Ur-Ring

Urdu Finger Reader Ring

By

Aysha Mughal

Maha Farhat

Zehaib Mohsin

Submitted to the Faculty of Computer Software Engineering

National University of Sciences and Technology, Islamabad in partial fulfillment

For the requirements of a B.E. Degree in Computer Software Engineering

JUNE 2016

CERTIFICATE

Certified that work contained in this thesis “Ur-Ring” carried out by Aysha Mughal, Maha Farhat and Zehaib Mohsin, under the supervision of Dr. Seemab Latif for partial fulfillment of Degree of Bachelor of Software Engineering is correct and approved.

Approved by ______________________

Dr. Seemab Latif

Department of CS

Military College of Signals

Dated: JUNE 1, 2016

ABSTRACT

Visually impaired people report numerous difficulties with accessing printed text using existing technology, including problems with alignment, focus, accuracy, mobility and efficiency. Many people also fail to get basic education because of their disability.

We present a finger mounted device that assists the visually impaired and illiterate people with effectively and efficiently reading paper-printed text in Urdu language (i.e. Nastaleeq Font).

Ur-Ring is an index-finger wearable device that supports the visually impaired and illiterate in reading printed text by scanning with the finger. The camera captures the video in real time and after processing at the backend desktop application, it converts it into sound to be heard by the user. The device does not have many buttons or parts in order to provide a simpleinterface for users and easily orient the device.

According to an estimate about 1.5 million people are visually impaired and 45% people are illiterate in Pakistan, hence the target audience of our project are the visually impaired and illiterate people who face difficulty in reading printed text in their daily lives.

COPYRIGHTS NOTICE

No portion of the work presented in this dissertation has been submitted in support of another award or qualifications either at this institution or elsewhere.

DEDICATION

In the name of Allah, the Most Merciful, the Most Beneficent

To our parents, without whose unflinching support and cooperation,a work of this magnitude would not have been possible.Our supervisorDr. Seemab Latiffor her countless hours of support and encouragement.And special feelings of gratitude to our beloved siblings andfriends.

ACKNOWLEDGEMENTS

There is no success without the will of ALLAH Almighty. We are grateful to ALLAH, who has given us guidance, strength and enabled us to accomplish this task. Whatever we have achieved, we owe it to Him, in totality. We are also grateful to our parents and family and well-wishers for their admirable support and their critical reviews. We would like to thank our supervisor. Dr. SeemabLatif, for her continuous guidance and motivation throughout the course of our project. Without their help we would have not been able to accomplish anything.

Table of Contents

Table of Figures

Table of Tables

CHAPTER:1

INTRODUCTION

1 INTRODUCTION

1.1 Overview

Visually impaired people report numerous difficulties with accessing printed text using existing technology, including problems with alignment, focus, accuracy, mobility and efficiency. We present a finger mounted device that assists the visually impaired and illiterate people with effectively and efficiently reading paper-printed text in Urdu language (i.e. Nastaleeq Font).

1.2 Problem Statement

Some people with visual impairment find it difficult to access text documents in different situations. The existing technologies does not provide a suitable solution for this, especially for printed text written in Urdu language. We are making a finger mounted device for the problem stated above, which can read Urdu text (i.e. Nastaleeq Font) in real time.

1.3 Approach

The visually impaired people face many difficulties in their daily life because of their disability. Mostly people remain uneducated because they cannot read. The literacy rate in Pakistan is also only 55%, so we have devised a solution that illiterate people and people with visual impairment can also read printed text using this device.

1.4 Scope

The Ur-Ring will help the book reader specifically visually impaired people in reading the books written in Urdu language (Nastaleeq Font). It has 4 modules. The first module will make the video in real time through the camera mounted on the ring. The next module will do image segmentation on the video. In the third module the segmented images will be converted into the text using Urdu OCR. Finally in the last module the text will be converted to speech through the desktop application for the user to listen.

1.5 Aims &Objectives

The objectives of project include:

- Using software engineering techniques for gathering requirements during the development process, designing the software, implementing and testing requirements gathered.
- To learn real time processing of video using image segmentation techniques.
- To learn the optical character recognition techniques using OCR.
- To help the visual impaired people overcome their disability.

1.6 Organization

The first part of thesis is the abstract which describes the main details of Ur-Ring (Urdu Finger Reader Ring for Visually Impaired), followed by the introduction section which specifies the problem statement, approach, scope and objectives. The literature review section state the various resources read online before the commencement of the project. The design and development part illustrate the diagrams which describe the detailed design of the Ur-Ring its components, interface and data necessary for the implementation phase. The analysis and evaluation part give details of the black box testing, unit testing and system integration testing; actual results against expected results. The future work gives states the enhancements that can be applied to the application.

1.7 Deliverables

Deliverable Name	Deliverable Summary Description
Software Requirements Specification(SRS) Document	Complete Description of what the system will do, who will use it. Detailed description of functional and non-functional requirements and the system features.
Design Document	Complete description of how the system will be implemented i.e. the detailed design.
Code	Complete code.
Testing Document	The whole system is tested according to the specification described in the SRS document. Black box, unit and System integration testing is done.
Complete System	Complete working system.

Table 1-1 : Deliverables

CHAPTER: 2

LITERATURE REVIEW

2 LITERATURE REVIEW

The first step was to study in detail about the number of visually impaired people and the need of a reading device for the visually impaired people. The next step was to start building the device which will help the visually impaired and illiterate people in their daily lives. To do so, it was essential to study about the MATLAB techniques of working on real time video, image segmentation, OCR, natural language processing and text to speech synthesis.

2.1 Visually Impaired People and Loss of Productivity in Pakistan

An estimated 1.5 million visually impaired people in Pakistan- out of 40 to 45 million total are causing an annual loss of Rs63 billion in term of loss of productivity. It is projected that by the year 2020, there will be about 54 million blind persons with the age of 60 worldwide. As well as there are a total of 45% illiterate people in Pakistan which is also a hindrance in the development of the country. The major problem they face in their daily life is to read printed text, which is a great hindrance in their lives.

2.2 Devices for Reading Printed Texts

There are some devices already present for reading printed texts. There are pen readers, different apps that help the visually impaired people to read the printed text. But there are not many devices that help to read printed text in real time. A finger reader device is under development phase in MIT, but that is only for English language. As there is not much work done in Urdu so it’s a new device merrily for helping the users read printed text in Urdu language in real time.

2.3 Processing video in real time

Advance Video Processing requires an accurate segmentation of foreground objects. Therefore detecting regions of moving objects in scenes. Due to different nature of different backgrounds, some static, some dynamic and some quasi-stationary motion detection posed a difficult problem. A variety of algorithms have been proposed in the literature to suit scenarios in different environments. Different techniques are included in the class of parametric algorithm, and the major one is background subtraction algorithm.These techniques were assed to work in real time with videos.

2.4 Urdu Writing System

Urdu is an Indo-Aryan language and spoken by about 104 speakers all over the world. It isthe National language of Pakistan. Urdu is also spoken in Afghanistan, Bahrain, Bangladesh, Botswana, Fiji, Germany, Guyana, India, Malawi, Mauritius, Nepal, Norway, Oman, Qatar, Saudi Arabia, South Africa, Thailand, the UAE, the UK and Zambia. Urdu is written in Arabic script with some additional characters which are not present in Arabiclanguage. It is a bidirectional language. Words start form right side and numbers are written from left to right. This bidirectional nature of language increases the complexity of Urdu writing system.Urdu alphabet is comprised of 38 basic characters. These characters are joined together tomake words of the language. The very interesting phenomenon that Urdu script exhibits is change in basic shapes of characters. This change of shape is dependent on the position of the character in ligatures. This leads to four possible positional categories in which shapes of acharacter can be divided. These are initial, medial and final positions of a character in a ligatureand the isolated one. One character can acquire several shapes in each position. The shapes of a character are dependent on characters coming before and after it. Some characters can only come at the final position, so they can be called ligature finishers.

2.5 OCR System

The main objective of OCR system is to import printed text into machine, and enabling user to edit it with minimal effort. But there is a lot more modern usage of this application. Some ofthem are listed below.

1) OCR can create editable text document with little human effort that can save space andtime as well. It can be used to publish the online content as text instead of images. Itcan help not only in saving the storage space but also a huge collection of Urdu corporacan be generated. It can aid language processing applications which requires reasonableamount of data for statistical calculations.
2) Modern search engines can be equipped with OCR system. Hence, user can search onlineimages with required text in them. The significance increases with the fact that still majorityof news websites use images to display Urdu text. Online content is hard to search.
3) Text to speech application can use OCR system to get the textual form of an image out of a document. It can help to enhance the applicability of TTS system. Applications developed to aid visually impaired people can use OCR system to increase their versatility.

2.6 Urdu OCR

Development of OCR system for Urdu language has remained a challenging task for Urduresearchers during last few years. The main reason for its challenging nature is complexityof Urdu writing system. As discussed earlier Urdu is widely printed in

Nastaleeq writing style. A detailed survey carried out by CRULP showed that almost all of the books (Printed Text) are published in Noori Nastaleeq font. The complex nature of Nastaleeq was analyzed in Section 2.6. The cursive and context based shape changing phenomenon of Urdu makes the isolation of characters harder as compared to Latin script. So segmentation itself is proven to be an error prone and difficult task. Hence, most of the developed systems are based on ligature recognition methodology.

A segmentation free approach was used by Hussain in which ligature segmentation was not involved. The main bodies were extracted from diacritics and then statistical features were acquired. These identified features were used to recognize the ligature through a multi-tier neural network model. The trained system showed accuracy higher than 95% on test samples of two character ligatures. In 2005, same feature extraction technique was used to develop a ligature based recognition system. Spatial features were extracted from the isolated ligatures of Urdu text and then Hidden Morkov Model was trained for the recognition process. Nearly 1500 ligatures were trained and the system showed accuracy above 92%. The limitation of this system was shape dependence, which means only trained ligatures were recognized by that OCR. To cover current corpus of language, all ligatures have to be trained to the system with high accuracy which itself is a tedious work. In order to cut down the training work, segmentation was necessary to encounter the shape changing phenomenon of Urdu. In 2007, research was extended to develop a segmentation based Noori Nastaleeq OCR system. It was the first attempt to divide ligatures of Nastaleeq into smaller pieces for recognition purpose. The proposed technique has following major modules.

- Line Separation
- Ligature and Diacritic Separation

• Thinning

• Segmentation

• Framing

• Recognition

• Post Processing

This system was trained and tested with 60 different ligatures, made by combinations of 6 Urdu classes (Alif, Bay, Dal, Swad, Ain, and Choti Ye). The research showed very promising results by using workable solutions for Urdu OCR. The system is discussed here in details. The line separation process and baseline detection has been done by finding the vertical histogram of image. Clear white spaces are present between two text lines in an image. Vertical histogram shows a band of zero black pixel values and text line is separated from there. After separation of lines, the maximum value of vertical histogram is marked as the baseline for text. Instead of using single pixel baseline, a band of 5-10 pixels is chosen to fulfill the purpose.

The baseline identification was required to isolate main bodies of ligatures from their diacritic marks. All of the connected components present on baseline are considered as main bodies and rest of them are marked as diacritics. In a more recent research it has been experimented that the method is error prone because some diacritics reside on baseline as well. The proposed solution of this problem is setting a threshold on size of connected body to distinguish between diacritics and main bodies. After identification of main bodies, next step in the given research is segmentation. The technique of segmentation is based on single pixel stroke so thinning of main bodies is introduced. The detail of thinning process is not elaborated in the research but it was done with the help of Matlab2. The discussed segmentation technique works with stroke cutting at the junction point i.e. during the traversal of thinned body, when more than one neighboring pixel is found a segment is separated. Some problem of under and over segmentation of this technique has also been highlighted by the researchers of this study.

As soon as segments are acquired, the next proposed step was performing framing operation to divide the segment in equal windows. A frame window of 8*8 pixels has been. An application which provides image processing tools used for division of segments. A static disjoint framing technique is chosen and thin stroke is kept in the middle of 8*8 sized window. Frames are numbered with the order of segment stroke sequence. After filling all frames DCT is applied and output is stored in a single file. The system was trained to recognize only classes which are used to construct a particular ligature. Character recognition was not performed in this research. 6 different classes (ابدعی) were used to construct different ligatures and system was trained for those ligatures. Segmentation yielded 60 unique segments and HMM model was developed. A rule based post processing unit was developed to construct the ligature. Test results of all classes have been reported in the given study. Almost all samples of ا and د have been recognized showing accuracy of 99%. Classes of ع and ب shows nearly 95% accuracy on the other hand ص showed 80% accuracy.

The research discussed above shows very encouraging results and a devised method for segmentation based Urdu OCR system. One of the challenges described by the researcher of that study is lengthy training procedure. Grouping of segments, and manual training took very long time for building HMM model. Also the absence of sufficient features of thinned stroke was pointed out.

2.7 Text to Speech

Text-to-speech synthesis is logically divided into two stages. The first stage takes raw text input, processes it and converts it into precise phonetic string to be spoken, appropriately annotated with prosodic markers (e.g. stress and intonation). The second stage takes this phonetic representation of speech and generates the appropriate digital signal using a particular synthesis technique. These stages may be referred to as Natural Language Processing (NLP) and Speech Synthesis (SS). For SS, formant based techniques or diaphone based techniques are normally employed and are generally script independent (as they are only dependent on temporal and spectral acoustic properties of the language and take input in script-neutral form, e.g. in IPA). However, NLP is very dependent on cultural and linguistic specific usage of script. NLP may also be divided into further parts. The first component is dedicated to pre-processing, ’cleaning’ and normalizing input text. Once the input text is normalized, the second component does phonological processing to generate a more precise phonetic string to be spoken. One of the first tasks in the Phonological Processing

Component is to convert the input text into a phonemic string using Letter-to-Sound (LTS) rules. This string is then eventually converted to precise phonetic transcription after application of sound change rules and other annotations.

.

CHAPTER: 3

OVERALL DESCRIPTION

3 OVERALL DESCRIPTION

This part of the document contains information about the product, its features, perspective, users' characteristics and constraints.

3.1 Product Perspective

Ur-Ring is basically built to help the reader in reading text written in Urdu language (Nastaleeq font) especially for visually impaired and elderly people as well as illiterate and people having eyesight problems which causes difficulty in reading.

3.2 Product Features

The main features of Ur-Ring are highlighted below:

1. A camera mounted on the ring to be worn on index finger.
2. Video made in real time by camera mounted on the ring.
3. Segmenting video to images.
4. Uses Urdu OCR to convert image to text.
5. Recognizes only Urdu Nastaleeq font.
6. Converting text to speech using text to speech conversion.
7. Desktop Application for image processing and text to speech conversion.
8. Uses system speakers to produce sound.

3.3 User Classes and Characteristics

3.3.1 Summary of User Classes

The following section describes the types of users of the Ur-Ring. There are explanations of the user followed by the interactions the user(s) shall be able to make with the software.

3.3.2 Book Reader

The user in this is basically the Book Reader who are visually impaired, illiterate or elderly people, wearing the ring with the camera mounted on it to make video and give it as input to the system. The reader will place the finger wearing the ring in line with the text to make video of the text and the system after processing will read it out loud for the reader to hear.

3.4 Operating Environment

3.4.1 Hardware

The Ur-Ring operates, either directly or indirectly, with the following external hardware:

- **Camera Mounted on Ring**: The camera is used to make video in real time when aligned with the text while the user is reading.
- **Data Cable**: The data will be transferred from the data cable to the computer.
- **Computer**: Software installed which will process the data obtained through the camera and will convert it into speech.
- **Speakers**: Computer speakers will be used to hear the sound produced through software.

3.4.2 Software

- Windows: 7, 8, 8.1, 10.
- IDE: MATLAB

3.5 Design and Implementation Constraints

- Ur-Ring will only process captured video when aligned with the text.
- It will only read the text written in Urdu Nastaleeq font.
- It will only read text having font size 14 to 44.
- It can only be used by one person at a time.
- Input may contain noise along with the data.
- There will be time delay and less fluency in speech.

3.6 User Documentation

There will be a user manual provided to help the users in getting familiar with the application. It will also provide users with the description of the interface and how the application works.

3.7 System Features

This section describes in detail the system features of the Ur-Ring. System features are classified into following use cases:

1. Video Input
2. Sound Alert
3. Divide into Frames
4. Convert to text

5. Convert Text to speech

3.7.1 Video Input

3.7.1.1 Description and Priority

This will take video input from the camera mounted on the ring when the camera is attached to the system.

This function is of high priority because without the video no further processing can take place.

3.7.1.2 Stimulus/Response Sequences

- The camera should make the video.
- The video made in real time should be sent for further processing.

3.7.1.3 Functional Requirements

REQ-1: The camera should be attached to the system and turned on.

REQ-2: The finger on which the device is worn should be pointing to the paper with printed text.

REQ-3: The camera should be aligned with the text to be read.

REQ-4: The video made in real time should be sent for further processing.

3.7.2 Sound Alert

3.7.2.1 Description and Priority

The sound alert will be produced when the camera is not aligned with the text or there is the end of line and no text to be read further.

This has a medium priority as it will only take place if the user deviates from the line or there is the end of line.

3.7.2.2 Stimulus/Response Sequence

- The video should be taken as input.
- The sound alert should be produced if the system faces the conditions described above.
- The process should continue as usual if the camera is aligned again.

3.7.2.3 Functional Requirements

REQ-1: The camera should be making video.

REQ-2: Input should be fed to the system to check whether the input is aligned with the text.

REQ-3: The system should produce a beeping sound as an alert for the user.

3.7.3 Divide into Frames

3.7.3.1 Description and Priority

The video captured in the previous step should be segmented into the frames to be fed to OCR for further processing.

This function has also high priority because the segmented input is to be fed to the OCR for further processing in real time.

3.7.3.2 Stimulus/Response Sequence

- Get the video as input from the camera.
- Divide the video into frames by segmentation techniques.
- Feed the segments to OCR for further processing.

3.7.3.3 Functional Requirements

REQ-1: The Camera should be making video in real time.

REQ-2: The camera should be aligned with the text.

REQ-3: Segment the video into images.

3.7.4 Convert to Text

3.7.4.1 Description and Priority

The input from the segmented part should be converted into text using the OCR.

This has a high priority since the optical character recognition is done in this part which is converted to text for further processing.

3.7.4.2 Stimulus/Response Sequence

- The Segmented image should be fed to the OCR.
- The OCR should be able to recognize the text in the image.
- The OCR should convert the image to text.

3.7.4.3 Functional Requirements

REQ-1: The images should be segmented properly.

REQ-2: OCR should recognize the image with text accurately.

REQ-3: The OCR should convert the image to text for further processing.

3.7.5 Convert text to Speech

3.7.5.1 Description and Priority

The output from the last step is taken as input in the next step. The Image with text is converted into the sound. The sound is produced through the desktop speakers to be heard by the user.

This function has high priority as the user will only be able to hear the text if this function is completed successfully.

3.7.5.2 Stimulus/Response Sequence

- Get the output from the OCR as text.
- Convert the text into speech.

3.7.5.3 Functional Requirements:

REQ-1: The OCR should provide accurate output of recognized text.

REQ-2: The text should be accurately converted to speech.

REQ-3: The system speakers should be used to produce sound.

3.8 External Interface Requirements

3.8.1 User Interfaces

- Image converted to text will be displayed on the screen.

3.8.2 Hardware Interfaces

- Video input will be taken through camera.
- Communication between software and camera will be done using data cable.
- Speakers of computer will be used to get speech output.

3.8.3 Software Interfaces

- To convert images to text OCR will be used.
- To convert text to sound the text to speech system will be used.

3.8.4 Communications Interfaces

- Data cable will be used as medium of communication between camera and software. This data cable will transfer video from camera to computer.
- This video will be converted to frames which will then be sent to OCR as input.
- Frames will be then converted to text by OCR.
- Text will be segmented and converted to speech by text to speech system.
- Speech output will be given by inbuilt speakers of computer.

3.9 Other Nonfunctional Requirements

3.9.1 Safety Requirements

The use of the software product has no harms whatsoever; nor does it have any possibility of loss or damage that might be inflicted. If the app crashes there will be no change in its original processing algorithms.

3.9.2 Security Requirements

Application running on the computer should not need any additional information other than the collected data from the user or already present data. There are no connections to other devices or severs so no data will be sent or received or used in any way.

3.9.3 Software Quality Attributes

3.9.3.1 Usability

The graphical user interface of app is to be designed with usability as the first priority. The app will be presented and organized in a manner that is both visually appealing and easy for the user to navigate.

3.9.3.2 Accuracy

To ensure reliability and correctness, there will be zero tolerance for errors in the algorithm that computes results.

3.9.3.3 Portability

The device is not easily portable because one cannot carry the whole system with the device on which the processing is being done. It is not impossible because if the device is attached with a laptop which can perform the processing then this issue can be resolved.

3.9.3.4 Availability

The application will be available from boot to shutdown, provided the system does not crash.

3.9.3.5 Flexibility

The design and architecture of the application will be flexible enough for catering any new requirements, if any at some later stage or for the application enhancement.

3.9.3.6 Data Integrity

If the app crashes during addition, deletion or editing there will be no changes.

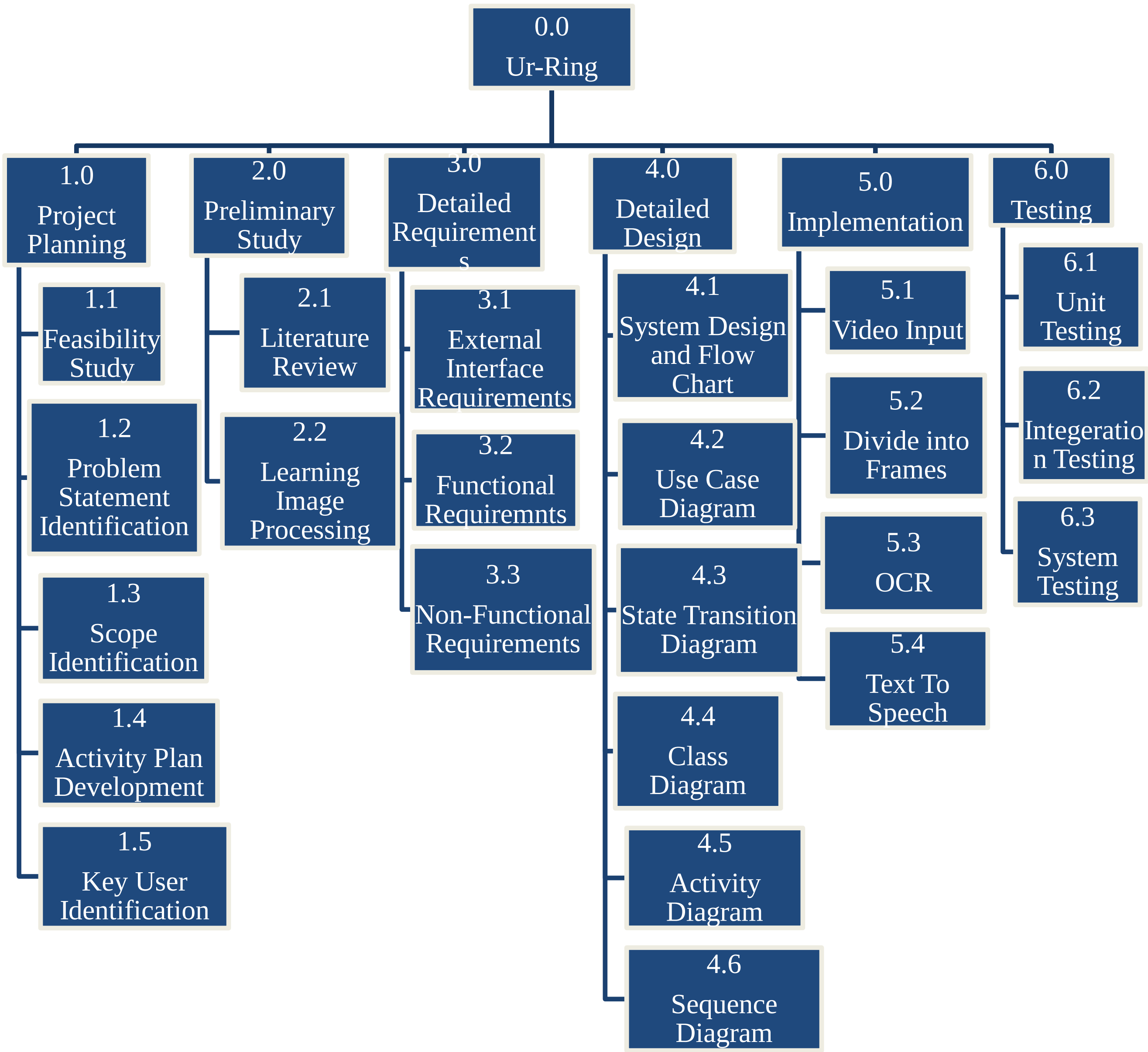
3.9.3.7 Scalability

The application is expected to handle one user at a time. The application can be used by a single device attached with the system.

CHAPTER: 4

DESIGN AND DEVELOPMENT

4 DESIGN AND DEVELOPMENT



4.1 Work breakdown structure

Figure 4-1: Work Breakdown Structure

4.2 Flow Chart

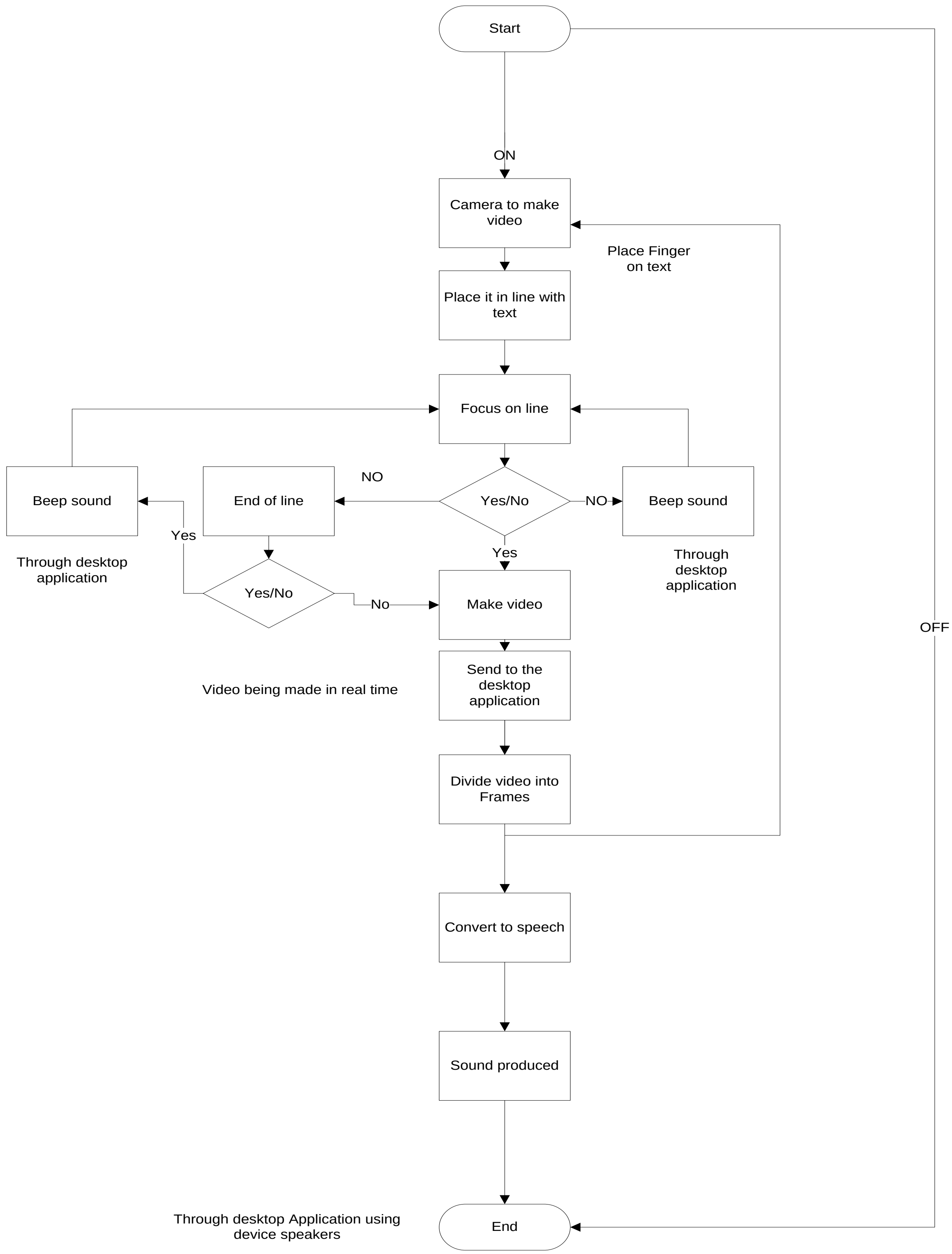


Figure 4-2: Flow Chart

4.3 System Architecture Description

Detailed description of system architecture and design pattern which this system is going to use is discussed later in the document in section ‘Design Decisions and Tradeoffs’.

This Section overview of application, its higher and lower levels details and user interfaces.

4.3.1 Overview of Modules/Components:

The purpose of the component diagram is to show the relationship between different components in a system. The main components of the system and their interactions are shown in the component diagram below.

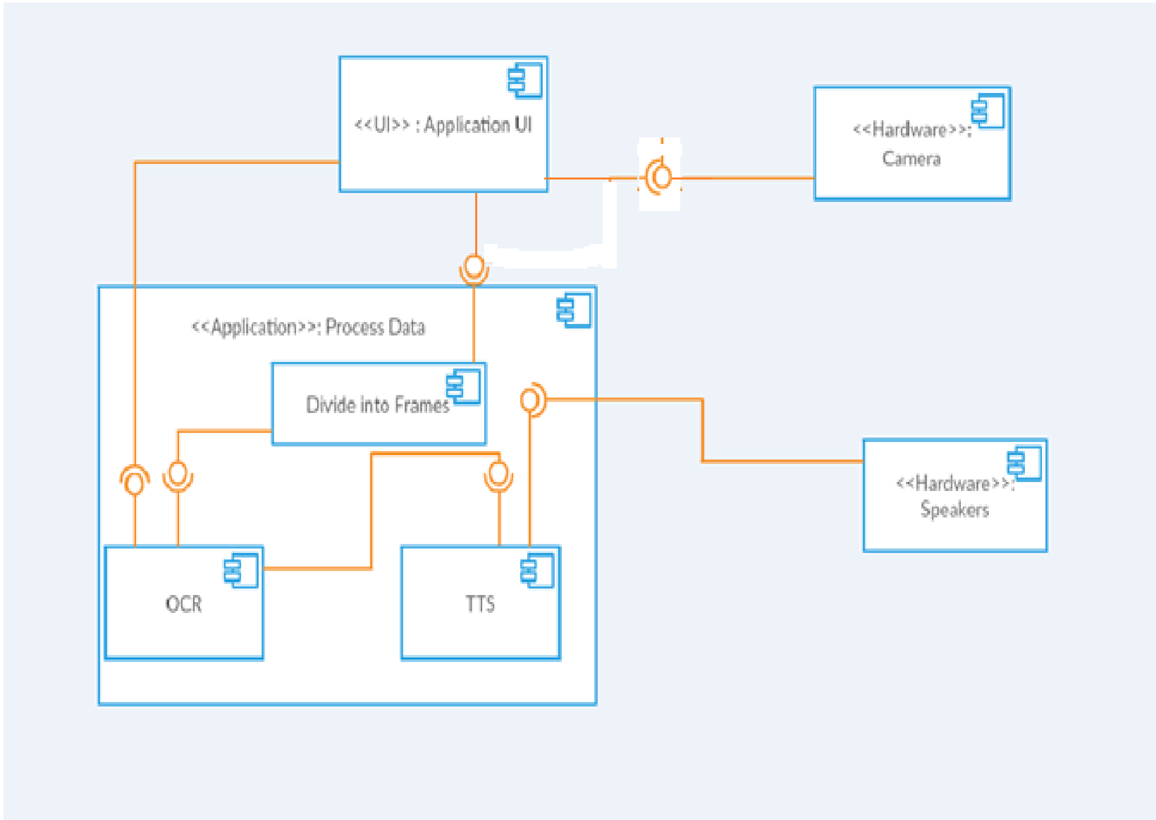


Figure 4-3: Component Diagram

Ur-Ring comprises of following components:

1. UI Application
2. Camera
3. Process Data
   - Divide into frames
   - OCR

- TTS

4. Speakers

The **Camera** gives input to the **Process Data** module. There are three sub components of Process Data module. The video from the camera goes to the **Divide into frames** component, where the video is split into frames for further processing. These frames are then fed to the **OCR** component. This component converts the input frames into text. This text is further sent to the **TTS** component for converting text to speech as well as to the **UI Application** module to display the converted text on the screen. The output of TTS component is then sent to the **Speakers** from which sound is produced.

4.4 Structure and Relationships:

This section covers the overall technical description of Ur-Ring (refer Section 1.3). It shows the working of application in perspective of different point-of-views and also shows relationships between different components.

4.4.1 System Block Diagram

This diagram shows the higher level description of the application. It shows generic working of the application and interaction with the user. In this the user will make video through the camera mounted on the ring which will be send to the desktop application for further processing. If the user deviates from the line, or if the line ends while making the video, beep sound will be generated as an alert. Once the video is sent to the desktop it will be divided into frames which will be fed as an input to the OCR. The OCR will convert the image to text which after further processing will be converted into speech to be heard by the user.

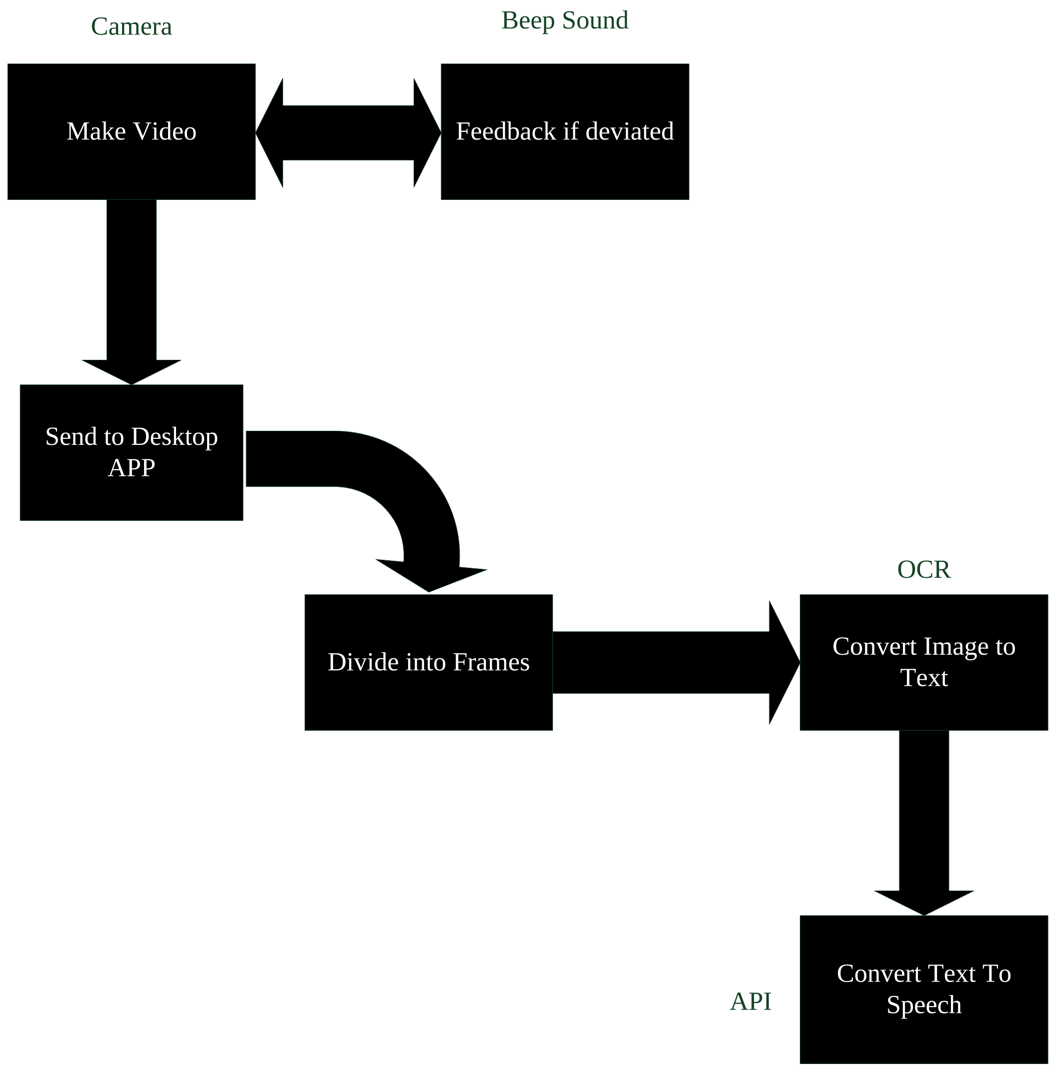


Figure 4-4: Block Diagram

4.4.2 Architectural Design

We have followed the Pipe and Filter architectural design for our application in this as shown below:

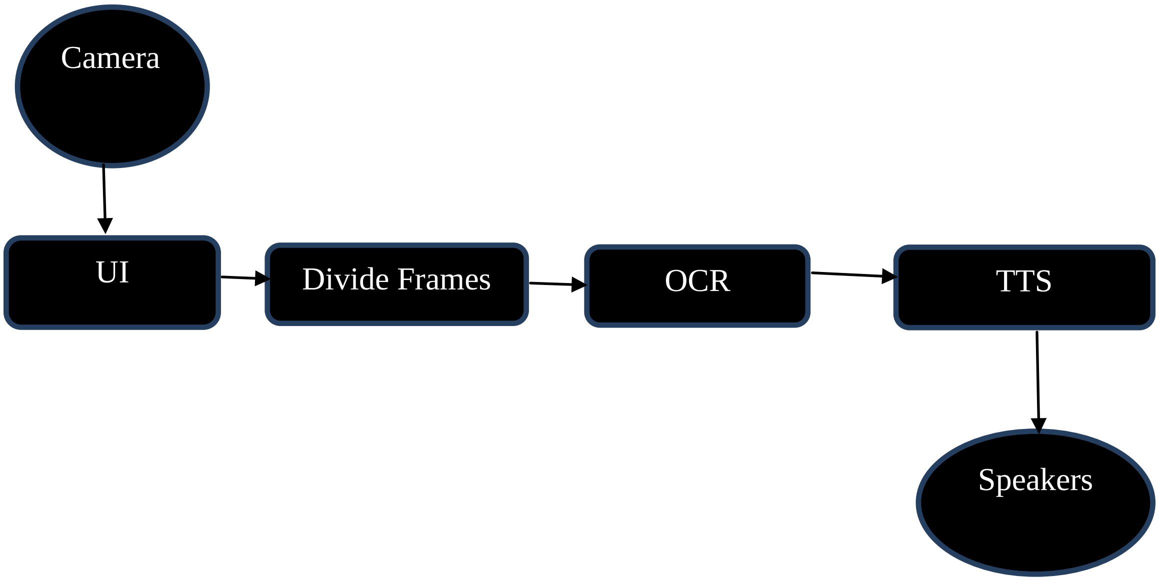


Figure 4-5: Pipes and Filter Diagram

The filter transforms or filters the data it receives via the pipes with which it is connected. A filter can have any number of input pipes and any number of output pipes. The pipe is the connector that passes data from one filter to the next. It is a directional stream of data that is usually implemented by a data buffer to store all data, until the next filter has time to process it. The pump or producer is the data source. It can be a static text file, or a keyboard input device, continuously creating new data. The sink or consumer is the data target. It can be another file, a database, or a computer screen.

Because one component is dependent on previous component hence data cannot be processed until whole processing has been done by the previous component on data. This is the reason we are using the pipes and filters architecture. In the diagram shown above the **Pump** is the camera**, Filters** are UI, Divide Frames, OCR and TTS and **Sink** is Speakers and the arrows represents the **Pipes.**

4.4.3 User View (Use Case Diagram)

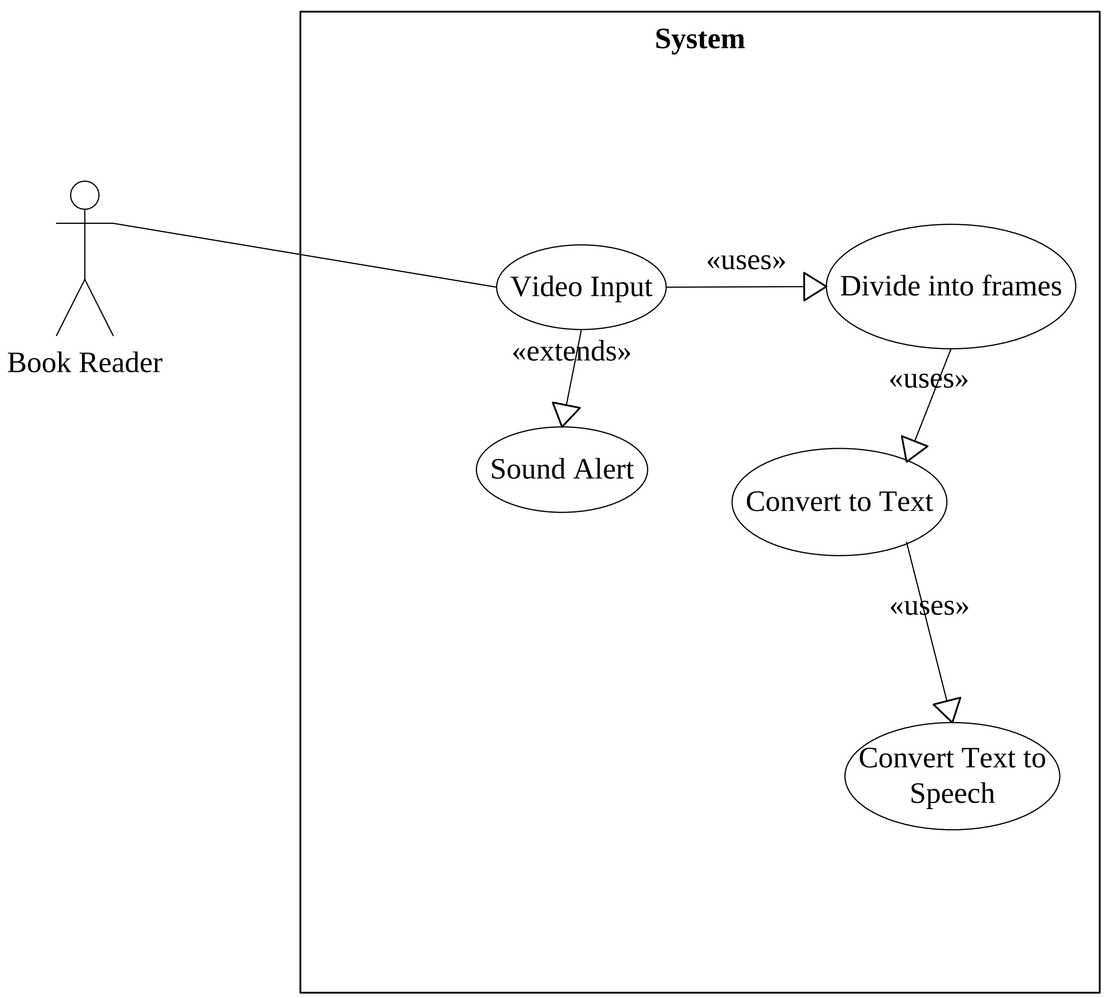


Figure 4-6: Use Case Diagram

Actors:

- Book Reader

Use Cases:

- Video Input
- Divide into Frames
- Convert to Text
- Convert Text to Speech
- Sound Alert

Use Case Description:

- ### Use Case 1: Video Input

Use Case Requirement: The camera mounted on the ring will capture the video in real time. This video will be used for further conversion of text into speech.
Use Case Paths • Normal:Video sent for processing. • Exceptional:Beep Sound is produced
Normal Path: Video sent for processing
Externals: Data Cable
Preconditions: The camera captures the video in real time.
Interactions: The captured video is sent to the system for processing.
Postconditions: Captured video is divided into frames
Categorization • **Frequency**: High

• **Criticality**: High • **Probability of Defects**: Medium • **Risk**: High
Exceptional Path: Sound Alert is produced
Externals: Speakers
Preconditions: The camera is not aligned with the text while video is being captured or end of line.
Interaction: An error signal is sent to the system
Postconditions: The beep sound is produced through the speakers of the computer.
Categorization • **Frequency**: High • **Criticality**: High • **Probability of Defects**: Low • **Risk**: High

Table 4-1 : Video Input

- **Use Case 2: Divide into Frames**

Use Case Requirement: The video captured through the camera when successfully sent for processing is divided into frames. These frames are then sent for further processing into the system.
Use Case Paths • Normal:Successfully divided into frames
Normal Path: Successfully divided into frames
Externals: Video input
Preconditions: The video is sent to the system for processing.

Postconditions: The divided frames are used for further processing.
Categorization • **Frequency**: High • **Criticality**: Medium • **Probability of Defects**: Medium • **Risk**: Medium

Table 4-2: Divide into Frames

- **Use Case 3: Convert to Text**

Use Case Requirement: The divided frames obtained are used to convert the image to Text using OCR. The converted text will be sent for further processing for its conversion into speech.
Use Case Paths • Normal:Image Converted to text • Exceptional:OCR generates error
Normal Path: Image converted to text
Externals: Frames
Preconditions: Image are obtained to convert to text.
Interactions: OCR is used for conversion
Postconditions: Text obtained is used for further processing.
Categorization • **Frequency**: High • **Criticality**: High • **Probability of Defects**: High • **Risk**: High
Exceptional Path: OCR generates Error

Externals: Frames
Preconditions: Image are obtained to convert to text.
Interactions: OCR is unable to convert image to text.
Postconditions: Error generated and beep sound is produced.
Categorization • **Frequency**: Low • **Criticality**: Low • **Probability of Defects**: Low • **Risk**: Low

Table 4-3: Convert to Text

- **Use Case 5: Convert Text to Speech**

Use Case Requirement: The text obtained after segmentation is utilized by the system for conversion into speech.
Use Case Paths • Normal:Sound is produced
Normal Path: Sound is produced
Externals: Text
Preconditions: The text is obtained for conversion.
Interactions: System converts text into speech
Post conditions: The user is able to hear the text.
Categorization • **Frequency**: High • **Criticality**: High • **Probability of Defects**: Medium

• **Risk**: High

Table 4-4: Convert Text to Speech

- **Use Case 6: Sound Alert**

Use Case Requirement: The video Input Obtained by the Camera. The video will be checked by the margins specified.
Use Case Paths • Normal: Beep Sound is produced
Normal Path: Beep Sound is produced
Externals: Sound
Preconditions: The video obtained is not in line with the text or end of line.
Interactions: Sound alert is produced by the system speakers.
Post conditions: The user repositions the ring finger for capturing the video.
Categorization • **Frequency**: High • **Criticality**: High • **Probability of Defects**: High • **Risk**: High

Table 4-5: Sound Alert

4.4.4 Sequence Diagram

The sequence diagram of the UR-Ring is given below:

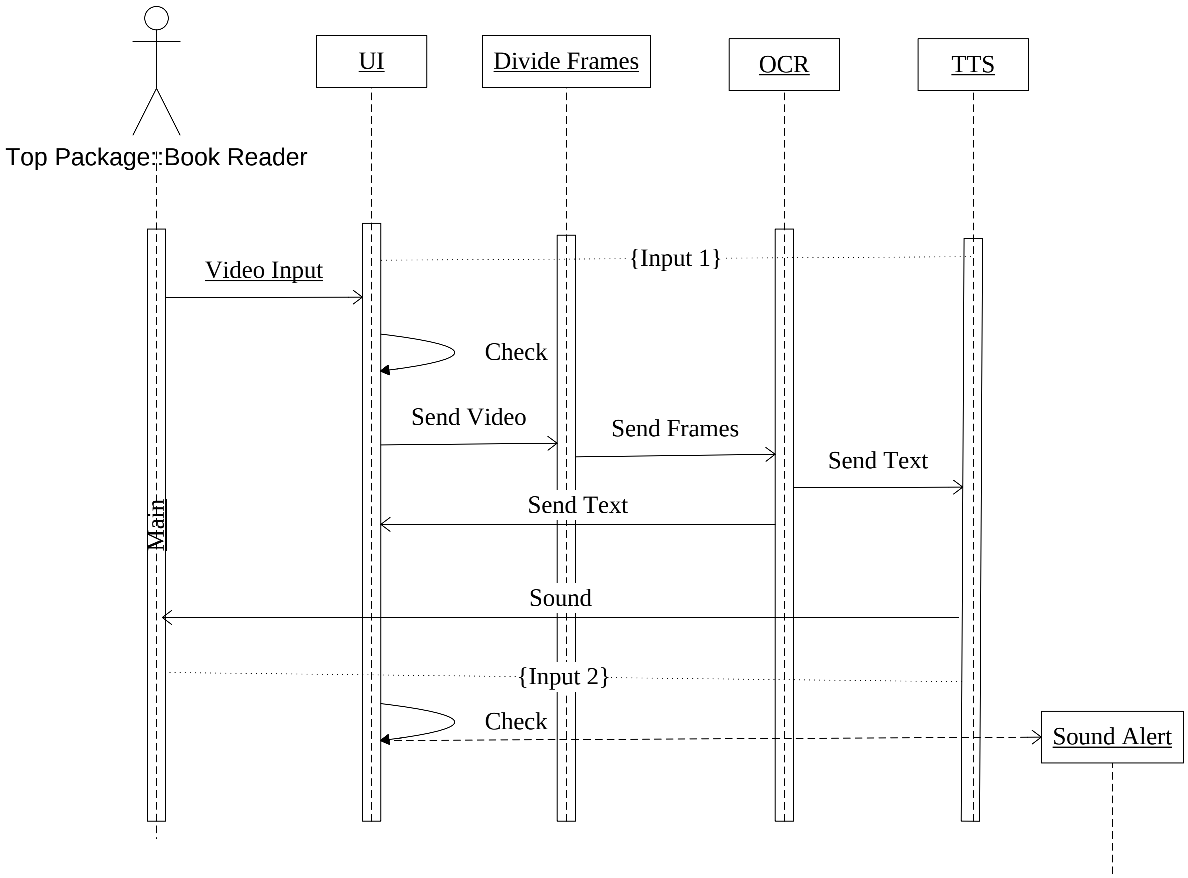


Figure 4-7: Sequence Diagram

4.4.5 Logical View (State Transition Diagram)

The State Transition diagram is shown in the figure below:

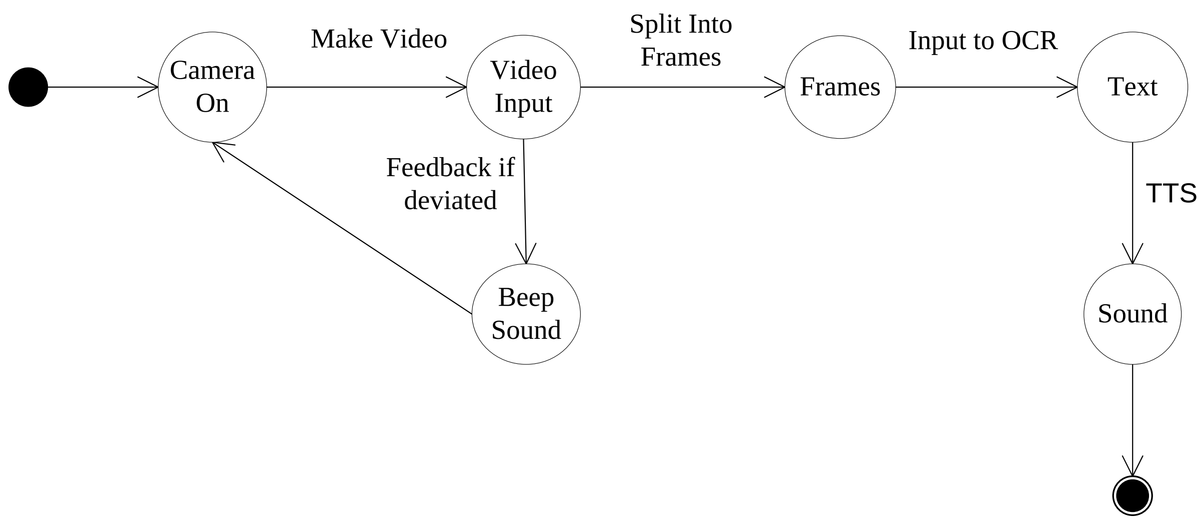


Figure 4-8: State Transition Diagram

4.4.6 Dynamic View (Activity Diagram)

In activity diagram, the dynamic view of the system is shown. All the activities are shown concurrently with their respective start and end states.

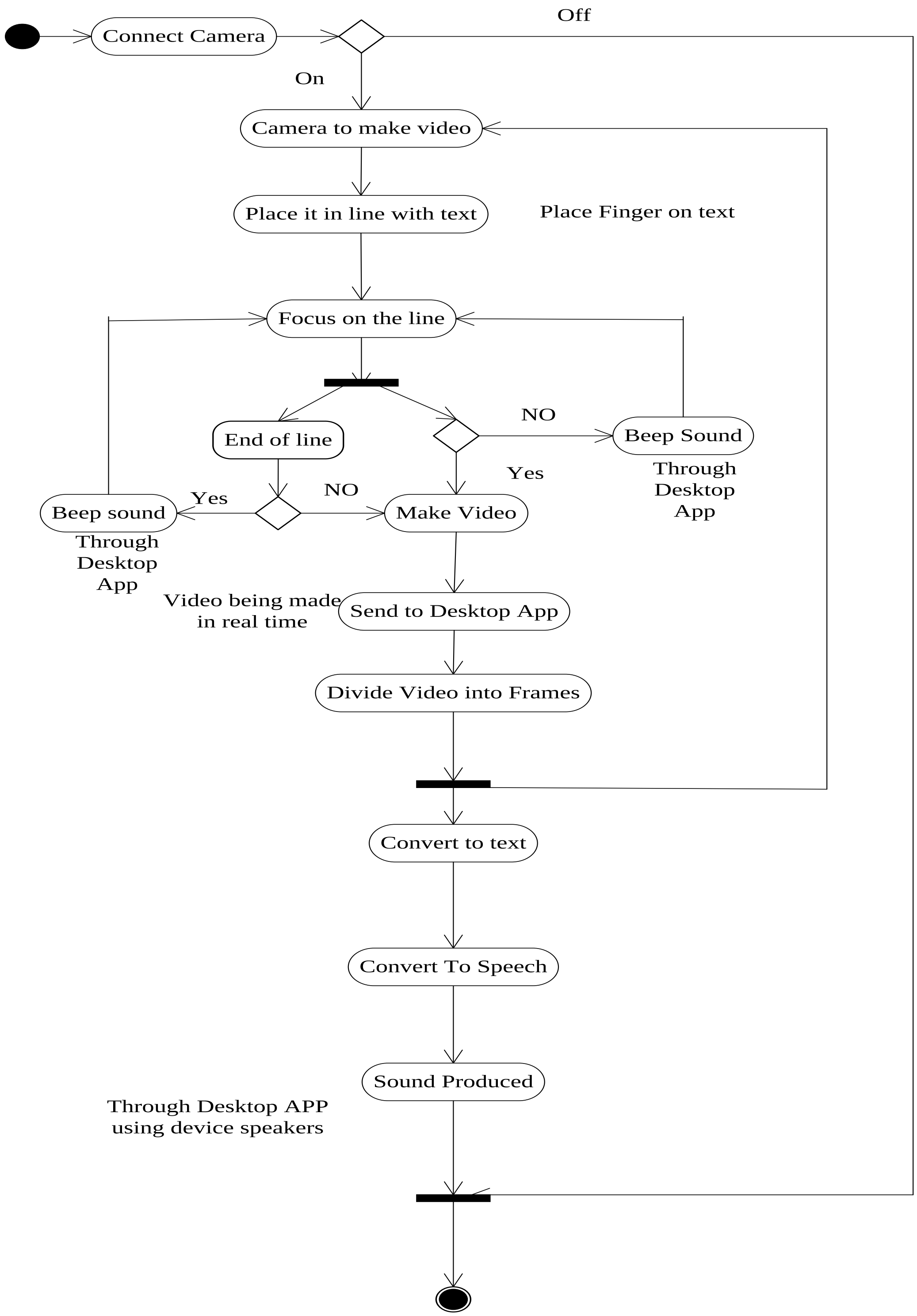


Figure 4-9: Activity Diagram

4.4.7 Implementation View (Class Diagram)

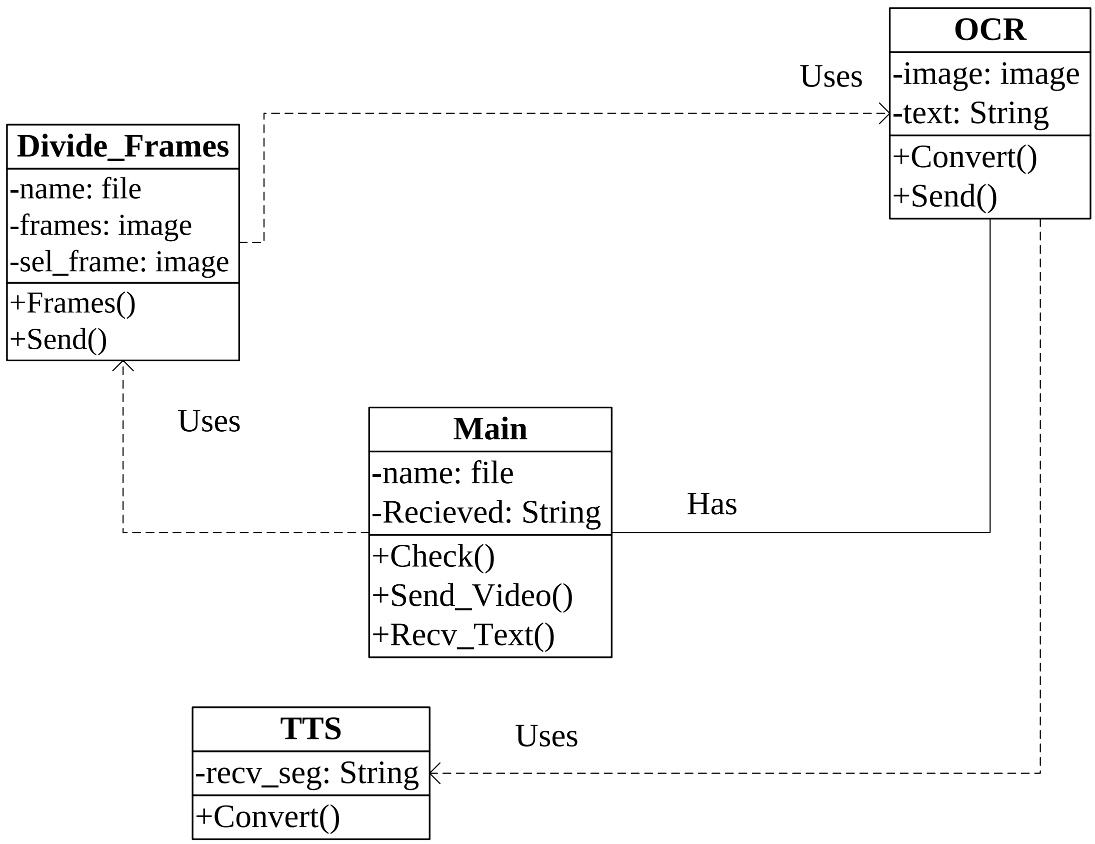


Figure 4-10: Class Diagram

Classes	Description
Main	The main class of the application, which'll be executed first in when the application is run. It receives the video from the camera. The check function checks whether the video being made is in line with

	the text or not. This video is then send to the Divide_Frames class for further processing.
Divide_Frame	This class first split the video into frames Then the frames are sent to OCR class for further processing.
OCR	It converts the frames into text and send the output to the main class to be displayed through the app to the user as well as to the TTS class.
TTS	This class receives the text from the OCR class which converts it into the speech to be heard by the user.

Table 4-6: Class Table

4.4.8 Structure Chart

This chart shows the breakdown of the application to its lowest manageable levels. It shows the modules and their corresponding functions which this application will implement. This chart basically shows the structure breakdown of the application starting from main modules to specific functions.

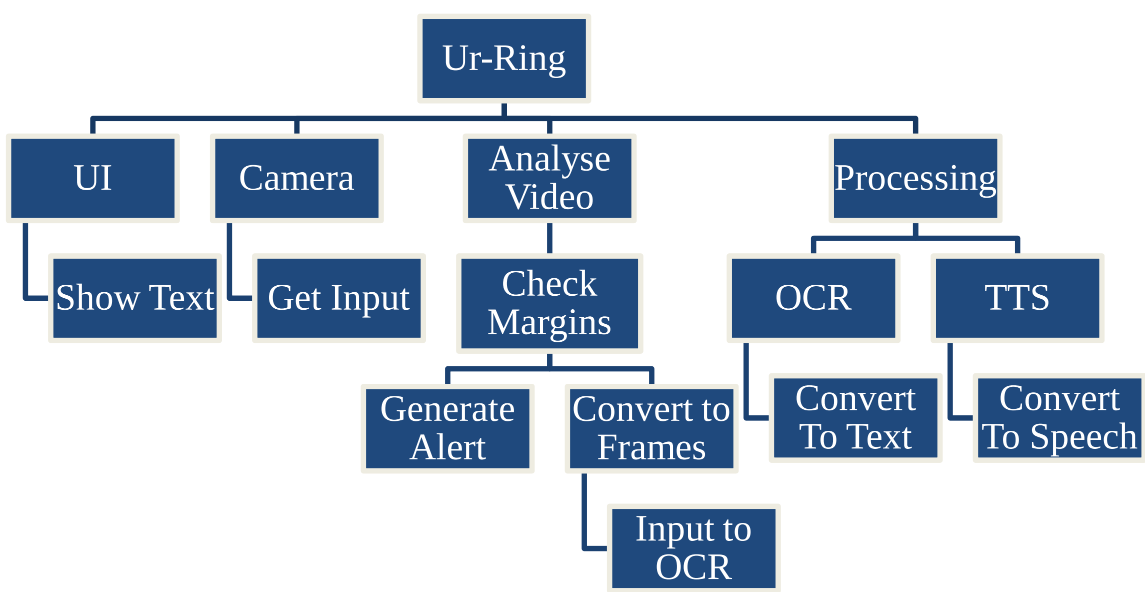


Figure 4-11: Structure Chart

4.5 User Interface

The user Interface of the Ur-Ring is as follows:

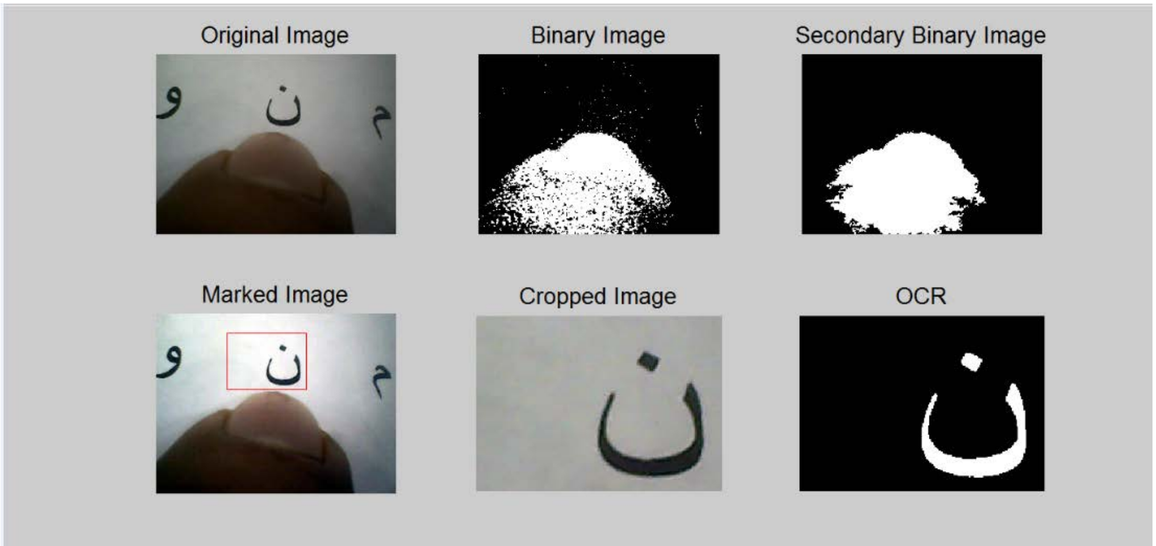


Figure 4-12: GUI

The Screen will show the original image being obtained through the video in the left most frame, in the next frame it will show the conversion into binary image, after that secondary binary image, in the bottom line the left most frame shows the marked window, the next one the cropped image and finally the input image that is to be fed to OCR.

4.6 Detailed Description of Components:

4.6.1 Application UI

Identification	*Name:* Application UI *Location:* Presentation layer of the system architecture
Type	UI component
Purpose	The User is able to see the converted text which is being read on the screen while the video is being made by the camera mounted on the ring. This component fulfills following functional requirements (as specified in SRS Document) related to user interface in the application: **REQ-1:** Application should be able to show the text of the line on which the user has placed its finger. **REQ-2:** If the user deviates from the line or end of line sound alert should be generated by the system. **REQ-3:**The application should be able to get the video being made in real time and send it for further processing in the system.
Function	This component has two functions, the first one is to send video being captured in real time for further processing in the system. The second one is to display the output of image conversion to the text on the screen.
Subordinates	This component has two subordinates; one is responsible for input, the other for output. The input subordinate satisfies all functional requirements that require user input.

	The output subordinate satisfies all functional requirements that provide output.
Dependencies	It interacts with *Process Data*, whenever a user interacts with the application. It also gets its Input from the hardware component Camera.
Interfaces	The User interface is part of this section. It will display the text and produce sound alert when needed.
Resources	**Hardware:** Camera will be used to get video input. **Software:** MATLAB and C++ code
Processing	Take Input from the camera making video. Provide feedback if deviated.
Data	Video

Table 4-7: Application UI

4.6.2 Camera

Identification	*Name:*Camera *Location:* Presentation layer of the system architecture
Type	Hardware component
Purpose	Following functional requirements mentioned in SRS are fulfilled by this component: **REQ-4:**The application should get input video in real time through the camera mounted on the ring.
Function	This component makes video in real time of the line under which finger is placed to be read.
Subordinates	N/A
Dependencies	This component isn’t dependent on any other component, but provide

	video to the UI Application for further processing.
Interfaces	The video will be sent to the UI Application.
Resources	**Hardware:** Camera
Processing	Camera records video and send it to the UI Application for further processing.
Data	Video

Table 4-8: Camera

4.6.3 Process Data

This component has 3 sub-components:

4.6.3.1 Divide Frames

Identification	*Name:*Divide Frames *Location:* Application Logic layer of the system architecture
Type	Sub-component
Purpose	Following functional requirements mentioned in SRS are fulfilled by this sub-component: **REQ-5:**It is going to split video into frames. **REQ-6:** The new frame will be sent further for conversion into text.
Function	The function of this sub component is to provide frames of video in form of images to be used to convert into text for further processing.
Subordinates	It has one subordinate which splits the video into images and feeds it to the OCR.
Dependencies	This is dependent on the UI Application to provide the video.

	Other components are also dependent on it for further processing.
Interfaces	The UI Application provides it the video for processing.
Resources	Hardware: Camera Software: MATLAB Code
Processing	The component splits the video into frames. The frames obtained are used for further processing.
Data	Video

Table 4-9: Divide Frames

4.6.3.2 OCR

Identification	*Name:*OCR *Location:* Application Logic layer of the system architecture
Type	Sub-component
Purpose	Following functional requirements mentioned in SRS are fulfilled by this sub-component: **REQ-7:**Convert image frames into text.
Function	The function of this sub component is to convert the received frames into text.
Subordinates	N/A
Dependencies	This is dependent on the Divide frames component to provide frames for further processing, whereas UI application is dependent on it to provide text to be displayed on the screen.
Interfaces	The UI Application to display the converted text on the screen.
Resources	Hardware: Camera Software: MATLAB Code
Processing	The component converts the image into text to be sent for further

	processing.
Data	Text

Table 4-20: OCR

4.6.3.3 TTS

Identification	*Name:*TTS *Location:* Application Logic layer of the system architecture
Type	Sub-component
Purpose	Following functional requirements mentioned in SRS are fulfilled by this sub-component: **REQ-10:**Converts text to speech.
Function	The function of this sub component is convert text to speech to be heard by the user.
Subordinates	N/A
Dependencies	This is dependent on the OCR component to provide text on which the processing is to be done.
Interfaces	N/A
Resources	Hardware: Speakers Software: C++ code
Processing	The text is converted to the speech to be heard by the user.
Data	Text

Table 4-31: TTS

4.6.4 Speakers

Identification	*Name:*Speakers *Location:* Presentation layer of the system architecture
Type	Hardware component
Purpose	Following functional requirements mentioned in SRS are fulfilled by this component: **REQ-11:**The converted text should be heard by the speakers of the system.
Function	This component enables the user to hear the text.
Subordinates	N/A
Dependencies	This component is dependent on the TTS to provide the data to be voiced through the speakers.
Interfaces	N/A
Resources	**Hardware:** Speakers
Processing	Using the system speakers to voice the converted text and enable the user to hear it.
Data	Converted text to sound

Table 4-42: Speakers

4.7 Reuse and Relationships to Other Components

The Ur-Ring is a new device. No work has been done on this in Pakistan. We have taken OCR as a black box which has already been developed previously and is needed for its functionality. We are going to enhance its efficiency by further developing it to full fill our requirements. TTS system is also used as a black box and its efficiency will also be increased by further developing the system.

We are making a desktop application but it can be further developed by making it a mobile app so as making it easy to be carried anywhere. Also we are just working on the Nastaleeq font, it can be further developed by including different Urdu fonts in it to increase its usability.

4.8 Design Decision and Tradeoffs

The Design of the device is very handy it is a camera mounted on the ring to be worn on the index finger. The finger is to be placed under the line of text which the user wants to read. The rest of the processing is done on the desktop application. The device will be connected to the laptop through the USB cable. So any laptop or computer on which the application is installed the user will be able to use it by connecting the device through the cable.

CHAPTER: 5

SYSTEM IMPLEMENTATION

5 SYSTEM IMPLEMENTATION

5.1 Ur-Ring: An Overview

The device consists of 4 modules:

1) Camera
2) Application UI
3) Process Data
   i) Divide Frames
   ii) OCR
   iii) TTS
4) Speakers

The video is made in real time by connecting the camera to the system and pointing the finger wearing it at the printed text to be read. The video made in real time through the camera is sent to the process data module for further processing. Here it is checked first if the finger is aligned with the text or not. If yes it is further processed, if not a sound alert is generated to give the user feedback. The video obtained is converted to images and is fed to the OCR to convert it into text. The OCR is to recognize the printed text in the images and convert it into editable text. This text is displayed on the screen of the Application UI module. This text is then converted into speech to be heard by the user.

Feedback is given to the user if the text is not aligned or there is an end of line by sound alert.

5.2 Main Flow

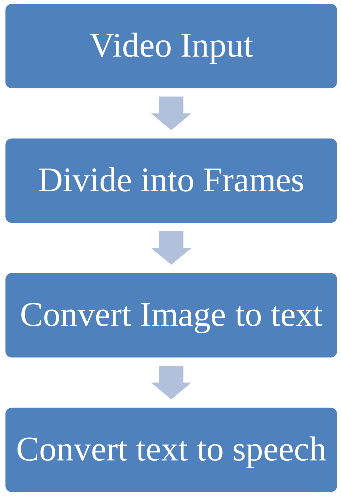


Figure 5-1:Ur-Ring Main Flow

5.3 Pseudo Code

5.3.1 Application UI

```
Begin

Get_video (); //method to get video input from the camera

Send_Video(); //method to send video to Divide_Frame class for further processing

Get_text(); // method to receive text from OCR class

Display_Text(); //method to display text on the screen

End
```

5.3.2 Process Data

Process data has following three components:

5.3.2.1 Divide Frames

```
Begin

Get_video(); //method to get video from the Application UI

Split(); //method to split video into frames

Send(); //method to send frames to OCR

End
```

5.3.2.2 OCR

```
Begin

Get(); //method to get frames from Divide_Frames class

Convert(); //method to convert the frames in form image to text

Send(); //method to send the converted text to UI Application and TTS class

End
```

5.3.2.3 TTS

```
Begin

Get(); //method to get text from OCR class

Convert(); //method to convert text to speech

End
```

CHAPTER: 6

ANALYSIS AND EVALUATION

6 ANALYSIS AND EVALUATION

6.1 BLACK BOX TESTING

6.1.1 Camera Connection

Test Case Name	Camera Connection
Test Case Number	1
Description	Camera should be connected to the system
Testing Technique Used	Black box testing
Preconditions	The desktop or laptop should be on and in the working environment
Input Values	The camera should be connected to the device
Valid Inputs	The Camera is on and is making video for input
Steps	Turn the Camera on Check if the video is made
Expected Output	Video Frames
Actual Output	Video Frames
Status	Pass

Table 6-1: Test Cases for Camera Connection

6.1.2 OCR output as text

Test Case Name	OCR output as text
Test Case Number	2
Description	Recognize and convert the image of text into

	text file
Testing Technique Used	Black Box Testing
Preconditions	Image with focused text
Input Values	Image to be fed to the OCR of the focused text
Valid Inputs	Image with focused text to be fed to the OCR
Steps	Image fed as input to OCR OCR detects the word in the image OCR converts the detected word to text
Expected Output	Image to text conversion
Actual Output	Image to text conversion
Status	Pass

Table 6-2: OCR output as text

6.1.3 GUI

Test Case Name	GUI
Test Case Number	3
Description	Converted image to text displayed on the screen
Testing Technique Used	Black Box Testing
Preconditions	Image to Text conversion from OCR
Input Values	Text input to be displayed on the screen
Valid Inputs	Converted text to be displayed on the screen
Steps	Get the converted image from OCR to text Display it on the screen of the system

Expected Output	Text is displayed
Actual Output	Text is displayed on the screen
Status	Pass

Table 6-3: Test Cases for GUI

6.1.4 TTS Output

Test Case Name	TTS output
Test Case Number	4
Description	Convert the Text to speech
Testing Technique Used	Black Box Testing
Preconditions	Text input
Input Values	Text input from the OCR to the TTS system
Valid Inputs	Text input from the OCR
Steps	Get the text from OCR Take the text as input to the TTS System Convert the text to Sound to be heard by the user
Expected Output	Text to speech conversion
Actual Output	Text to speech conversion
Status	Pass

Table 6-4: Test Cases for TTS output

6.2 White Box Testing

6.2.1 Scene Detection

Test Case Name	Scene detection

Test Case Number	5
Description	Camera should detect the background i.e. a page with text written on it
Testing Technique Used	White box testing
Preconditions	Camera should be on
Input Values	Video Frames
Valid Inputs	Video Frames detection
Steps	Camera should be making video System should compare and detect the scenes
Expected Output	Scene detected
Actual Output	Scene detected
Status	Pass

Table 6-5: Test Cases for Scene detection

6.2.2 Finger Detection

Test Case Name	Finger detection
Test Case Number	6
Description	Finger extraction from the scene detected
Testing Technique Used	White Box Testing
Preconditions	Scene is detected
Input Values	Finger should be clearly visible through the camera
Valid Inputs	Finger is visible through the camera on the paper
Steps	Camera should be making video

	System should detect the finger through skin color detection
Expected Output	Finger detection
Actual Output	Finger detection
Status	Pass

Table 6-6: Test Cases for Finger Detection

6.2.3 Tip Detection

Test Case Name	Tip detection
Test Case Number	7
Description	Tip of the finger is detected
Testing Technique Used	White Box Testing
Preconditions	Finger is detected
Input Values	Binary image of the detected finger
Valid Inputs	Binary image of the detected finger
Steps	Finger is detected Search for first white pixel in binary image of finger Get the coordinates of the pixel Insert marker at the tip of finger
Expected Output	Insert marker at the detected tip
Actual Output	Marker inserted at the detected tip
Status	Pass

Table 6-7: Test Cases for Tip detection

6.2.4 Line Detection

Test Case Name	Line detection
Test Case Number	8
Description	Detect the line below which the finger is placed
Testing Technique Used	White Box Testing
Preconditions	Tip detection of the finger
Input Values	Row coordinates according to the tip of the finger as well as the number of lines visible in the scene.
Valid Inputs	Get row coordinates and number of lines visible in the scene.
Steps	Finger is detected with a marker on the line to be read The focused line is fed further for processing
Expected Output	Line is detected for further processing
Actual Output	Line detected
Status	Pass

Table 6-8: Test Cases for Line Detection

6.2.5 Feedback at end of line

Test Case Name	Feedback at end of line
Test Case Number	9
Description	Sound alert generation at the end of line
Testing Technique Used	White Box Testing

Preconditions	Line detection
Input Values	Focused line in the video frame
Valid Inputs	Video frame with the detected line
Steps	Detect the line to be read Analyze end of line Generate feedback to user when the line ends
Expected Output	Sound alert for the user to know when the line ends
Actual Output	Feedback generated
Status	Pass

Table 6-9: Test Cases for feedback at end of line

6.2.6 Image Extraction

Test Case Name	Image extraction
Test Case Number	10
Description	Extract image after the word detection to be fed to OCR for further processing
Testing Technique Used	White Box Testing
Preconditions	Video Frames with focused text
Input Values	Focused word in the video frames
Valid Inputs	Focused words in the video frames
Steps	Get the frame with the text to be read Draw a window on the text to be read Save the image to be sent to OCR for further processing
Expected Output	Image with specified text to be read

Actual Output	Image with specified text to be read
Status	Pass

Table 6-10: Test Cases for Image Extraction

6.3 Integration Testing:

This application consists of four modules which are Application UI, Camera, Process Data and Speakers. These modules are integrated by using top down integration technique. Top down integration testing is an incremental integration testing technique which begins by testing the top level module and progressively adds in lower level module one by one. Top Down integration can be performed and tested in breadth first or depth first manner.

First of all, we enable the Camera by plugging it through the data cable into the system. After this we test if the camera is making video and if the video is of printed text. Then we detect the scene and finger in process data module and convert the video into images. These images are then fed to OCR and checked. The OCR converts it into text and display it on the screen in Application UI module. Furthermore, integration testing is done to check if the text is converted into speech accurately. We carried out black box tests after implementing every module, and the results were successful after every phase.

6.4 System Testing:

The System Test for Ur-Ring system will verify that the requirements to system have been implemented as described in the System Requirements and Design Specifications documents.

6.4.1 General Functionality Testing:

Test scenarios will be established based on the systems requirements and design Specifications documents.

6.4.2 Performance Testing:

Performance testing will verify that the text is converted into speech with accuracy.

6.5 Acceptance Testing:

This is the testing of the product that was performed by the developing team to check if the requirements of the system has been met.

CHAPTER: 7

FUTURE WORK

7 FUTURE WORK

Future work that can be done on this project is to enhance the accuracy of the OCR and to minimize the delays in processing. Also the major work that can be done is to add more words of Urdu Language so the system is efficient and workable on any text. As well as different fonts can be introduced as we are only working on the Nastaleeq font. Since the system we are developing is desktop based application, in future work it can be made wireless so as to make convenient for the user to take it anywhere for usage. This system can also be developed for other languages so as to not restrict it for the usage by people of some specific area. As well as android and iOS applications can be made for this so as to make it more accessible to the user.

CHAPTER: 8

CONCLUSION

8 CONCLUSION

8.1 Overview

The motivation of the project is to provide a feasible solution to the visually impaired and illiterate people to help them in their daily lives to read printed text.

This device is simple to use and does not have many hardware components.

Since there are 1.5 million visually impaired people and 45% illiterate people in Pakistan this device can help them read printed text in Urdu language as it converts printed text to speech which can be heard by the user.

8.2 Objectives Achieved

The Ur-Ring application helped to achieve the objectives of learning image processing and neural network training. It also helped in devising a solution by gathering requirements regarding the problem and solving it in an efficient manner.

BIBLIOGRAPHY

9 BIBLIOGRAPHY

1. http://fluid.media.mit.edu/projects/fingerreader
2. http://news.mit.edu/2015/finger-mounted-reading-device-blind-0310
3. http://www.cle.org.pk/
4. http://www.mathworks.com/matlabcentral/fileexchange/32710-ocr-urdu-compound-optical-character-recognition-code-and-thesis
5. http://www.mathworks.com/matlabcentral/newsreader/view_thread/276951?requestedDomain=www.mathworks.com
6. http://www.mathworks.com/examples/simulink-computer-vision/mw/vision_product-vipscenechange-scene-change-detection
7. http://www.mathworks.com/help/images/examples/contrast-enhancement-techniques.html
8. www.imlijh.com/projects/arabichandwritingrecognition/index.xml
9. http://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_natural_language_processi
10. https://www.researchgate.net/publication/4079707_Natural_language_processing_for_Urdu_TTS_system
11. http://www.cle.org.pk/Publication/papers/2005/Phonological_Processing-4-Urdu_Text-to-Speech_System.pdf
12. http://www.cle.org.pk/Publication/theses/2010/OCRMUAZ.pdf
13. http://ieeexplore.ieee.org/xpls/icp.jsp?arnumber=5625694
14. http://iovs.arvojournals.org/article.aspx?articleid=2125031

APPENDIX A

USER MANUAL

<u>Ur-Ring</u>

1) Overview

Ur-Ring (Urdu Finger Reader Ring) is a device which reads printed text written in Urdu language. It consists of a camera with a data cable which is attached to the system in which the processing of the video is to take place. The system converts the processed image into sound to be heard by the user.

2) Mounting the Camera

The camera mounted on the ring should be worn on the index finger and should be attached to the system with the data cable and turned on. The finger should be pointing at the printed text on paper.

3) System Processing

The system should be on with the camera attached to it so that the processing of the images can take place and the user can hear the written text through the system speakers.

The system speakers should be turned on and medium volume should be maintained so that the voice produced is clearly heard by the user.

4) Safety and Precautions

- The device is not to be used by children under 6 years of age.
- The device is not water proof.
- The camera is fragile and not to be mishandled.
- It can only be used to read Urdu printed text.
- The device will only operate with the software provided with it.