Abstract— It is a challenge for visually impaired people (VIPs) to navigate independently whenever they attempt to find their way in unfamiliar buildings searching for amenities (i.e. exits, ladies/gents toilets) even with a walking stick or a guide dog. Camera-based computer vision systems have the potential to assist VIPs in independent navigation or way finding in unfamiliar places. To leverage on previous research of Signage Recognition Framework which could only recognize public signage with slanted angle less than 30°, an improved OCR signage recognition model with skew and slant correction in public signage is presented. The proposed OCR method consists of Canny edge detection algorithm, Hough Transformation and Shearing Transformation were used to detect and correct skewed and slanted images. The proposed model would capture a public signage image, compare the image in the database using template matching algorithm and convert to machine readable text in a text file. The text will then be processed by Microsoft Speech Application Program Interface (SAPI) speech synthesizer and translated to voice as output. Experiments were conducted on 5 blind folded subjects to test the performance of the model. The proposed OCR recognition model has achieved satisfactory recognition rate of 82.7%.

Keywords: Optical Character Recognition, Hough Transformation, Shearing Transformation, slant correction, visually impaired.

I. INTRODUCTION

This research was inspired by special group of people in National Council of the Blind Malaysia (NCBM) who have difficulties in way findings independently due to their vision [1]. There are an estimated 45 million blind people and 135 million visually impaired people (VIPs) worldwide [2]. The aim of the research is to initiate an affordable gadget in order to help VIPs in their daily activities. The objective of the research is to initiate and improve the accuracy of recognition rate of the OCR system. Other factors contributing to low recognition are image quality and resolution, layout of the character, amount of noise, skewed and slanted angles, type of script being used and also type of character either printed or handwritten.

II. RELATED WORK

Extensive research has been conducted on OCR using different algorithms. OCR is the process is which a paper document is optically scanned and then converted into electronic format by recognizing and associating symbolic identity with every individual character in the document [4]. OCR processing includes several different steps that serve different functions which are 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.

There are several basic steps in image preprocessing which tailors to the OCR. The image and data preprocessing is the first important step in image recognition. It serves 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 [5]. A common problem in image preprocessing is the skew detection and correction or also known as tilt detection and correction. 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 [6]. The reason that skew or tilt should be eliminated is because it could largely reduces the accuracy of recognition rate of the OCR system.

There are many methods which have been identified by various researches. The proposed methods in the literature are...
rotation, nearest-neighbor clustering, Hough Transformation cross-correlation and morphological transformation [7] [8].

Saragiotis et al. proposed method in skew detection called connected component analysis [9]. The method works on grayscale image, where Gaussian filter is used to remove all the noise contained in the image. He attempts to find all connected elements from image lines to constructing 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.

To address the problem of skewed angle, a rotation algorithm is used to perform skew correction. The skew correction is computed using Eq.1 [10]:

\[
\begin{pmatrix}
    x' \\
    y'
\end{pmatrix} =
\begin{pmatrix}
    \cos \theta & \sin \theta \\
    -\sin \theta & \cos \theta
\end{pmatrix}
\begin{pmatrix}
    x \\
    y
\end{pmatrix}
\]

(1)

The character inclination found in images will also result in a low accuracy rate of the OCR system. Slavik et al. applied shear transformation to estimate and correct the slant images [10].

Many attempts for using OCR to help the VIPs have been published with relevant literature. One efficient way to use OCR to help the VIPs is by converting the text from physical document into a machine readable text, and then converts the recognized text into speech using speech synthesizer. Dumitras et al. developed a phone-based text-recognition framework for the visually impaired people using a mobile phone [11]. 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. Similarly, Gaudissart et al. built an automatic text reading assistant for VIPs via mobile camera-phone [12]. Nowadays, iPhone has launched its iPhone Apps namely ZoomReader to help the VIPs to read with the same framework. These research projects attempted to provide autonomy for VIPs in extracting textual information from banknotes, train schedules, restaurant menu, printed instructions and so forth.

Kurzweil et al. proposed a reading machine for the VIPs where it is being used to read printed documents such as newspapers, books or journals [13]. 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.

There are also multiple usages for OCR which suit different purpose. OCR has also been extensively used in car plate recognition [14]. Moreover in order to achieve the different desired results, different popular techniques have also been implemented. There are Neural Network, Hidden Markov Models and Support Vector Machines. OCR implementation in Neural Network is discussed in [15] and [16], while discussion for Support Vector Machines is being used in [17]. Prasad et al. used Hidden Markov Model algorithm for Arabic OCR [18].

### III. METHODOLOGY

RAD Prototyping is used to perform the analysis, design and implementation concurrently and repeatedly until the completed final implementation.

The advantage of using Prototyping rather than other methods is that, developer will be able to modify the system at an early stage of system development to overcome any shortcomings, and enable them to test the prototype system continuously. The flowchart of proposed OCR recognition model is illustrated in Figure 1:

![Flowchart of Proposed Hybrid OCR Recognition Model](image-url)
The proposed OCR Recognition Model in Figure 1 demonstrates the overall process of OCR recognition model from capturing image, to character preprocessing and character recognition and finally the speech synthesis.

A. Capturing Image

User will capture image of public signage using Panasonic Lumix DMC-FS62 phone camera with 10 Megapixels. The image will be stored in .jpeg format. 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 VIPs.

B. 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. The image in public signage contains high amount of noise and possibility of slight to moderate skew and slanted character. The type of script being used is alphabetic script and most of the public signage are using printed characters. Thus, several image preprocessing techniques are applied to tailor to these factors.

The improved model is implemented in 5 processes as shown in Figure 2. The processes include convert to grayscale image, Canny edge detection, Hough Transformation, Shearing Transformation, convert to binary image and noise reduction.

![Figure 2: OCR hybrid recognition model for character preprocessing](image)

First, the image taken will be converted into grayscale image and Canny edge detection algorithm is used to detect all edges of characters in the image using Matlab 7.10. The Canny algorithm has been extensively because it involves multistage edge detection which yields better results. The image is convolved using a Gaussian filter based on Eq. 2:

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

The intensity of the image is being computed to detect all edges of the characters which are horizontal, vertical and diagonal edges using Eq. 3:

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

Hough Transformation is deployed 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 edge detection algorithm. To compute the straight lines in Hough Transformation, parameters identified are \(r\) and \(\theta\) where \(r\) is the distance from the origin and character’s line, and \(\theta\) being the angle from origin to the closest point respectively. Hough Transformation is shown in Eq. 4 [8]:

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

The characters in the image will be determined regardless they are slanted or not. The slant is estimated by calculating the coordinate for four lines surrounding the characters using the following Eq. 5 [5]:

$$y = x + \beta_1$$

$$y = x + \beta_2$$

$$y = x + \beta_3$$

$$y = x + \beta_4$$

The slant angle is then calculated as \(\theta = \arctan(B/A)\), where \(B = \frac{\beta_3 + \beta_1 - \beta_2 - \beta_4}{2}\) and \(A = \text{Height of the character} \). After that, the horizontal shear transformation is applied to the image to correct the slant by shifting its pixels to the left or right according to the estimated \(\theta\). The shear transformation algorithm is computed using Eq. 6 [5]:

$$x' = x - y \cdot \tan \theta$$

$$y' = y$$

In Figure 3, Shearing Transformation with Affine Transformation is applied to correct and rotate skewed or slanted characters into horizontal aligned characters using Eq. 7 [5]:

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

![Figure 3: Corrected character image after Shearing Transformation](image)
The grayscale image will be converted into binary image. Noise in the image will be removed where all objects containing fewer than 50 x 50 pixels will be eliminated. This is because each character image (A-Z, a-z, 0-9) was constructed and normalized into 50 x 50 pixels in the database.

After character preprocessing is done, character recognition is performed by comparing the character image with those in the template database. Template matching algorithm is applied to determine the similarity between the binary image and the character image in the database. The recognized character will be stored in text.txt file.

C. Text To Speech

After the character preprocessing and recognition, text in the text.txt file will be converted into speech using Microsoft Speech Application Program Interface (SAPI). The command to read character string from text.txt file is:

\[ A = \text{textread('text.txt', '%c');} \]

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

IV. RESULT & DISCUSSIONS

A. Experimental Setup

There are 15 different samples of public signage and some of them are skewed or slanted as shown in Table I. The samples were tested by 5 subjects, where all of them were blind folded. The subjects comprised of 4 females and 1 male age ranged from 20 to 25 years old. If the signage is recognized correctly (T = True), a tick (\(\bigvee\)) is put in the entry. Otherwise the signage is not recognized (F = False).

We also collected image OCR dataset which is a database that contains 26 upper case letter images in the dataset, 26 lower case letter images in the dataset and 10 numeric images in the dataset. All of the images are stored in .jpg file format. All captured images must be normalized (50 x 50 pixels) to facilitate template matching process.
B. Experimental Results

From the results in Table II, 15 samples of public signage were tested. The subjects (blind-folded) snapped pictures of public signage in different angles. To find the Recognition Rate (RR), Eq. 8 is used:

\[
RR = \frac{\text{Classified Character}}{\text{Total Number of Characters}} \times 100\%
\]  

A total of 62 out of 75 samples were recognized:

\[
RR = \frac{62}{75} \times 100\% = 82.7\%
\]

C. Discussion

The experiment was repeated a few times in trial and error. It was noted that errors seem to always occur for item 4 in Table I. The proposed OCR prototype had mistakenly generated a character 'O' instead of the character 'C' during the recognition process for the skewed and slanted signage in item 4. Further testing and research are required on more signage samples to improve its recognition rate.

From the preliminary study, the current OCR signage recognition prototype has recognition rate of 82.7%.

V. CONCLUSION

The proposed model shows the proof of concept for OCR recognition model where it has successfully achieved a recognition rate of 82.7%. Although the character is skewed or slanted in various angles, the proposed prototype can detect and correct the skewed and slanted characters in an image. If the character is successfully recognized, it produces a voice output corresponding to the input character.

For future work, it is recommended to design a wearable camera or compact gadget for the convenience of VIPs. It is our hope that the model will be further improved by translating English word into commonly used languages in Malaysia such as Malay, mandarin and Tamil languages.

ACKNOWLEDGMENT

Special thanks to the National Council of the Blind Malaysia (NCBM) for providing support in this research. We also thank all staff who either directly or indirectly involved in the project.