PUBLIC OCR SIGNAGE RECOGNITION WITH SKEW & SLANT CORRECTION FOR VISUALLY IMPAIRED PEOPLE

by

INTAN FARIZA BINTI HAIRUMAN

Dissertation submitted in partial fulfilment of
the requirements for the
BACHELOR OF TECHNOLOGY (Hons)
(BUSINESS INFORMATION SYSTEM)

MAY 2011

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

PUBLIC OCR SIGNAGE RECOGNITION WITH SKEW & SLANT CORRECTION FOR VISUALLY IMPAIRED PEOPLE

by

INTAN FARIZA BINTI HAIRUMAN

A project dissertation submitted to the
Business Information System Programme
Universiti Teknologi PETRONAS
in partial fulfillment of the requirement for the
BACHELOR OF TECHNOLOGY (Hons)
(BUSINESS INFORMATION SYSTEM)

Approved by,

____

(Mrs. Foong Oi Mean)

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK
May 2011

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the reference and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

INTAN FARIZA BINTI HAIRUMAN

ABSTRACT

This paper presents an OCR hybrid recognition model for the Visually Impaired People (VIP). The VIP often encounters problems navigating around independently because they are blind or have poor vision. They are always being discriminated due to their limitation which can lead to depression to the VIP. Thus, they require an efficient technological assistance to help them in their daily activity. The objective of this paper is to propose a hybrid model for Optical Character Recognition (OCR) to detect and correct skewed and slanted character of public signage. The proposed hybrid model should be able to integrate with speech synthesizer for VIP signage recognition. The proposed hybrid model will capture an image of a public signage to be converted into machine readable text in a text file. The text will then be read by a speech synthesizer and translated to voice as the output. In the paper, hybrid model which consist of Canny Method, Hough Transformation and Shearing Transformation are used to detect and correct skewed and slanted images. An experiment was conducted to test the hybrid model performance on 5 blind folded subjects. The OCR hybrid recognition model has successfully achieved a Recognition Rate (RR) of 82.7%. This concept of public signage recognition is being proven by the proposed hybrid model which integrates OCR and speech synthesizer.

ACKNOWLEDGEMENT

In the name of Allah s.w.t., the Most Gracious, the Most Merciful.

First and foremost, the author would like to express greatest gratitude to Allah the Almighty. A thousand thanks goes to Mrs. Foong Oi Mean. The author would like to give her a sincere appreciation for her role as a project supervisor, for her kindness and guidance throughout completing this final year project, and also for her endless encouragement and patience.

Next, the author would also like to give outmost appreciation to her friends and family for their continual support. Thank you for willing to share all kinds of resources, knowledge and experience. The author appreciates the time that they had scarified for this project. It would be impossible for the author to work on this without any help from all of them.

Last but not least, the author's big gratitude goes to Universiti Teknologi PETRONAS (UTP) for providing this conventional place with superb facilities for learning.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

OCR	Optical Character Recognition
VIP	Visually Impaired People
RR	Recognition Rate
MATLAB	Matrix Laboratory
RAD	Rapid Application Development
SAPI	Speech Application Program Interface

CHAPTER 1
INTRODUCTION

1.1 Background of Study

The project was inspired by the special group of people who have difficulties with their vision where it investigates a method to aid these VIP to read public signage to improve their daily lives or work. Often, these people known as the Visually Impaired People (VIP); has the inability to see and have tremendous reduction in vision. However, people often misunderstood the VIP for having a total blindness. The VIP is actually inclusive of the blind and people with low vision. Thus, the project will focus on the category 1 and 2 of visually impaired people which will be discussed further in this paper.

"The blindness refers to a condition where a person suffers from any of the following conditions, namely total absence of light; or visual acuity not exceeding 6/60 or 20/200 (snellen) in the better eye even with correction lens; or limitation of the field of vision subtending an angle of 20 degree or worse." (Dijik, 2002). However, "a person with low vision is one who has impairment of visual functioning even after treatment, and/or standard refractive correction, and has a visual acuity of less than 6/18 to light perception or a visual field of less than 10 degrees from the point of fixation, but who uses, or is potentially able to use, vision for the planning and/or execution of a task." (WHO, 1992).

World Health Organization (WHO) provides the standard for each of VIP's category based on the visual acuity and is extensively used in medical reports and publications. WHO defines blindness as visual acuity of less than 3/60 in the better eye with the best possible correction while a person with more than or equal with 3/60 visual acuity and less than 6/60 is defined as severe visual impairment. "Visual acuity refers to the ability of the eye to see details. The visual acuity for distance is measured as the maximum distance at which person can see a certain object, divided by the maximum distance at which a person with normal eyesight can see the same object." (Dijik, 2002). The methods in measuring acuity are by using the E-chart and Finger Count. E-chart is extensively being used in optometry shop to measure customer's visual acuity before ordering their lens. It involves letters from the biggest to the smallest and customer will have to tell each letter as their lens is corrected and measuring the visual acuity. Second simpler method is by Finger Count where at a distance of 6 meters, a person will have to guess the number of fingers of other person. The person has a perfect vision of 6/6 if he can see and count the fingers clearly and vice versa. Nevertheless, below are the categories of visual impairment according to the visual acuity measurement which compiles to WHO's standard.

Category	**Corrected VA – Better Eye**	**WHO Standard/Definition**
0	6/6-6/18	Normal
1	<6/18-6/60	Visual Impairment
2	<6/60-3/60	Severe Visual Impairment
3	<3/60-1/60	Blind
4	<1/60-PL	Blind
5	NPL	Blind

Table 1: Categories of visual impairment

Some of the VIP's vision is still can be used but with very low in performance, therefore; there is a rising need in technological assistance in aiding the VIP to lead a normal life. This is due to the modernized infrastructures which also lead to the lacking of handicapped-friendly environment. Despite of all the assistances they received, the VIP might have difficulties in reading public signage in the public environment and sometimes they are out without any company to rely on. The help of technology advancement for the VIP will tremendously help them to be more independent and avoid error in direction or while purchasing goods or services.

1.2 Problem Statement

Figure 1: Example of public signage

The VIP usually have problem navigating around independently because they are blind or have low vision. VIP often accompanied by relatives or guardian. Without their guardian, they could not move freely and will face difficulty using public facilities such as toilet. For example the public signage as depicted in figure 1 is often seen, and the VIP might sometimes mistaken the toilet either for men and women, hence the need of technological assistance to help them in differentiating the signage. Other than that, the VIP would appreciate assistance during selecting food orders from menu in a restaurant. This could avoid them from ordering the wrong food or to keep track the price for each food available.

Other than that, often VIPs are being discriminated due to their inability to see. This could lead to depression and sometimes will result in high danger for their health. Thus, having a technological assistance can also help them to regain confidence especially going out in the public.

The OCR system should be able to recognize even skewed or slanted characters because the VIP will snap pictures randomly; and the possibility that the characters taken in an image to be skewed or slanted is high.

Figure 2: Example of slanted characters

Figure 2 portrays the example of slanted characters of "CAUTION" or "WET FLOOR". Usually the VIP will randomly capture the picture of public signage and the picture might be slightly or excessively slanted, and a system that can process slanted characters of more than 25^{o} will be more efficient and useful.

Lastly, available aids for the VIP are already marketed but often the price is relatively expensive where not all VIP could afford one and some of them are even difficult to use. Examples of VIP technological assistance are Global Positioning System (GPS) for the blind, wireless walking stick for the blind and mobile phones character recognition for the VIP.

1.3 Objectives

1. To propose a hybrid model for Optical Character Recognition (OCR) skew & slant correction on public signage.
2. To integrate speech synthesizer with the hybrid model for VIP signage recognition.

1.4 Scope of Study

The main scope of the project is to focus on how OCR can help VIP in their daily activities. The project works in a situation where a VIP will capture the picture of random public signage and the picture will later be processed using OCR system that will recognize the character contained in the picture and convert it into a machine readable text. The text will be processed again using the speech synthesizer to produce a voice output.

Usually public signages have a combination of images and characters. However, the project only focuses on characters contained in an image. Moreover, characters in the images taken by the VIP usually are skewed or slanted; where an efficient algorithm should be implemented in order to overcome the problem.

CHAPTER 2

LITERATURE REVIEW

2.1 Optical Character Recognition (OCR) Research Review

Extensive research has been done on OCR in offering different algorithm which suits specific needs of the project. "OCR is the process is which a paper document is optically scanned and then converted into computer process able electronic format by recognizing and associating symbolic identity with every individual character in the document." (Ghosh et al., 2010). OCR includes several processes which will transform characters in images into text:

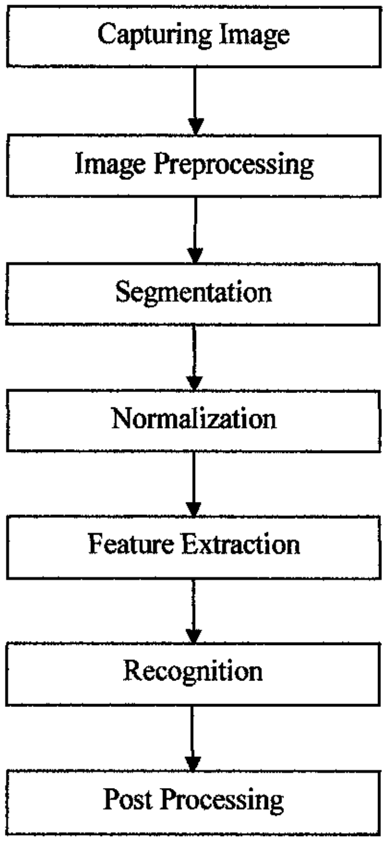


Figure 3: State diagram of the OCR process

OCR process as depicted above display the different steps in OCR processing which include the capturing the image, image preprocessing, segmentation, normalization, feature extraction, recognition and then post processing. The process begins in capturing image, where images which contain characters from public signage are transformed into scanned or digital format. The image taken could result in slight to moderate skewed and slanted image which requires image preprocessing to correct the problem.

In image preprocessing of OCR, several basic steps are always implemented to ensure the image is ready to be successfully recognized. The next step will be the segmentation process where the character splitting process begins. The splitting process involves determining rectangular region of each of the object character to identify the characters or words.

Segmented characters which have already been enveloped into rectangular regions would have to go to normalization process, where algorithms for resizing the characters whether to enlarge or reduce the size to fit into processes-able input.

Next step is the feature extraction process. "The purpose of feature extraction is the measurement of those attributes of patterns that are most pertinent to a given classification task." (Cheriet et al., 2007). This means that, feature extraction describe for methods for simplifying complex resources such as computational power; while processing the large set of data accurately.

Then, the input would go to the character recognition process. In that particular step, the character in processed image will be analyzed and recognized; the character will be match with possible results in the character database to confirm the character and lastly to return the result.

Lastly in post processing stage, the recognized character will be used in various purposes. Some of them are being used commercially for digital library, and other are being used to convert the character into speech using speech synthesizer.

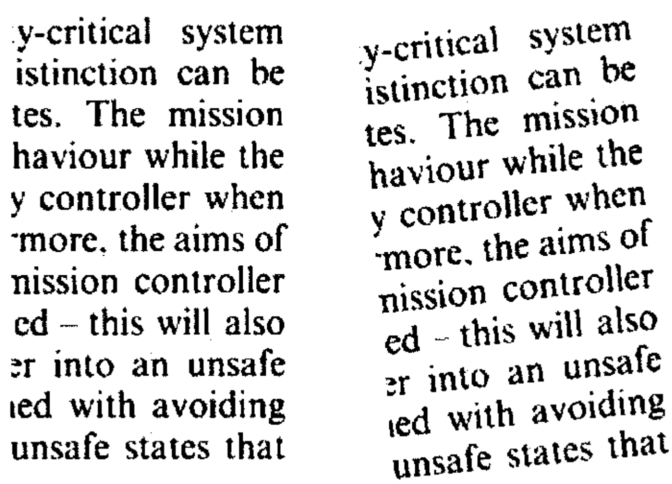
2.2 Image Preprocessing

"As the first important step, image and data preprocessing serve the purpose of extracting regions of interest, enhancing and cleaning up the images, so that they can be directly and efficiently processed by the feature extraction component." (Cheriet et al., 2007). There are several basic steps included in image preprocessing and the number of steps being used should tailor to the OCR purpose. Other than that steps included also should be based on several factors such as image quality and resolution, layout of the character, the amount of noise, skew and slant, type of script being used and also type of character either printed or handwritten.

Some of the steps in image preprocessing include image acquisition, noise removal, skew and slant correction. The image acquisition acquire captured image in the form of grayscale and binary format. Other than that, in order to remove noise in an image; a smoothing operation is performed. According to (Cheriet et al., 2007), "smoothing operations are used to reduce the noise or to straighten the edges of the characters, for example, to fill the small gaps or to remove the small bumps in the edges (contours) of the characters."

2.2.1 Skew Detection and Correction

Another important step in image preprocessing is the skew detection and correction or also known as tilt detection and correction. According to (Liu et al., 2003), "Tilt of a word is defined as the general ascending or descending trend of the writing with respect to the horizontal line or the x-axis." The reason that skew or tilt should be eliminated is because it could largely reduces the accuracy of recognition rate of the OCR system. Therefore, in order to achieve a desired result; skew detection and correction should be applied in the image preprocessing.



(a) (b)

Figure 4: Correctly aligned image (a) and skewed characters (b)

There are many methods which have been identified by various researches. The proposed methods in the literature are rotation, nearest-neighbor clustering, Hough Transform (Singh et al., 2008), cross-correlation and morphological transform.

There is also proposed method in skew detection called connected component analysis proposed by (Saragiotis et al., 2008). The method works on grayscale image, where Gaussian filter is used to remove all the noise contained in the image. Then the analysis has to find connected element from image lines to construct a bounding rectangle. Another straightforward method is the projection profiles analysis, where an image is projected at several angles to compute the variance in the number of black pixels per projected lines.

After the skew angle is estimated, a rotation algorithm needs to be implemented to correct the skew. Using the new acquired coordinates, the following formula (Cheriet et al., 2007) is used for skew correction:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

2.2.2 Slant Detection and Correction

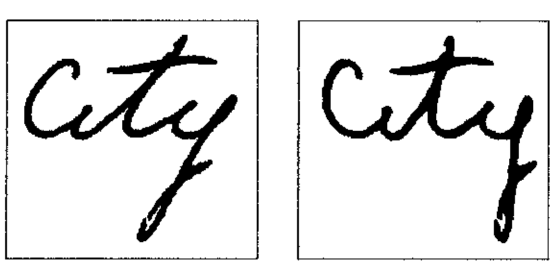
"The character inclination that is normally found in cursive writing is called slant." (Cheriet et al., 2007). The character inclination found in images also will result in a low accuracy rate of the OCR system. Thus, finding a method for slant correction is highly anticipated.

There are also several useful methods to conduct the slant correction. The method mentioned in (Slavik et al., 2001), is shear transformation to estimate and correct slant. Other than that, slant can also be estimated first by calculating the coordinate for four lines surrounding the characters using the following formulas (Cheriet et al., 2007):

$$y = x + \beta_1,$$
$$y = x + \beta_2,$$
$$y = x + \beta_3,$$
$$y = x + \beta_4,$$

The slant angle is then being calculated as $\theta = arctan(B/A)$, where $B = (\beta_4 + \beta_1 - \beta_3 - \beta_2)/2$ and $A = Height\ of\ the\ character$. After that, the horizontal shear transformation is being applied on the image to correct the slant by shifting its pixels to the left or right according to the estimated θ. The shear transformation algorithm is computed as:

$$x' = x - y \cdot \tan(\theta),$$
$$y' = y$$



(a) (b)

Figure 5: Slanted character (a) and after slant correction (b)

2.3 Matlab

Matlab stands for Matrix Laboratory, is highly interactive software developed by Math Works mainly used for scientific and engineering computation. Matlab is a powerful tool because it is able to perform complex mathematical equation efficiently and faster than earlier developed programming language such as C and C++. For image processing, Matlab will be a suitable platform to build the OCR system because the image processing toolbox is readily available inside the software.

2.4 Related Work on OCR

Many attempts for using OCR to help the VIP have been published with relevant literature. One efficient way to use OCR to help the VIP is by converting the text from physical document into a machine readable text, and then converts the recognized text into speech using speech synthesizer. Prior work from (Dumitras et al., 2006) who developed a phone-based text-recognition for the visually-impaired has applied this framework by using a mobile phone. The phone camera will capture an image with characters and the built-in OCR system will translate the character into text and the output will be voice. Similar project has also been done by (Gaudissart et al., 2004). Nowadays, iPhone has launched its iPhone Apps namely ZoomReader to help the VIP to read with the same framework. These projects done were to read expiry date, restaurant menu, printed instructions and packaged goods.

Other than that, (Kurzweil et al., 2006) proposed a reading machine for the VIP where it is being used to read printed documents such as newspapers, books or journals. The reading machine requires the user to manually scan the book face down on the scanner. The project does a tilt or skew correction to make sure scanned text is horizontally aligned before it goes through the OCR system. OCR is also useful in car plate recognition which is discussed in (Qadri et al., 2009).

To achieve the different desired results, different popular techniques such as Neural Network, Hidden Markov Models and Support Vector Machines have also been implemented. OCR implementation in Neural Network is discussed in (Deng et al., 2009) and (Ganapathy et al., 2007), while discussion for Support Vector Machines is being used in (Ramanathan et al., 2010). There is also discussion regarding improved Hidden Markov Models algorithm for Arabic

OCR which is used in (Prasad et al., 2009). All of these projects implement different methods in achieving desired result. This shows the proof that various methods can be used to suit the different needs of each project.

2.5 OCR for Multiscript Recognition

(Ghosh et al., 2010) emphasize on methodologies which can accommodate recognition of different scripts in the world. The OCR architecture has a bank of multiple OCR language technique to correspond to all different scripts expected, which can identify different scripts and transform them into text including Logographic, Alphabetic, Bharmic, Abjads, Featural and Syllabic scripts.

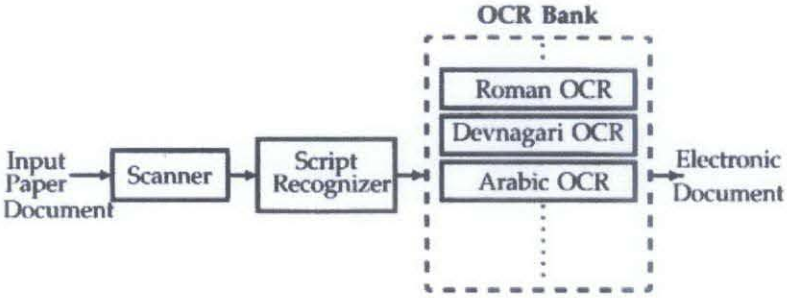


Figure 6: Document processing in a multi-script environment (Ghosh et al., 2010)

2.5.1 Script Identification at Paragraph and Text Block Level

Other researchers have been using various techniques in multiscript recognition where (Patil et al., 2002) has developed the system architecture using neural network technique to identify Latin, Devnagari and Kannada scripts.

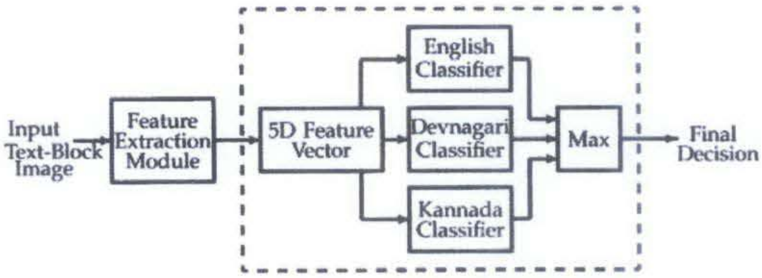


Figure 7: Neural network-based architecture for script identification (Patil et al., 2002)

"In the feature extraction stage, a feature vector corresponding to pixel distributions along specified directions is obtained via morphological operations. The modular neural network structure consists of three independently trained feed-forward neural networks, one for each of the three scripts under consideration." (Ghosh et al., 2010). Recognition rate was tested and is said to reach 100% accuracy for recognizing English and Kannada scripts and 97% for Devnagari script. Another research is using the neural network technique in (Chi et al., 2003) to recognize Latin and Han scripts, but without feature extraction module. Moreover, (Kanoun et al., 2002) and (Zhou et al., 2006) have also contributed their work to recognize Arabic, Latin, English and Bengali scripts.

2.5.2 Text-line-wise Script Identification

A research in (Pal et al., 2002) automates OCR system for Latin, Urdu, Devnagari and Bengali scripts as shown below:

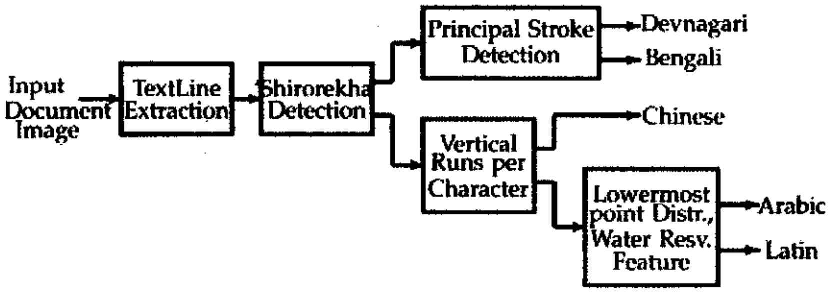


Figure 8: Script line separation from multiscript documents in India (Pal et al., 2002)

In Shirorekha Detection phase, Devnagari and Bengali script text-lines are separated from Latin, Chinese and Arabic script text-lines. "Next, Bengali script lines are distinguished from Devnagari by observing the presence of certain script-specific principal strokes. Similarly, Chinese text-lines are identified by checking the existence of characters with four or more vertical runs. Finally, Latin (English) text-lines are separated from Arabic using statistical as well as water reservoir-based features." (Ghosh et al., 2010). Deeper research has been made in (Pal et al., 2003) which able to classify 12 different Indian scripts.

CHAPTER 3
METHODOLOGY

3.1 Rapid Application Development (RAD)

The project adopts Prototyping in one of the categories in RAD which enables developers to develop the system quickly in parts and hand in to users for better understanding to make further improvements. Prototyping category in RAD involves performing the analysis, design and implementation concurrently and repeatedly until the completed final implementation.

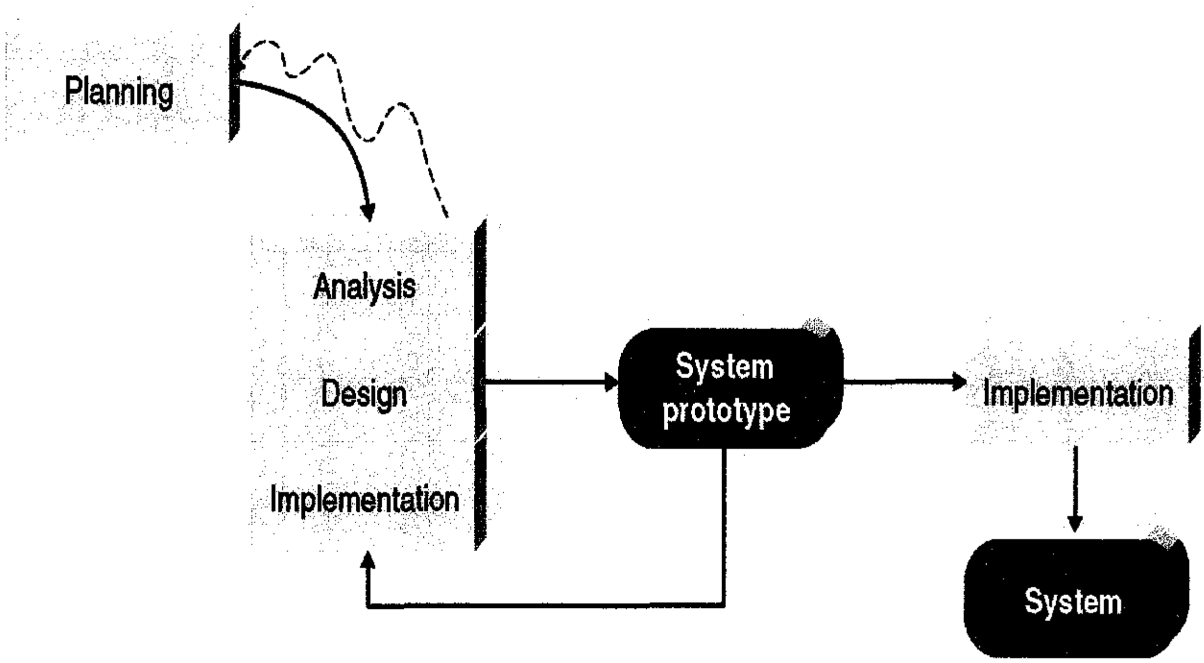


Figure 9: RAD - Prototype model

The advantage of using Prototyping rather than other methods is that, Prototyping is able to provide end users with the real system at an early stage of system development, and enable them to test the system continuously. The end users could comment on the prototype for any shortcoming and let developers make amendments quickly and efficiently.

3.2 Flow Chart of OCR Hybrid Recognition Model

The process and activities involved in the project's execution is done in stages and is illustrated as below:

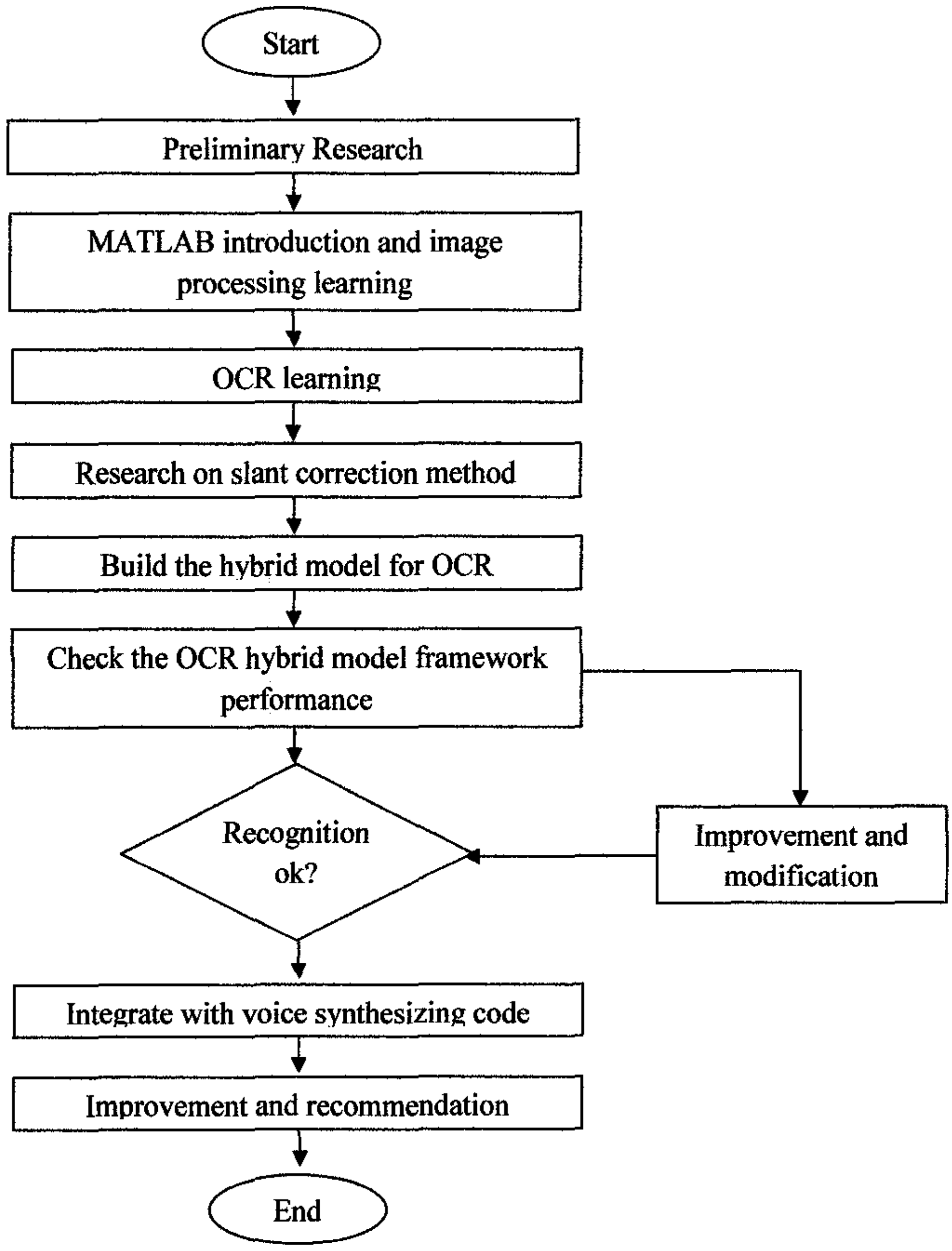


Figure 10: Flow Chart of OCR Hybrid Recognition Model

3.3 Proposed System Architecture of OCR Hybrid Recognition Model

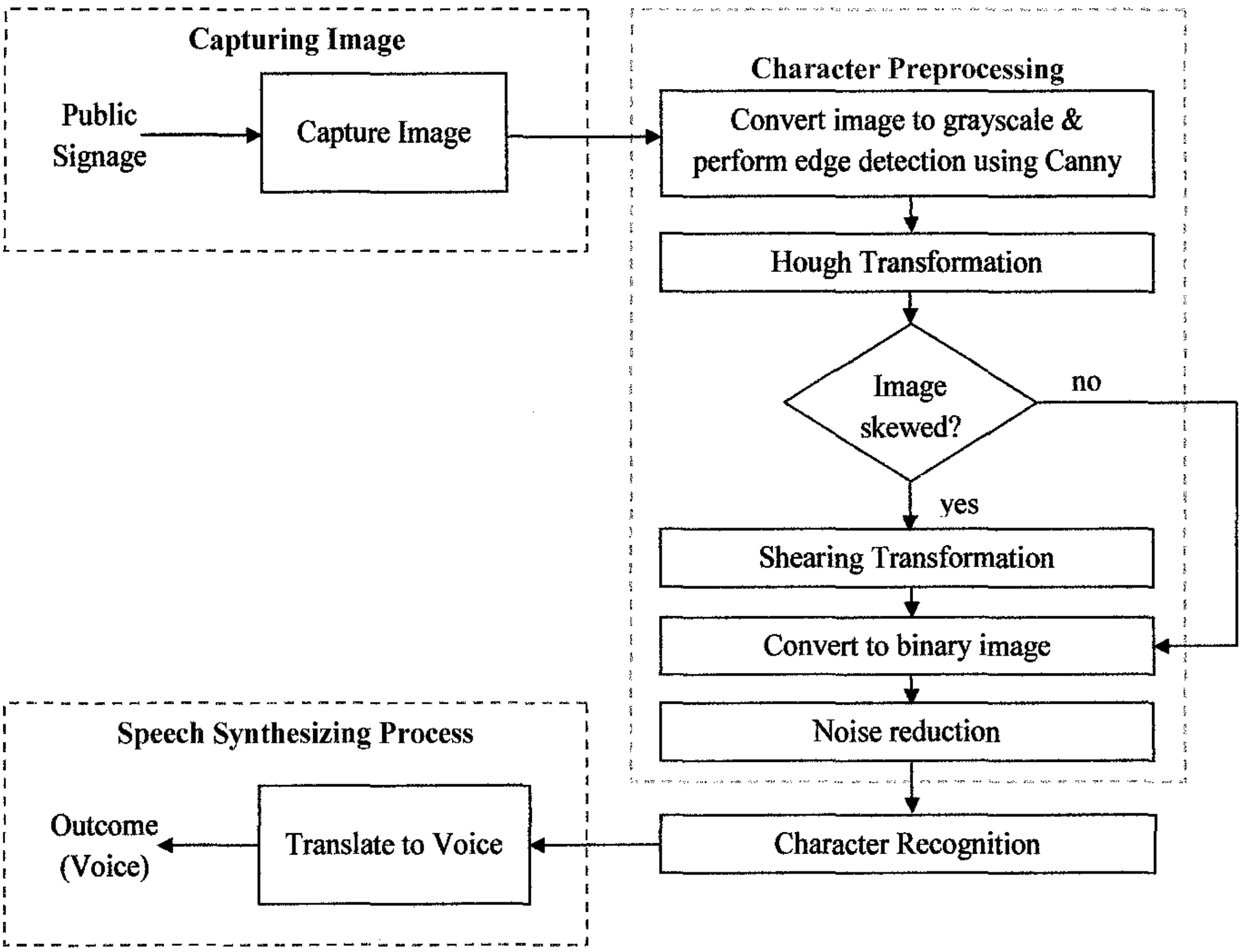


Figure 11: Proposed system architecture of OCR hybrid recognition model

The system architecture as shown in Figure 11 above demonstrates the overall process of OCR hybrid recognition model from capturing image, to character preprocessing and recognition and finally the speech synthesizing process.

3.3.1 Capturing Image

The VIP will capture image of public signage with characters and the image will then be stored in .jpeg format in specific folder. Usually the captured or scanned image contains noise as well as skewed and slanted characters resulting from non-aligned image capturing technique especially by the VIP. Therefore, it is important to do preprocessing first before character recognition to achieve a more accurate result.

3.3.2 Character Preprocessing & Recognition

The most significant part for this paper is in the preprocessing part where the steps included in preprocessing are based on several factors. The influential factor including moderate image quality due to inconsistent image capturing using different types of capturing methods such as hand phone camera, compact camera or any portable camera. Another factor is the layout of the character where the public signage in Malaysia is usually in horizontal layout. Moreover, image in public signage contain high amount of noise and possibility of slight to moderate skew and slanted character. The type of script being used is alphabetic script and mostly the public signage is using printed character. Thus, several image preprocessing techniques are applied to tailor to these factors.

The hybrid model is implemented in 5 processes. The processes include convert to grayscale image, Canny Method edge detection, Hough Transformation, Shearing Transformation, convert to binary image and noise reduction.

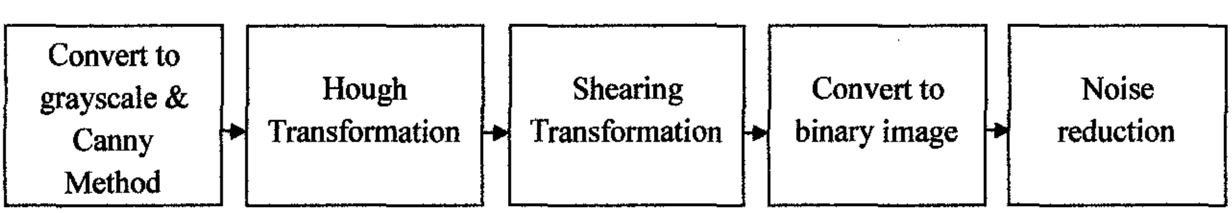


Figure 12: OCR hybrid recognition model for character preprocessing

First, the image taken will need to be converted into grayscale image and Canny Method is used to detect edge of characters inside of the image. Canny Method has been extensively used in image preprocessing because it gives a smoother finish than any other methods. It uses multistage algorithm and the outcome will be image showing position of tracked intensity. First the image is convolved using a Gaussian filter based on the following formula:

$$G_\sigma = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{m^2 + n^2}{2\sigma^2}\right]$$

Next, the intensity of the image is being computed as below to detect four edges of the characters which are horizontal, vertical and diagonal edges using below formula:

$$G = \sqrt{{G_x}^2 - {G_y}^2}$$

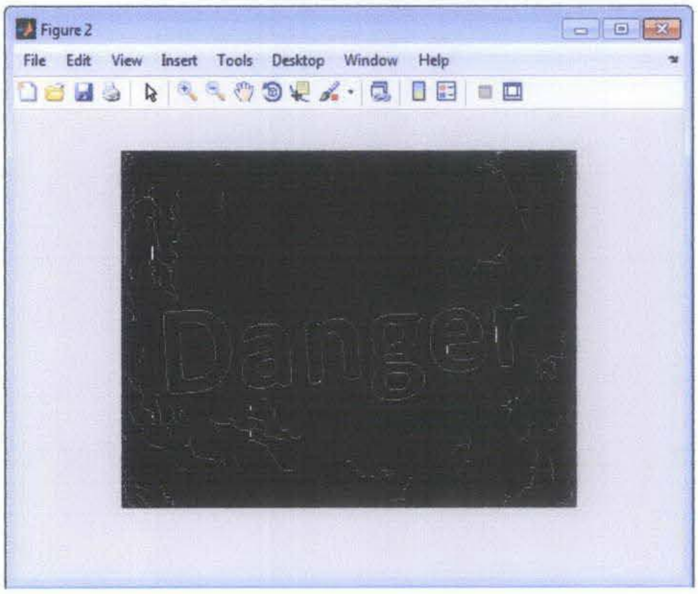


Figure 13: Screenshot of edge detection

Hough Transformation is being used to detect arbitrary shapes in the image by finding straight lines of characters in image. The parameters of curve are found based on given edge point which was derived from Canny Method. To compute the straight lines in Hough Transformation, parameters identified are r and θ where r is the distance from the origin and character's line, and θ being the angle from origin to the closest point. The formula for Hough Transformation is shown below:

$$y = \left(-\frac{cos\theta}{sin\theta}\right)x + \left(\frac{r}{sin\theta}\right)$$

Next, the characters in the image will be determined where it is slanted or not using these code. If the characters are slanted, Shearing Transformation with Affine Transformation will be done to correct and rotate skewed or slanted characters into horizontal aligned characters. Formula used in rotation is shown below:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = s \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}$$

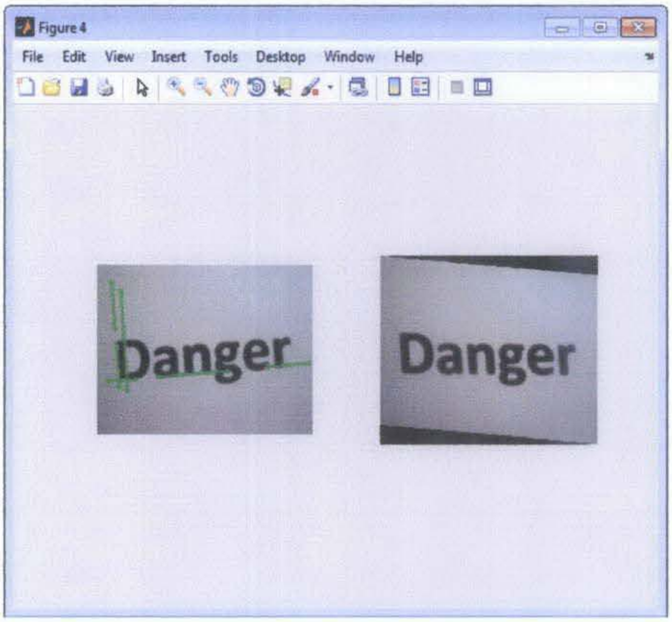


Figure 14: Corrected character image after Shearing Transformation

The grayscale image will be converted into binary image. After that, the noise in the image will removed where all objects containing fewer than 50 pixels will be eliminated.

After character preprocessing is done, character recognition is performed by matching the recognized character with the template database. The template database contains 26 upper case letter images in the dataset, 26 lower case letter images in the dataset and 10 numeric images in the dataset. The matching process will be determined based on the characters similarity from recognized character and template database. The output will be stored in text.txt file.

3.3.3 Speech Synthesizing Process

After the character preprocessing and recognition, text in the text.txt file will be converted into speech through speech synthesizing process. A command below is being used to read character string from text.txt file:

```
A = textread('text.txt', '%c');
```

The parameter identified is A where a text-to-speech function processes input string of A and convert it into speech. Thereafter a command “tts(A)” is used to convert the text to speech. Final output will be voice generated through a speaker.

3.4 Tools Required

This project requires an integration of hardware and software to develop the system. Table below shows tools or components required in the project:

No	Hardware Required	Software Required
1	Minimum Windows XP Operating System with minimum 2GB RAM	Matlab 7.10.0 (R2010a)
2	Camera	SAPI
3	Speaker	Speech SDK 5.1/5.3

Table 2: Tools required

CHAPTER 4

RESULT AND DISCUSSION

4.1 Experimental Setup

The OCR hybrid recognition model is further tested to calculate its Recognition Rate (RR). There are 15 different samples of images and some of them are skewed or slanted. The samples are to be tested by 5 respondents, where all of them were blind folded to act as a VIP. The respondents comprise of 4 females and 1 male age ranged from 20 to 25 years old. All of the images are stored in .jpg file format.

Samples	Number of Respondents									
	1		2		3		4		5	
	T	F	T	F	T	F	T	F	T	F
	√			√	√			√	√	
	√		√		√		√		√	
	√		√		√		√		√	
		√		√		√		√		√
	√		√		√		√		√	

CAUTION WET FLOOR	√		√		√		√			√
BUS	√		√		√		√		√	
OPEN	√		√		√		√		√	
Danger	√		√		√		√		√	
MEN	√		√		√		√		√	
TAXI	√			√	√		√		√	
ROAD	√		√		√		√		√	
CLOSED	√		√		√		√			√
STOP	√		√		√		√		√	
EXIT	√			√	√			√		√

Table 3: Evaluation on the public OCR signage recognition

Respondents	Number of Samples	Number of Recognized Samples
1	15	14
2	15	11
3	15	14
4	15	12
5	15	11
Total	**75**	**62**

Table 4: Evaluation from the respondents

4.2 Experimental Result

From the result in table 4 above, 15 samples of images are tested. To find the Recognition Rate (RR), the following formula is being used:

$$RR = \frac{Classified\ Character}{Total\ Number\ of\ Character} \times 100\%$$

$$RR = \frac{62}{75} \times 100\%$$

$$RR = 82.7\%$$

There are a total of 62 out of 75 samples which were recognized. The calculation is being shown as above where 62 is divided by 75 samples then to be multiply with 100% to get the percentage of accuracy rate. Therefore; the current public OCR signage recognition has 82.7% Recognition Rate (RR).

4.3 Recognition Error Analysis

Based from observation throughout the experiment process, error is always shown for characters which have similar shape where the OCR will mistakenly generate O or 0 instead of D. Moreover, OCR is also having difficulty to differentiate l character from I or i, where it is actually a small letter of L. Other than that, OCR will not be able to recognize character that is cut or incomplete as shown in figure below:

Figure 15: Incomplete character

The R character is cut from the picture which makes is unrecognized by OCR as 9. OCR also could not recognize l word that is assumed as i by OCR and D character that is recognized as 0.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

The proposed hybrid model is integrated into three parts which are image capturing, character preprocessing & recognition and speech synthesizing process. The character preprocessing includes several methods to correct skewed and slanted characters namely Canny Method for edge detection, Hough Transformation for finding straight lines of character in images and Shearing Transformation for skew and slant correction.

The hybrid model shows the proof for OCR hybrid recognition model where the hybrid model has successfully achieved a RR of 82.7%. Although the character is skewed or slanted, the proposed project can detect and correct the skewed and slanted characters in an image. If the character is successfully recognized, the model will produce a voice output corresponding to the input of recognized character.

5.2 Recommendation for Future Enhancement

For future enhancement, there are several recommendations proposed such as a compact gadget, the OCR algorithm to recognize different scripts and the speech synthesizer could translate English to different language. The OCR should able to recognize different types of scripts because all of the scripts commonly used in Malaysia are logographic, alphabetic and abugidas (brahmic) scripts. Other than that, the speech synthesizer only pronounce English word; the hybrid model should be improve in translating English word into commonly used language in Malaysia such as Cantonese, Mandarin, Malay and Tamil.

REFERENCES

[1] K. V. Dijik, "Definition: Visual Impairment Handbook ", New Vision Consultant, 2002.

[2] World Health Organization (1992): Management of Low Vision in Children P. 7

[3] Ghosh, D.; Dube, T.; Shivaprasad, A.P.; , "Script Recognition—A Review, "*Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.32, no.12, pp.2142-2161, Dec. 2010

[4] M. Cheriet, N. Kharma, C. Y. Suen. Character Recognition Systems: A guide for students and practitioners. USA: Wiley, 2007, pp. 30-36.

[5] Z.-Q. Liu, J. Cai, R. Buse. Handwriting Recognition: Soft computing and probabilistic approaches. Germany: Springer, 2003, pp. 17-29.

[6] C. Singh, N. Bhatia, and A. Kaur, "Hough transform based fast skew detection and accurate skew correction methods." Pattern Recognition, vol. 41, New York, USA, Dec. 2008.

[7] P. Saragiotis and N. Papamarkos, "Local Skew Correction in Documents." International Journal of Pattern Recognition and Artificial Intelligence, vol. 22, pp. 691-710, 2008.

[8] Slavik, P.; Govindaraju, V., "Equivalence of different methods for slant and skew corrections in word recognition applications," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.23, no.3, pp.323-326, Mar 2001.

[9] Dumitras, T.; Lee, M.; Quinones, P.; Smailagic, A.; Siewiorek, D.; Narasimhan, P,; "Eye of the Beholder: Phone-Based Text-Recognition for the Visually-Impaired," Wearable Computers, 2006 10th IEEE International Symposium, pp. 145-146, 11-14 Oct. 2006.

[10] V. Gaudissart et al., "SYPOLE: mobile reading assistant for blind people," In International Conference on Speech and Computer (SPECOM), 2004, pp. 538-544.

[11] R. C. Kurzweil, P. Albrecht, L. Gibson, L. Lvovsky, "TILT ADJUSTMENT FOR OPTICAL CHARACTER RECOGNITION IN PORTABLE READING MACHINE", United States Patent Application Publication, US2006/0071950 A1, Apr. 6 2006.

[12] Qadri M.T., Asif M., 2009, "Automatic Number Plate Recognition System for Vehicle Identification Using Optical Character Recognition" International Conference on Education Technology and Computer, 2009. ICETC '09, pp.335 – 338.

[13] Deng H., Stathopoulos G., Suen, C.Y., 2009, "Error-Correcting Output Coding for the Convolutional Neural Network for Optical Character Recognition" International Conference on Document Analysis and Recognition, 2009. ICDAR '09. 10th , pp. 581 – 585.

[14] Ganapathy V., Lean C.C.H., 2007, "Optical Character Recognition Program for Images of Printed Text using a Neural Network," IEEE Conference on, Industrial Technology, 2006. ICIT 2006, pp.1171–1176.

[15] Ramanathan R., Ponmathavan S., Valliappan N., Thaneshwaran L., Nair A.S., Soman K.P., 2010, "Optical Character Recognition for English and Tamil Using Support Vector Machines," International Conference on Advances in Computing, Control, & Telecommunication Technologies, 2009. ACT '09, pp. 610 – 612.

[16] Prasad R., Saleem S., Kamali M., Meermeier R., Natarajan P., 2009, "Improvements in hidden Markov model based Arabic OCR," International Conference on Pattern Recognition, 2008. ICPR 2008. 19th, pp.1-4.

[17] S.B. Patil and N.V. Subbareddy, "Neural Network Based System for Script Identification in Indian Documents," Sadhana, vol. 27, no. 1, pp. 83-97, Feb. 2002.

[18] Z. Chi, Q. Wang, and W.-C. Siu, "Hierarchical Content Classification and Script Determination for Automatic Document Image Processing," Pattern Recognition, vol. 36, no. 11, pp. 2483-2500, Nov. 2003.

[19] S. Kanoun, A. Ennaji, Y. Lecourtier, and A.M. Alimi, "Script and Nature Differentiation for Arabic and Latin Text Images," Proc. Int'l Workshop Frontiers in Handwriting Recognition, pp. 309-313, Aug. 2002.

[20] L. Zhou, Y. Lu, and C.L. Tan, "Bangla/English Script Identification Based on Analysis of Connected Component Profiles," Proc. Int'l Workshop Document Analysis Systems, pp. 243-254, Feb. 2006.

[21] U. Pal and B.B. Chaudhuri, "Identification of Different Script Lines from Multi-Script Documents," Image and Vision Computing, vol. 20, nos. 13/14, pp. 945-954, Dec. 2002.

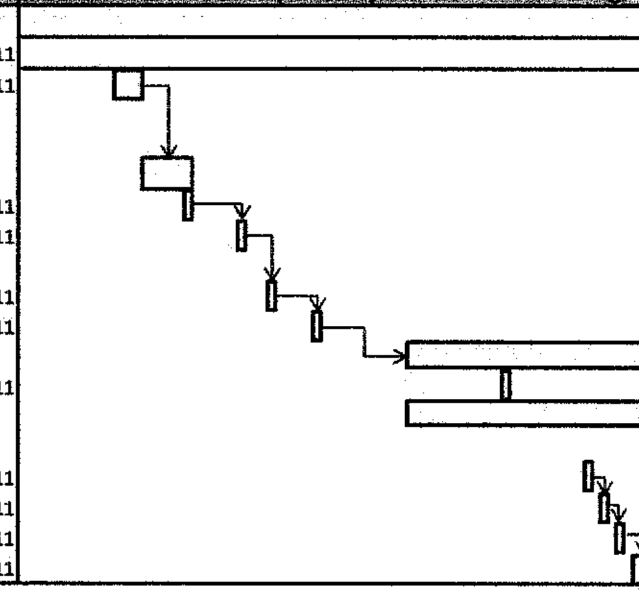
[22] U. Pal, S. Sinha, and B.B. Chaudhuri, "Multi-Script Line Identification from Indian Documents," Proc. Int'l Conf. Document Analysis and Recognition, pp. 880-884, Aug. 2003.

APPENDIXES

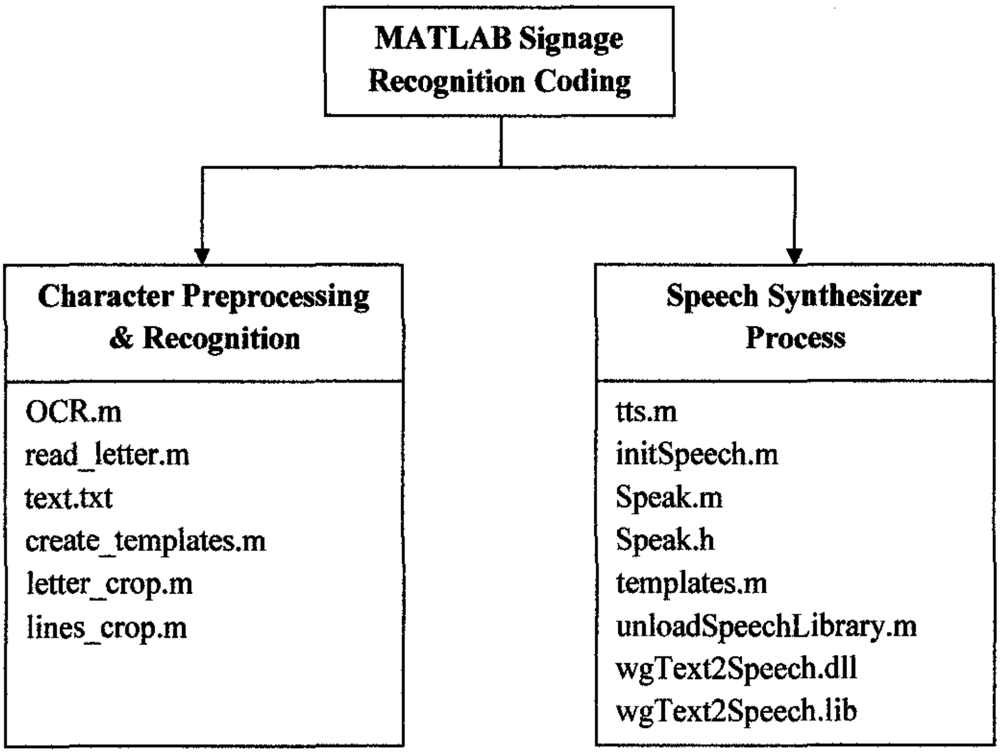
Appendix A – FYP II Gant Chart

Public OCR Signage Recognition with Skew & Slant Correction for Visually Impaired People - FYP II

No.	Task Name	Start	Finish	Jan-11 Feb-11 Mar-11 Apr-11 May-11 Jun-11 Jul-11 Aug-11
1	Start Project	24/1/2011	30/9/2011	
2	Submit Proposal and Get Project Approval	7/2/2011	7/2/2011	
3	Research and Development			
	Phase 1: Planning			
	Critical Review of Related Work on OCR; Skew & Slant Correction			
	Submission of Extended Proposal	4/3/2011	4/3/2011	
	Submission of Proposal Defend and Progress Evaluation	25/3/2011	25/3/2011	
	Phase 2: Analysis and Tools Used			
	Submission of Interim Report	8/4/2011	8/4/2011	
	Submission of Technical Report	22/4/2011	22/4/2011	
	Phase 3: Design and Development			
	Submission of Progress Report	4/7/2011	4/7/2011	
	Phase 4: Evaluation, Testing and Further Enhancement			
4	Project Submission			
	Pre Edx	3/8/2011	3/8/2011	
	Submission of Dissertation	9/8/2011	9/8/2011	
	VIVA	19/8/2011	19/8/2011	
	Submission of Final Dissertation and Technical Report	24/8/2011	26/8/2011	



Appendix B – OCR Hybrid Recognition Model Coding



Appendix C – MATLAB Coding

Character Preprocessing & Recognition - OCR.m

```
warning off %#ok<WNOFF>
% Clear all
clc, close all, clear all

%-------------------------------------------------------------------
%Start Image Pre-processing
%Step 1: Read image
%Step 2: Convert to grayscale image and perform egde detection using Canny
%        method
%Step 3: Perform Hough transformation to find prominent lines
%Step 4: Perform Shearing transformation if skew/slant is greater than or
%        less than 50 degree
%Step 5: Convert image to binary image
%Step 6: Noise reduction: Remove all objects containing fewer than 50 pixels
%-------------------------------------------------------------------

% Read image
[inputfilename,dirname] = uigetfile('*.*');
inputfilename = [dirname, inputfilename];
imagen = imread(inputfilename);
imagen = imagen(:,1:end-3,:); % remove small white band on the side
imshow(imagen);
% Convert to grayscale and perform edge detection using Canny method
Iedge = edge(rgb2gray(imagen), 'canny');
figure,imshow(Iedge)
```

```
% Hough transformation
[H T R] = hough(Iedge);
P  = houghpeaks(H, 4, 'threshold',ceil(0.75*max(H(:))));
lines = houghlines(Iedge, T, R, P);
theta = 90;
if(theta > 89 || theta < -89)
% Shearing transformation
slopes = vertcat(lines.point2) - vertcat(lines.point1);
slopes = slopes(:,2) ./ slopes(:,1);
TFORM = maketform('affine', [1 -slopes(1) 0 ; 0 1 0 ; 0 0 1]);
imagen2 = imtransform(imagen, TFORM);
% show accumlation matrix and peaks
figure, imshow(imadjust(mat2gray(H)), [], 'XData',T, 'YData',R, 'InitialMagnification','fit')
xlabel('\theta (degrees)'), ylabel('\rho'), colormap(hot), colorbar
hold on, plot(T(P(:,2)), R(P(:,1)), 'gs', 'LineWidth',2), hold off
axis on, axis normal
% show image with lines overlayed, and the aligned/rotated image
figure
subplot(121), imshow(imagen), hold on
for k = 1:length(lines)
   xy = [lines(k).point1; lines(k).point2];
   plot(xy(:,1), xy(:,2), 'g.-', 'LineWidth',2);
end, hold off
subplot(122), imshow(imagen2)
end
% Convert to binary image
threshold = graythresh(imagen);
imagen =~im2bw(imagen,threshold);
% Noise reduction: Remove all object containing fewer than 50 pixels
imagen = bwareaopen(imagen,50);
% End of Image Pre-processing
```

```
%----------------------------------------------------------------------
%%

%Storage matrix word from image
word=[ ];
re=imagen;
%Opens 'text.txt' as file for write
fid = fopen('text.txt', 'wt');
% Load templates
load templates
global templates
% Compute the number of letters in template file
num_letras=size(templates,2);
while 1
   %Fcn 'lines_crop' separate lines in text
   [fl re]=lines_crop(re); %fl= first line, re= remaining image
   imgn=fl;
   n=0;
   %Uncomment line below to see lines one by one
   %figure,imshow(fl);pause(2)
   %------------------------------------------------------------
     spacevector = [];     % to compute the total spaces betweeen
                     % adjacent letter
   rc = fl;
     while 1
      %Fcn 'letter_crop' separate letters in a line
     [fc rc space]=letter_crop(rc); %fc =  first letter in the line
                            %rc =  remaining cropped line
                            %space = space between the letter
                            %  cropped and the next letter
     %uncomment below line to see letters one by one
```

```
%figure,imshow(fc);pause(0.5)
img_r = imresize(fc,[42 24]);  %resize letter so that correlation
                               %can be performed
n = n + 1;
spacevector(n)=space;

%-------------------------------------------------------------------
% Call fcn to convert image to text
letter = read_letter(img_r,num_letras);
    %letter concatenation
word = [word letter];
    if isempty(rc)  %breaks loop when there are no more characters
  break;
 end
end
    %-------------------------------------------------------------------
      %
max_space = max(spacevector);
no_spaces = 0;
  for x= 1:n  %loop to introduce space at requisite locations
 if spacevector(x+no_spaces)> (1.0 * max_space) %highlight here where I change the
maximum space X by 100%
   no_spaces = no_spaces + 1;
    for m = x:n
     word(n+x-m+no_spaces)=word(n+x-m+no_spaces-1);
    end
   word(x+no_spaces) = ' ';
   spacevector = [0 spacevector];
 end
end
```

```
    %fprintf(fid,'%s\n',lower(word));%Write 'word' in text file (lower)
    fprintf(fid,'%s\n',word);%Write 'word' in text file (upper)
    % Clear 'word' variable
    word=[ ];
    %When the sentences finish, breaks the loop
    if isempty(re)  %See variable 're' in Fcn 'lines'
        break
    end
end
% Read character string from text.txt
A = textread('text.txt', '%c');
% Open 'text.txt' file
winopen('text.txt')
% Text to speech function
tts(A)
% Close the 'text.txt' file
fclose(fid);
% Clear all
clc, clear all
```