2019 IEEE International Conference on Signal Processing, Information, Communication & Systems(SPICSCON)
28-30 November,2019, Dhaka, Bangladesh

A Dual-Purpose Refreshable Braille Display Based on Real Time Object Detection and Optical Character Recognition

K M Naimul Hassan*, Subrata Kumar Biswas†, Md Shakil Anwar, Md Shakhrul Iman Siam, Celia Shahnaz ‡

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
*naimul.hassan@ieee.org, †subratakumarbiswas@ug.eee.buet.ac.bd, ‡celia.shahnaz@ieee.org

***Abstract*—This paper proposes a dual-purpose braille system for the visually impaired people. There are two main features of this system- *object detection* and *optical character recognition.* Real time object detection will help a visually impaired person to know about the things around him and optical character recognition will help him reading characters in both international (English) and local community (Bengali) language. In this paper, the detailed methodology of our proposed method is described. A pre-trained convolutional neural network (AlexNet) is used for classifying the objects and an OCR engine (Tesseract) along with basic image processing is used for optical character recognition. A refreshable braille display is also designed to show the braille characters.**

***Keywords*—braille, convolutional neural network, optical character recognition, control, refreshable braille display**

I. INTRODUCTION

Visual impairment and illiterate, or having a learning disability is one of the biggest drawback for humanity. According to WHO (World Health Organization), an estimation of 2.2 billion people live with vision impairment [1]. Although blindness holds a natural attraction for technological originality, innovation of accessibility and educational aid for blind people has been the subject of many investigations. The main purpose of our methodology is to provide a convenient book reader and surroundings detector that is reasonable and readily available to the blind community for making their life easier, self-dependent and moreover converting them into human resources.

Currently available braille books are bulky and very costly. However a very large portion of books are not available as braille books due to the costly braille printing system. To overcome these shortcomings, refreshable braille display has been introduced. All the refreshable braille display currently available work with offline data. For example, textual content in Portable Document Format (PDF), word documents etc. can be read using the display for which data must be stored in the device memory prior to its use. Also adding to that, there is no such availability of refreshable braille display for local community language (Bengali).

Our system consists of two main units- *Control unit* and *Braille Display Unit*. The control unit can be subdivided into two parts- *Image acquisition* and *Processing*. The process of image acquisition is conducted by an external camera mounted on spectacles. This image is processed for either object detection or optical character recognition according to the choice of the user. The braille display part can also be subdivided into two parts- *Transistor control* and *Solenoid driver circuit*. Transistors are controlled through a microprocessor according to the predicted labels or characters. These transistors then meet the power requirements of the solenoids and refresh them after every use.

II. PROPOSED METHODOLOGY

An eye-ware mounted camera takes raw images and sends image data to the processor. After necessary preprocessing, the processor either detects object or read characters from the image according to the user's choice. The

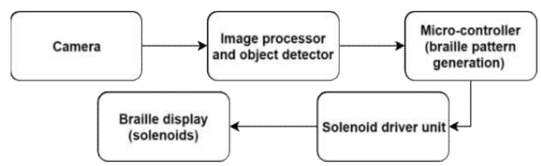

Fig. 1. Block diagram of braille display unit (hardware)

predicted label is sent to the microcontroller character by character. The microcontroller converts the letter into six pin braille output. All six pin braille outputs are sent to the driving circuit through the output pin of the microcontroller. Details of the control unit and braille display unit are described in sub-section A and B. Block diagram and workflow of the proposed system are shown in Fig. 1 and 2 respectively.

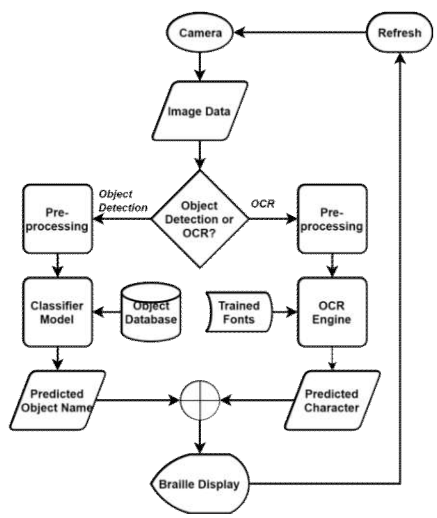

Fig. 2. Workflow of the proposed method

A. Control unit

Control unit is the brain of our proposed method. It works based on two features- *object detection* and *optical character recognition*.

Object Detection:

A handcrafted dataset consisting of 9 objects was trained using a pre-trained convolutional neural network. This trained model was then used to predict objects from real time image data (camera). The dataset was split into train and validation in a ratio of 80:20.

The main preprocessing for the image dataset was to resize each and every image. As the input size of the images for the classifier model is 227x227, every image from the train dataset had to be resized to 227x227.

Pre-trained *AlexNet* [2] was used to train the objects. AlexNet is a convolutional neural network that is trained on more than a million images from the *ImageNet* database [3]. It is consisted of 5 convolutional layers and 3 fully connected layers. In the network, *ReLU* [4] is applied after every convolutional and fully connected layer except the last fully connected layer. *Softmax* [5] is used after the last fully connected layer as activation. After ReLU activation, each convolutional layer is followed by batch normalization and a *max pooling* [6] layer of pool size 3. *Dropout* [7] is applied before the first and second fully connected layer. This pre-trained AlexNet was used to train images of 9 objects (bottle, calculator, cube, flower, gamepad, keyboard, mouse, mug, and stapler*)* for our purpose. The model architecture is shown in Fig. 3.

Spectacles mounted camera was used for capturing real time data. Images were taken frame by frame and this images were the test images. The model gives its predictions on the test images. In this case, prediction label is the name of the predicted objects in the test images.

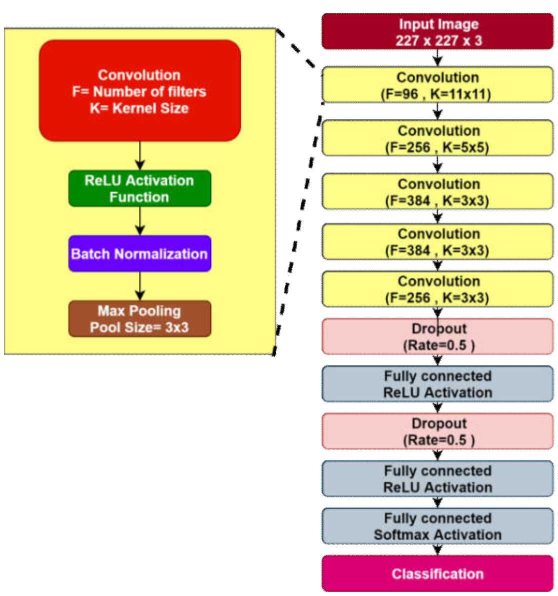

Fig. 3. Model architecture of the classifier model, AlexNet (5 convolutional layers and 3 fully connected layers)

Optical Character Recognition:

Optical character recognition (OCR) was used for reading both international (English) and local community (Bengali) language. *Tesseract-OCR* [8] was used in this regard.

Pepper noise was removed using a median filter after converting the image into black and white. It was fine-tuned via *PSM* (Page Segmentation Mode) parameters [9].

The language setting of the OCR engine was set for both English and Bengali. The fonts of the test images were predicted by matching them with the already trained tesseract fonts. Predictions were done based on the confidence score.

B. Braille display unit

6 solenoids were used for the purpose of RBD. Power requirements of 6 solenoids were met by microcontroller controlled driver circuit. For diving solenoids, transistor were used as switch. We used switching transistor (2N2222) for fast switching of solenoids to form braille letter. Fast switching transistors were being driven by a microcontroller. The micro controller was directly connected to main processor (Broadcom BCM2837).

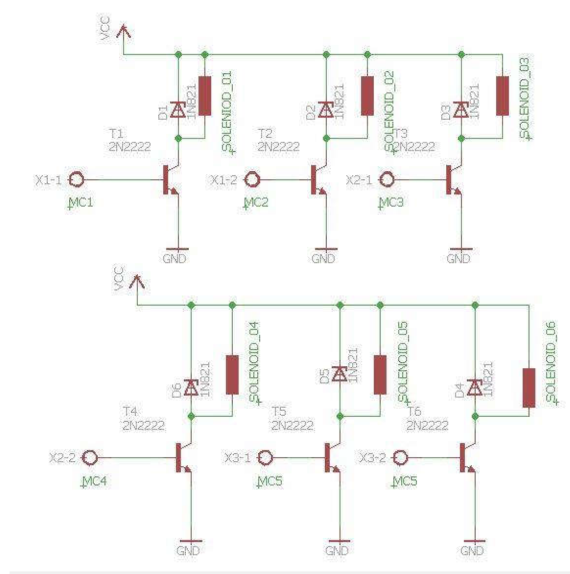

Fig. 4. The driver circuit of the refreshable braille display

The main processor works as image processor and then finally gives signal to microcontroller to operate the solenoids. To capture image, camera module (Sony imx219) is mounted on spectacles of blind people. It has fixed focal length with simultaneous video recording option at 720p, 60FPS.The captured image data is then sent to the main processor wirelessly or by a wired connection. For processor, we had some options such as, Raspberry Pi, Nvidia Jetson Nano etc. We chose Raspberry Pi as main processor because it is cheap, resourceful developer support, available in local market. 11.1V battery was used as primary power source. The battery is consisted of 3 cells each of 3.7volt. For low power microcontroller and driver, a buck converter using LM2596 IC was used. It can handle maximum 40V as input voltage with output current of 3A. This can deliver as low as 1.25V output depending on device requirement. The inner view of the braille display module is shown in Fig. 5.

Fig. 5. Inner view of the braille display module

III. RESULTS

Object Detection:

AlexNet has already learned a rich set of image features but the network was fine-tuned by training it on our dataset so that it could learn features specific to the new dataset. The loss function used was *Categorical crossentropy* [10] and optimizer was *Adam* [11]. Initial learning rate was 0.001. The model was trained using *SGDM* (stochastic gradient descent with momentum)[12]. A summary of training accuracy is shown in Table I.

TABLE I. TRAINING ACCURACY

Epoch	Iteration	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	6.25%	4.8374	0.0010
10	50	100%	0.0006	0.0010
20	100	100%	0.0002	0.0010

Optical Character Recognition:

Tesseract-OCR is trained with fonts of different languages. English and Bengali fonts are also included in it. We used this pre-trained OCR Engine to detect the characters. For the proper detection we needed to do some pre-processing (filtering, black and white conversion, page segmentation mode). An example of the outcome of optical character recognition using Tesseract-OCR is shown using Fig. 6 and Fig.7.

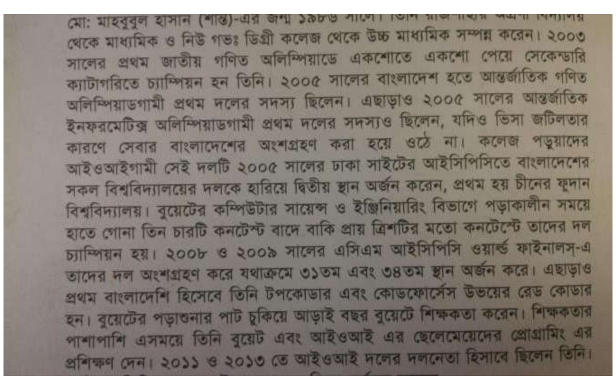
মো: মাহবুবুল হাসান (শান্ত)-এর জন্ম ১৯৮৬ সালে। তিনি রাজশাহীর অগ্রণী বিদ্যালয়
থেকে মাধ্যমিক ও নিউ গভঃ ডিগ্রী কলেজ থেকে উচ্চ মাধ্যমিক সম্পন্ন করেন। ২০০৩
সালের প্রথম জাতীয় গণিত অলিম্পিয়াডে একশোতে একশো পেয়ে সেকেন্ডারি
ক্যাটাগরিতে চ্যাম্পিয়ন হন তিনি। ২০০৫ সালের বাংলাদেশ হতে আন্তর্জাতিক গণিত
অলিম্পিয়াডগামী প্রথম দলের সদস্য ছিলেন। এছাড়াও ২০০৫ সালের আন্তর্জাতিক
ইনফরমেটিক্স অলিম্পিয়াডগামী প্রথম দলের সদস্যও ছিলেন, যদিও ভিসা জটিলতার
কারণে সেবার বাংলাদেশের অংশগ্রহণ করা হয়ে ওঠে না। কলেজ পড়ুয়াদের
আইওআইগামী সেই দলটি ২০০৫ সালের ঢাকা সাইটের আইসিপিসিতে বাংলাদেশের
সকল বিশ্ববিদ্যালয়ের দলকে হারিয়ে দ্বিতীয় স্থান অর্জন করেন, প্রথম হয় চীনের ফুদান
বিশ্ববিদ্যালয়। বুয়েটের কম্পিউটার সায়েন্স ও ইঞ্জিনিয়ারিং বিভাগে পড়াকালীন সময়ে
হাতে গোনা তিন চারটি কনটেস্ট বাদে বাকি প্রায় ত্রিশটির মতো কনটেস্টে তাদের দল
চ্যাম্পিয়ন হয়। ২০০৮ ও ২০০৯ সালের এসিএম আইসিপিসি ওয়ার্ল্ড ফাইনালস্-এ
তাদের দল অংশগ্রহণ করে যথাক্রমে ৩১তম এবং ৩৪তম স্থান অর্জন করে। এছাড়াও
প্রথম বাংলাদেশি হিসেবে তিনি টপকোডার এবং কোডফোর্সেস উভয়ের রেড কোডার
হন। বুয়েটের পড়াশুনার পাট চুকিয়ে আড়াই বছর বুয়েটে শিক্ষকতা করেন। শিক্ষকতার
পাশাপাশি এসময়ে তিনি বুয়েট এবং আইওআই এর ছেলেমেয়েদের প্রোগ্রামিং এর
প্রশিক্ষণ দেন। ২০১১ ও ২০১৩ তে আইওআই দলের দলনেতা হিসাবে ছিলেন তিনি।

Fig. 6. An input image of a page from a Bengali language book for OCR

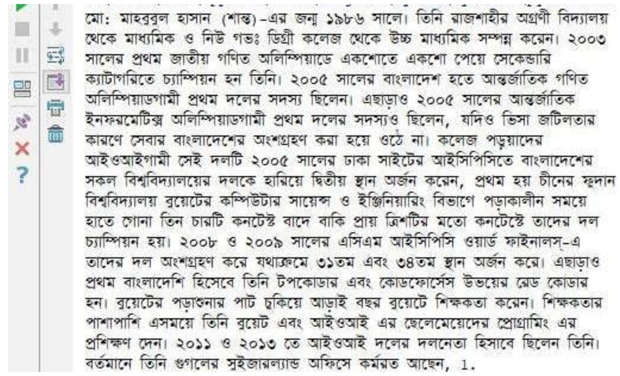
মো: মাহবুবুল হাসান (শান্ত)-এর জন্ম ১৯৮৬ সালে। তিনি রাজশাহীর অগ্রণী বিদ্যালয়
থেকে মাধ্যমিক ও নিউ গভঃ ডিগ্রী কলেজ থেকে উচ্চ মাধ্যমিক সম্পন্ন করেন। ২০০৩
সালের প্রথম জাতীয় গণিত অলিম্পিয়াডে একশোতে একশো পেয়ে সেকেন্ডারি
ক্যাটাগরিতে চ্যাম্পিয়ন হন তিনি। ২০০৫ সালের বাংলাদেশ হতে আন্তর্জাতিক গণিত
অলিম্পিয়াডগামী প্রথম দলের সদস্য ছিলেন। এছাড়াও ২০০৫ সালের আন্তর্জাতিক
ইনফরমেটিক্স অলিম্পিয়াডগামী প্রথম দলের সদস্যও ছিলেন, যদিও ভিসা জটিলতার
কারণে সেবার বাংলাদেশের অংশগ্রহণ করা হয়ে ওঠে না। কলেজ পড়ুয়াদের
আইওআইগামী সেই দলটি ২০০৫ সালের ঢাকা সাইটের আইসিপিসিতে বাংলাদেশের
সকল বিশ্ববিদ্যালয়ের দলকে হারিয়ে দ্বিতীয় স্থান অর্জন করেন, প্রথম হয় চীনের ফুদান
বিশ্ববিদ্যালয় বুয়েটের কম্পিউটার সায়েন্স ও ইঞ্জিনিয়ারিং বিভাগে পড়াকালীন সময়ে
হাতে গোনা তিন চারটি কনটেস্ট বাদে বাকি প্রায় ত্রিশটির মতো কনটেস্টে তাদের দল
চ্যাম্পিয়ন হয়। ২০০৮ ও ২০০৯ সালের এসিএম আইসিপিসি ওয়ার্ড ফাইনালস্-এ
তাদের দল অংশগ্রহণ করে যথাক্রমে ৩১তম এবং ৩৪তম স্থান অর্জন করে। এছাড়াও
প্রথম বাংলাদেশি হিসেবে তিনি টপকোডার এবং কোডফোর্সেস উভয়ের রেড কোডার
হন। বুয়েটের পড়াশুনার পাট চুকিয়ে আড়াই বছর বুয়েটে শিক্ষকতা করেন। শিক্ষকতার
পাশাপাশি এসময়ে তিনি বুয়েট এবং আইওআই এর ছেলেমেয়েদের প্রোগ্রামিং এর
প্রশিক্ষণ দেন। ২০১১ ও ২০১৩ তে আইওআই দলের দলনেতা হিসাবে ছিলেন তিনি।
বর্তমানে তিনি গুগলের সুইজারল্যান্ড অফিসে কর্মরত আছেন, 1.

Fig. 7. Detected characters by OCR Engine in raspberry pi (python) command window

Refreshable Braille Display:

All the components were assembled properly. The final refreshable braille display is shown in Fig. 8. The components were assembled in a square shaped box. The top of the box has six holes and the solenoid pins are placed under them. The system can be opened from one side for maintenance (which is also shown in Fig. 5).

Fig. 8. Proposed refreshable braille display

IV. DISCUSSION

The product is economical compared to the existing technologies we have. It is also capable of serving two different purposes.

TABLE II. COST ESTIMATION CHART

Category	Cost (USD)
Raspberry Pi Camera	100
Control Unit	40
Braille Display Unit	58
Total	198

Table II shows the estimated total component cost of the product. This cost includes the costs of a full set control unit, an eye ware mounted camera and a refreshable braille display. The total cost would be 248$ (198+50) including fixed and variable manufacturing cost. Commercial price of refrashable braille display varies from 500$ to 3500$. But this does not include any camera.

It provides a visually impaired person to feel his/her surroundings just by looking around through camera just like a normal human being which might help him/her to improve the way of living to some extent. As braille printed books are rare and hardly enough to meet the demand, our proposed system can be counted as the replacement of the bulky and costly braille books. One can look into the book through camera and the read the detected characters which will be shown in RBD one after another. Its ability to read books of local language proves its versatility which can serve the local community. The small stature of the device and the standalone nature provide its mobility and it is useable at anywhere.

V. CONCLUSION

The future target is to convert this dual-purpose braille display into a multi-purpose braille display which will help the visual impaired people lead a smart life. To fulfill the goal, it is needed to add some features. Device tracking system should be added for the device security. Over the Air (OTA) firmware update feature should be added so that we can update the trained model and have the chance to make it robust.

This device is an initiative to improve the conventional way of education for visually impaired people. The main advantage of the proposed method over other existing systems is that it is mobile, low-cost and easy to operate. People of many third world or underdeveloped countries where cost is one of their main concerns will be vastly benefited by this device. Reading native language is a unique feature which gives this device an edge over other conventional devices. The object detection module works as a third eye for them. Our aim is to build a better and comfortable world for the specially abled people.

VI. ACKNOWLEDGMENT

The authors of this paper would like to acknowledge the kind support from the supervisors and assistants of both *Communication Systems Laboratory* and *Microprocessor & Interfacing Laboratory* in the department of Electrical and Electronic Engineering of Bangladesh University of Engineering and Technology.

REFERENCES

[1] World Health Organization, "Blindness and vision impairment". In: (2019)

[2] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Networks". In: *volume-1; pages- 1097-1105; NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems.*

[3] Russakovsky, O., Deng, J., Su, H. et al. "ImagNet: A large Scale Hierarchical Image Database". In: *CVPR 2009.*

[4] Abien Fred M. Agarap, "Deep Learning using Rectified Linear Units (ReLU)", *arXiv:1803.08375 [cs.NE].*

[5] Chigozie Nwankpa, Winifred Ijomah, Anthony Gachagan, Stephen Marshall, "Activation Functions: Comparison of trends in Practice and Research for Deep Learning", *arXiv:1811.03378 [cs.LG]*

[6] Haibing Wu, Xiaodong Gu, "Max-Pooling Dropout for Regularization of Convolutional Neural Networks", *arXiv:1512.01400 [cs.LG]*

[7] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, Ruslan R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors", *arXiv:1207.0580 [cs,NE]*

[8] Google. *Google Tesseract OCR: An open source optical character recognition engine able to recognize more than 100 languages out of the box. 2018*. URL: https: //opensource.google.com/projects/tesseract

[9] Github. *Improving the Quality of the Output, Tesseract-OCR, Tesseract,* URL: https://github.com/tesseract-ocr/tesseract/wiki/ImproveQuality

[10] Zhilu Zhang, Mert R. Sabuncu, "Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels", *arXiv:1805.07836 [cs.LG]*

[11] Diederik P. Kingma, Jimmy Ba, "Adam: A Method for Stochastic Optimization", *arXiv:1412.6980 [cs.LG]*

[12] Sebastian Ruder, "An overview of gradient descent optimization algorithms", *arXiv:1609.04747 [cs.LG]*