Matching of individual sources listed must be less than 3% each, and (iii) Matching texts in continuous block must not exceed 8 words <i>Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.</i>	

Note Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

Signature of Supervisor

Name: Ng Hui Fuang

Date: 26/4/2024

Signature of Co-Supervisor

Name: _____

Date: _____



UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY (KAMPAR CAMPUS)

CHECKLIST FOR FYP2 THESIS SUBMISSION

Student Id	20ACB03614
Student Name	Khoo Zi Yi
Supervisor Name	Dr. Ng Hui Fuang

TICK (√)	DOCUMENT ITEMS
	Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item.
√	Title Page
√	Signed Report Status Declaration Form
√	Signed FYP Thesis Submission Form
√	Signed form of the Declaration of Originality
√	Acknowledgement
√	Abstract
√	Table of Contents
√	List of Figures (if applicable)
√	List of Tables (if applicable)
√	List of Symbols (if applicable)
√	List of Abbreviations (if applicable)
√	Chapters / Content
√	Bibliography (or References)
√	All references in bibliography are cited in the thesis, especially in the chapter of literature review
√	Appendices (if applicable)
√	Weekly Log
√	Poster
√	Signed Turnitin Report (Plagiarism Check Result - Form Number: FM-IAD-005)
√	I agree 5 marks will be deducted due to incorrect format, declare wrongly the ticked of these items, and/or any dispute happening for these items in this report.

*Include this form (checklist) in the thesis (Bind together as the last page)

I, the author, have checked and confirmed all the items listed in the table are included in my report.

(Signature of Student)

Date: 25/4/2024