Audio Assistance for Vision Impaired Individual To Recognize Graphical Content on Print Disable Documents

N.D.U.Gamage[1], K.W.C.Jayadewa[2], S.M.N.K.B.Senanayake[3] , K.L.A.D.Udeshitha[4], J.A.D.C.A.Jayakody[5]
Department of Information Technology and Department of Information Systems Engineering
Sri Lanka Institute of Information Technology (SLIIT), Malabe, Sri Lanka.
[1]narmadadg@gmail.com, [2]venura91@gmail.com, [3]bhagyasenanayake24@gmail.com, [4]ashan 901121@gmail.com
[5]anuradha.j@sliit.lk

***Abstract* - A print disabled person is a person who cannot effectively read printed documents because of visual impairment. The print disability prevents a person from gaining information from printed material in the standard way, and requires them to utilize alternative methods to access that information. Hence, this paper presents a mobile-based audio assistance to read textual documents, which contains graphical contents such as images, tables and mathematical equations to overcome above-mentioned challenge. Further, this paper discusses the test results and evaluations to justify feasibility of the proposed solution.**

Keywords—vision impairment, textual document, graphical contents, OCR technology

I. INTRODUCTION

According to the Rauha Maarmo, the project manager of Celia library "print disabled person is a person who cannot effectively read print because of a visual, physical, perceptual, developmental, cognitive, or learning disability" [1]. Print disability prevents a person from obtaining information from the printed material in a regular manner, and requires alternative methods to access that information. Print disabilities include visual impairments, learning disabilities, or physical disabilities that impede the ability to manipulate a book in some way [2].The term was coined by George Kerscher, a pioneer in digital talking books [3]. Visual impaired people require third party assistance or brail converted material to read printed documents even though the digital books are available. Further, there are technological advancements to scan and read the printed materials using a variety of software as well as apps. Existing applications are unable to read equations, images, tables in an accurate way as sighted people do. Therefore, a solution capable of delivering higher accuracy to read printed material with existing technology plays a significant role in an assistive technology research area.

This paper presents mobile-based feasible solution named as "Schmoozer" that provides audio assistance to navigate through mobile application, autofocused image capturing of printed papers, store captured images, classify selected text, images, tables and equations and read aloud generated digitized text. Therefore, "Schmoozer" would allow vision impaired individuals to unbraided document reading without others interaction. The remaining section of the paper contains the methodology, implementation and conclusions.

II. BACKGROUND

Vision is one of the main senses people use to see, grasp knowledge, experience and the world. According to WHO, it is estimated that [4] 285 million people suffer from vision impairment or blindness in the entire world. Sri Lankan contribution [5] to this statistic is 996,939 among 20 million populations. Mentioned statistics prove that a considerable amount of the population cannot view anything or view properly. However, people who suffer from incomplete vision or blindness also have rights to acquire information. "Audio assistance for visually impaired people to understand graphical content on textual documents", which is the research project, introduces a solution to ensure the grasping information right of visually impaired people. Moreover, the main target is visually impaired students.

Thousands of research have been carried out for many years to ensure the information, education and technology access for blind or visually impaired individuals. Evolution of technology leads to inventing a number of ways to connect visually impaired individuals with technical tools. The term "Assistive technology" refers to any "product, device, or equipment, whether acquired commercially, modified or customized, that is used to maintain, increase, or improve the functional capabilities of individuals with disabilities" [6]. Due to it, they require special attention [7] to educate vision impaired individuals. The assistive technology is recommended to eliminate barriers in education and employment for visually impaired individuals where they

could be able to complete schoolwork, explore, take tests or read books along with sighted people [8].

Audio assistance or text reader is a recognized assistive technology mechanism that supports visually impaired people to listen to the contents of printed or handwritten documents. As the main target is visually impaired students, it is important to have a reader, which enables to identify and state the contents in printed documents implicitly. It will allow students to listen to their textbooks, assignments, exam papers within a short period after analyzing.

The textual document contains images, mathematical equations and tables. Therefore, it is a must to recite those graphical contents for the visually impaired person to get the accurate idea of the document. Image processing is a method to translate an image into digital form and perform edit or searching operations on it, in order to acquire an improved image or to abstract useful information from the image [9], which is a highly recommended technology for autonomous graphical contents identifying the mechanism. Machine learning is another mechanism to train electronic devices like computers to grow and change when exposed to new data. It is a type of an artificial intelligence that allows computers to learn itself without explicitly programmed [10].Consequently, mentioned technics used to image recognition in an effective manner.

The research team provides a solution with a combination of above-mentioned mechanisms for autonomous document recognition to heighten the process of clasping knowledge by providing quick access to the information without the barrier of physical disability. “Schmoozer” is a mobile application that has the capability to auto capture a print disable document and upload it into server automatically, identify image regions accurately, generate digitized text from graphical images, mathematical equations and tables, and redirect digitized text into a text to speech engine that has the capability of reciting document content to the visually impaired individual.

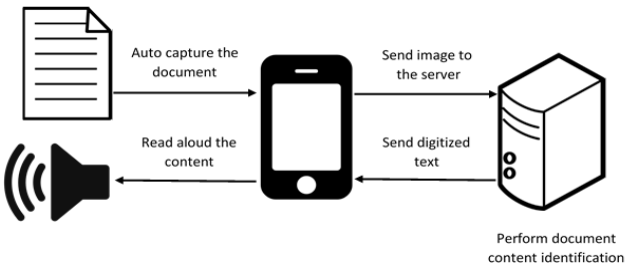

Fig. 1. System Overview

Figure 1 illustrates the system overview of the research solution. The research team hopes that the assistive technology document reading solution will lighten up the dreams of the visually impaired individuals in near future.

III. Research Objectives

The main objective of this application is to encourage visually impaired students to gather knowledge in the easiest way. Without reading document or books gaining knowledge is a very challenging task. Therefore, research group is introducing a solution to help the blind students to grab the knowledge of the documents without seeking others help. With this application, students can do their studies without any problem. The solution also will be very helpful to the teachers and parents because they do not want to read aloud all the text document contents to their children. As an additional feature, the application supports voice command inputs. While listening to the document, user can stop the reading by providing voice command as an input. All the implemented features will provide fascinating experience to vision impaired people to see the world by their own.

IV. METHODOLOGY

This section discusses the methodology carried out to implement the proposed portable solution as mentioned in the introduction.

Vision impaired individuals have to auto capture and upload an image of print disable document into the “Schmoozer IIS server” to start the print disable document reading process. Uploaded image go through image cropping function and graphical region identification function. When graphical region identified as a graphical image, mathematical equation or table, it will automatically redirect to graphical image identification, equation identification or table identification function accordingly. Functions implemented throughout the application development are listed down below.

A. Edge Detection and Image Auto Capturing

It is difficult for a blind person to capture an image by identifying edges of the document. Therefore, edge detecting, auto capturing and image uploading steps are used to complete the requirement. Fig. 2 shows the flow of auto capturing process

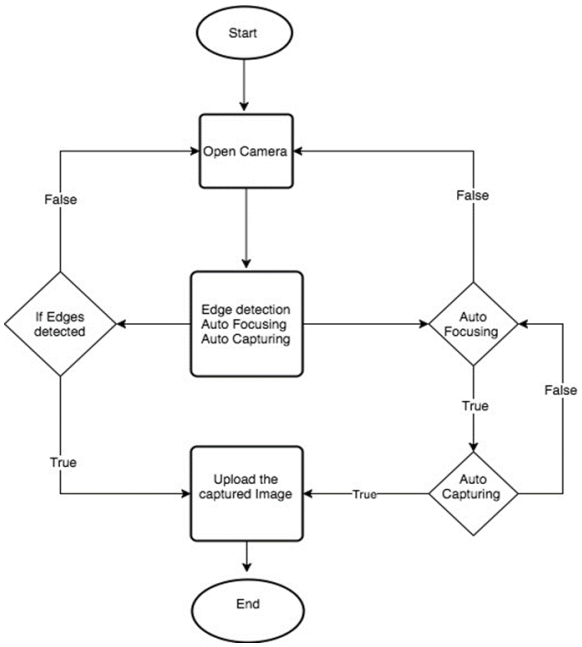

Fig. 2. The flow of image auto capturing process.

1) Edge detecting

The mobile application captures the printed document when the person places camera towards the document by detecting the edges of it. Canny edge detecting algorithm was used to detect the edges of the document and Gaussian filter applied to smooth the image in order to remove the noise. It is essential to avoid false detecting caused by noise. Then get the intensity gradient of the image. Non-maximum suppression is going to get rid of spurious response. After that, potential edges are determined to apply a double threshold. Finally, hysteresis tracks the edges smoothly.

2) Auto Capturing

The mobile application has the capability of auto capture and uploads the image of printed document to the server after detecting document edges smoothly. To auto focused the document, "Camera. Autofocus Callback interface" used to notify on completion of camera autofocus. The auto capturing function completes by saving a photo of the document as a byte array and automatically upload to the server using REST protocol via the internet.

B. Graphical Content Identification

Text, graphical images, mathematical equations and tables are the main graphical components in a document. Separate identification of document content is the core function in "Schmoozer" mobile application, which is totally based on image processing. Image processing functions are not suitable to execute inside the mobile phone due to mobile phone's processing speed is insufficient. Therefore, research group decided to use Internet Information Services (IIS, formerly Internet Information Server) [11] for the entire image processing related functions. At the end of the image processing the processed outputs were redirected to the Text to Speech (TTS) engine of the user's mobile phone, which delivers audio format of the document to the users. Following steps were carried out to identify graphical contents printed on documents.

1) Graphical regions identification

Text, graphical images, equations and tables are the main regions in print disable documents.

Authors collected images of mathematical equations, text, tables and graphical images and stored them in separate folders. Then apply HOG feature extraction and Support Vector Machine (SVM) algorithm on the collected set of images to convert them into a trained data set.

When a vision impaired individual uploads an image of a document to the IIS server, the image is cropped and going through "Graphical region identification function". The function has the capability of extracting HOG features of cropped image and suggest the region using SVM.SVM algorithm contains a predefined function called prediction. The trained dataset and a single test data (cropped image segment) input to the function. The function compares HOG features of the input segments with the HOG features stored inside trained dataset to classify and label the input segment. (Ex: Input segment is an equation, Input segment is a table).Based on the label generated by SVM, the "Schmoozer IIS server" redirect identified image segment to graphical image identification, mathematical equation identification or table identification functions. All steps in "Graphical region identification function" converted into DLL file to use it in the web service.

2) Graphical image identification

The main purpose of the "Graphical image identification function" is separately recognizes and label the images available in print disable document. Authors implemented a strong dataset, which contains a huge amount of images of living beings and many single objects with significant behavioral changes. Then applied HOG feature extraction to identify unique features in the graphical image and convert into a decimal value to store in a database. The database was trained using SVM algorithm to convert the dataset into the trained dataset.

When "Graphical image identification function" obtain a cropped and identified image segment from graphical region identification function it extracts HOG features and converts into a decimal value. The decimal value was compared with the trained data set to find the name of the object. Prediction quality of SVM algorithm used for decimal value comparison task. If the decimal value of input test image matches with any decimal value inside the trained database, the labeling value of that image is assigned to a variable and generated a meaningful digitized text that describes the graphical image. Then it automatically opens a text file and writes all the digitized outputs on it. All steps in "Graphical image identification function" converted into DLL file to use it in the web service.

3) Mathematical equation identification

The main purpose of the "Mathematical equation identification function" is to separately recognize characters and symbols in a mathematical equation that are printed on documents, and convert them into digitized text. The first step was designing of mathematical symbols and numbers using Photoshop in bitmap format. Bitmap files are easy to create and pixel data stored in a bitmap file could be accomplished by using a set of coordinates that allow the data to be conceptualized as a grid. Therefore, all the major mathematical characters were created in bitmap format and stored inside a folder called "letters_numbers. Then created a Matlab template and match each character /mathematical symbol/number into created bitmap images.

The "Mathematical equation identification function" obtained input images from the "Graphical regions identification function". The segmented image with equations went through "Equation" function and faced to set of preprocessing tasks. Read image, Convert to grayscale, Convert to BW, Remove all object containing less than 9 pixels, Resize letter (same size of the template) are the preprocess followed to adjust the input equation match with the created template. After preprocessing, each single character in the equation converted into a binary value. Finally match every character of the input image with the binary value obtained from the created template and regenerate the equation as digitized text. Function's output is set of single digitized values for single character/symbol in the equation. Then it automatically opens a text file and writes all the digitized outputs on it. All steps in "Mathematical equation identification function" converted into DLL file to use it in the web service.

4) Table data identification

The main purpose of the "Table data identification function" is to recognize the type of data table (whether it is 2columns or 3columns) and convert table data into meaningful digitized text. Implement a strong dataset with a huge number of 2 column and 3 column tables is the foremost step in table identification process. Then applied HOG feature extraction to identify unique features of 2 column and 3 column tables and convert into a decimal value to store in a database The database was trained using SVM algorithm method to convert the dataset into the trained dataset..(If user required identifying more columns, it can be done easily by adding more columns tables to the collected images of tables.)

When "Table data identification function" obtain a cropped and identified table segment from graphical region identification function it extracts HOG features and converts into a decimal value. The decimal value was compared in with the trained data set to find the type of the table whether the table has 2 or 3 columns. The data inside the image of the table was read using in build OCR function of Matlab and wrote the data into a text file.

All steps in "Table data identification function" converted into DLL file to use it in the web service

Table data reading functions were built inside the web service. "Tableread" function implementation identify the table type by observing Matlab "Table type identification" function's output text file. Text file, which contains table data redirect to the special functions that have the capability of generating a meaningful output according to the type of the table. If the input image is a two-column table, the text file obtains from Matlab OCR function pass through "TwoColTableReading" function and provide a meaningful sentence that implies table headers and table data separately by adding the values in a text file into two element array. If the input image is a three-column table, the text file obtains from Matlab OCR function pass through "ThreeColTableReading" function and generate a meaningful sentence that implies three table headers and table data separately by adding the values in text file into three element array.

C. Implement web service.

Mobile application and the server connectivity were established through web service developed using c# language. Generated DLL files for image regioning, graphical image identification, mathematical equation identification and table data identification are added to the web service.

Then all the functions were published in IIS server to host the document content reading functionalities. IIS server bridge the server processes with the mobile application to cater vision-impaired individuals.

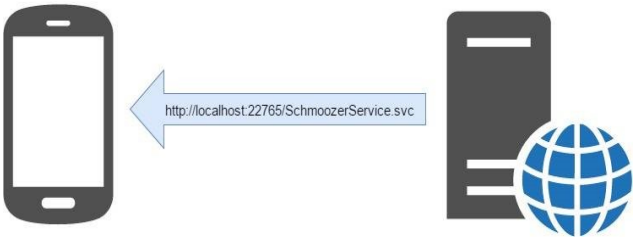

Fig. 3. Connectivity between mobile phone and the server

Figure 3 illustrates the connectivity between mobile phone and the server. When user uploads image to the server, it calls corresponding functions for creating objects.

D. Audio Assistance Functionalities

When a vision impaired individual captures an image of a document and auto uploads to the server, it goes through a set of graphical content identification functions mentioned above and generate digitized text that allow to read aloud by TTS engine. Audio assistance function collects all the digitized text files and arrange as a single paragraph of a document.

Then generated paragraph go through the implemented dictionary function to correct spelling errors. Finally, formatted sentence has been sent to the mobile phone.

The finalized text received by the mobile phone immediately redirects to the text to speech engine. The TTS engine has been designed with the compatible for Android devices above "Kit Kat" version. The process of the TTS engine is covered input text sentence into the speech format. Designed engine includes key features such as familiar speech accent with the adjustable speech rate. After generating the output from the TTS engine, it delivers to the user. To distribute the output with attractive and user-friendly style design the mobile app front end. The front end of the schmoozer application mainly contain two option such as (Vision Impaired) VI MODE and the SIGHTED MODE

1) VI mode

VI mode is specially designed to improve vision impaired students studying specifications. User Interface (UI) design with button click events is challenging task because of user unable to see the buttons on the application. Therefore, to avoid that problem as well as improve efficiency and effectiveness "Schmoozer" UI design follows four gesture recognition activities. Single tap event gives all the user guidelines for the application handling process to continue document reading. Long press event starts the reading document content to the user. The document reading synthesis is provided in clear and meaningful voice accent with the adjustable delivery speed. Therefore, the user can understand the document content smoothly. Scroll event use to stop the speech. Double tap event will open mic to input voice commands. By providing a voice command such as "STOP" or "PLAY" user can handle reading. Using this gesture events user can grasp the document contain knowledge without any difficulty.

Vibration is enabled with all these gestures to feel the gesture activities to the vision impaired individuals. To avoid close the application by accidentally pressing back button it disables the back button when the application is open. By pressing home button user able to exit the application.

2) SIGHTED mode

SIGHTED mode is designed for the ordinary people to read the document content in a meaningful manner. Sighted mode contains the simple button click events. Using this buttons user can handle the application very easily.

Without purchasing high cost, awkward or ineffective document reading product or application both sighted and visually impaired person can use schmoozer mobile application to gain the document content information and improve their depth of knowledge.

V. RESULT AND DISCUSSION

The main target of the research is to build an application that read document contents for vision impaired user groups. Therefore, the application consists of autofocusing and image capturing techniques to reduce the time taken to capture the image. The captured image converts into a byte array before sending it to the backend server. It could be able to increase the sending speed because it will compact the size of the file.

After the upload the captured image, graphical region identification is the foremost task of converting print disable document into digitized document. The accuracy of region identification was calculated with 40 test images

The accuracy of the function is in high volume. Therefore, the authors are able to prove that HOG feature extraction with the combination of SVM algorithm is a good solution for extract real content inside captured image.

TABLE I. ACCURACY CALCULATION FOR GRAPHICAL REGION IDENTIFICATION

Scenario	Number of occurrences	Number of Occurrence meets expected output
Identify image regions	10	10
Identify text regions	10	10
Identify table regions	10	8
Identify equation regions	10	10
Accuracy = $\frac{Number\ of\ correct\ output\ occurance}{Number\ of\ occurance}$ x 100% = $\frac{38}{40}$ x 100% = 95%		

The graphical image identification function is also totally based on HOG feature extraction and SVM. Even though accuracy value calculated for regions, the authors did the same kind of test with different types of images. The sample data size was 45

TABLE II ACCURACY CALCULATION FOR GRAPHICAL IMAGE IDENTIFICATION

Scenario	Number of occurrences	Number of Occurrence meets expected output
Identify significant images accurately	15	15
Identify behavior variations of an image accurately	15	14
Identify color variations of an image accurately	15	15
Accuracy = $\frac{Number\ of\ correct\ output\ occurance}{Number\ of\ occurance}$ x 100% = $\frac{44}{45}$ x100% = 97.778% ≈ 98%		

Mathematical equations in captured images should be converted into machine editable format in order to recite by the TTS engine. Due to some existing TTS engines were unable to read mathematical symbols like π,∞, √, ∑ authors created digitized text with an English word that describes mathematical symbols within "read_letter" function.

Ex: Values given to the mathematical symbols through "read_letter" function

π = pi ∞ = infinity

√ = square root ∑ = sum

Represent mathematical symbols with English word enhanced the performance of TTS engine.

The accuracy of the equation identification was calculated by providing weighted matrix. Ten images of mathematical equations were executed within test application to calculate the average accuracy.

Weighted Matrix

Expected output = 1

Number of operators and operands in equation = n

Weighted value for one operator or operand ($W_{n)}$ = 1/n

Number of accurate operators and operands in = A_n

Actual output

Accuracy of the digitization process (A_c) = $W_{n\ x}$ An

Maximum accuracy (A_m) = 100 x execution number

Average accuracy (A_a) = [(∑ $A_{c)}$ / A_m] x100

TABLE III. ACCURACY CALCULATION FOR MATHEMATICAL EQUATION IDENTIFICATION FUNCTION

Exe no	Expected digitize text	Actual digitized text	Wn	Accuracy (WnxAn)	Percentage Acx100%
1	2+ 3 = 5	2 + 3 = 5	1/5 = 0.2	0.2x 5 = 1	100%
2	10 + 5 -2 = 13	10 + 5 -2= 13	1/7 = 0.14286	0.14286x7 = 1	100%
3	2 x 10 = 20	2 x 10 = 20	1/5 = 0.2	0.2 x 5 = 1	100%
4	X + 2x + 5y = 53	X + 2x + 5y = 53	1/9	(1/9) x 9 = 1	100%
5	Y = mx + c	Y = mx + c	1/6	(1/ 6) x6 = 1	100 %
6	A = πrr	A = pi r toThePower 2	1/5 = 0.2	0.2 x 5 = 1	100 %
7	3/ 5	3 - 5	1/3	(1/3) x 2 = 0.6667	66.67 %
8	(10 x 5 + 3) /2	(10 x 5 + 3) - 2	1/9	(1/9)x8 = 0.889	88.9%
9	Sin θ	Sin θ	½ = 0.5	0.5x 2 = 1	100%
10	a2 +b2 = c2	A toToThePower 2 + B toToThePower 2 = C toToThePower 2	1/8	(1/8) x 8 = 1	100 %
Average accuracy (Aa) = 955.57 /1000 = 95.56%					

Table identification is a process that uses set of functions to extract characters inside a table and generate meaningful digitized text. 30 input images were tested with the application to calculate the average accuracy of "Table data

identification" function. Table IV contains the results of the function.

TABLE IV. ACCURACY CALCULATION FOR TABLE DATA IDENTIFICATION FUNCTION

Scenario	Number of inputs	Number of correct outputs
2 column tables with single line borders	6	6
2 column tables with double line borders	6	6
2 column single line border tables without 1 header	3	2
2 column tables without borders	5	4
3 column tables with single line borders	6	6
3 column tables with double line borders	4	4
Average accuracy (Aa) = $\frac{Number\ of\ correct\ outputs}{Number\ of\ total\ inputs}$ x 100% = $\frac{28}{30}$ x 100% = 93%		

Calculated accuracy values prove that the functions used to build the "Schmoozer" application have accuracy more than 90%. Therefore, authors assume that the combination of above-mentioned functions provides best autonomous, accurate document reading service to visually impaired individuals.

VI. CONCLUSION

Vision impairment is a huge barrier for a considerable portion of humans all over the world to move with their colleagues, competitors, parents or even with their partners. The situation becomes more sensitive when vision impaired students have to deal with many documents to gain knowledge. Requirements stated by vision impaired individuals for document reading was, a user-friendly mobile application which has the ability to auto capture document image, identify and read aloud documents that contain text, images, equations and table. The research team has developed a mobile application with all the requested facilities to provide user-friendly cost effective accurate document reading solution for them. The major benefit of building a mobile application is that the vision impaired individuals has no need to spend additional cost to by document readers separately or learn Braille technique from the beginning. The solution also has the capability to cut off printing Braille document that contains embossed text.

This paper provides the reasons behind choosing the topic, and how the background survey was done to finalize the solution and methodology for implementation with the testing process. Methodology, Testing and the result presented in this paper are totally based on auto focusing and auto capturing the image, image segmentation with images, mathematical equation, table reading and audio assistance.

Among the group of vision-impaired individuals, students are the most important segment due to they are the future of the nation. The research team identified that reading print disable documents without other's help is the most challenging problem due to documents that could contain text, images, mathematical equations and tables. The majority of existing products are unable to identify images, tables and mathematical equations. To overcome from the situation, research team build a solution that provides auto-capture an image of print disable document, Image processing to convert image content into digitized text and read the content to the visually impaired individuals. With all these key features, schmoozer application is able to provide better service to the visually impaired people to gain the document contain information and fulfill their knowledge.

ACKNOWLEDGMENT

Several people played an important role in accomplishing of this research. First, authors would like to thank the Sri Lanka Institute of Information Technology for the encouragement to pursue this study and the academic staff of Sri Lanka Institute of Information Technology for providing valuable guidance. Authors also wish to thank Ceylon Employees Federation for the support they provide to gather needs of visually impaired people and test the mobile application, without which this research would be incomplete. In addition, authors are grateful for the support, comments and advice received from faculty colleagues of Sri Lanka Institute of Information Technology for their enormous support and guidance towards the success of this product. The support received from all other parties is acknowledged as well.

REFERENCES

[1] R.Maarno, "A library for all – including people with print disabilities," Scandinavian Library Quarterly, 05 May 2017.

[2] Learning Ally, "Learning Ally, Together it is possible," [Online]. Available: https://www.learningally.org

/Educators/school-grants/Massachusetts/Getting-Started. [Accessed 03 May 2017].

[3] AFB Press Customer Service, "George Kerscher: A Pioneer in Digital Talking Books Still Forging Ahead," AccessWorldMagazine, May 2001.

[4] WHO, "Visual impairment and blindness," 2016. [Online]. Available: http://www.who.int / mediacentre /factsheets/fs282/en/. [Accessed 10 Mar 2017].

[5] Census of Population and Housing 2012 – Final Report, "Department of census and statics.," p. pp.117 –120., 2012.

[6] Whatis.techtarget.com, "What is assistive technology (adaptive technology)?," 2011. [Online]. Available: http://whatis.techtarget.com/definition/assistivetechnology- adaptive-technology. [Accessed 13 Sep 2016].

[7] www.afb.org, "Educating students with visual Impairments for inclusion in society - American foundation for the blind," 2005. [Online]. Available: http://www.afb.org/info/programs-and-services/professional-development/teachers/inclusive-education/1235. [Accessed 19 Sep 2016].

[8] www.afb.org, "Assistive technology - American foundation for the blind," 2005. [Online]. Available: http://www.afb.org/info/living-with-vision-loss/using-technology /assistive- technology/123. [Accessed 13 Sep 2017].

[9] www.engineersgarage.com, "Introduction to Image Processing.," 2014. [Online]. Available: http://www.engineersgarage.com/articles/image-processing-tutorial-applications. [Accessed 15 Sep 2016].

[10] WhatIs.com, "What is machine learning?," 2016. [Online]. Available: http:// whatis. techtarget . com/definition/machine-learning. [Accessed 15 Sep 2016].

[11] M.Rouse, "IIS (Internet Information Services)," 2017 Mar 05. [Online]. Available: http://searchwindowsserver.techtarget.com/definition/IIS. [Accessed 02 Apr 2017].

[12] S.Ray, "Understanding Support Vector Machine algorithm from examples (along with code)," 06 Oct 2015. [Online]. Available: https://www.analyticsvidhya.com/blog/2015/10/understaing-support-vector-machine-example-code/. [Accessed 02 Apr 2017].