Real time Signboard Reader to assist visually impaired

Prof.(Dr.) Shripad Bhatlawande, Sarvesh Patki, Niraj Wani, Aditya Wyawahare, Siddhesh Uppu
Electronics & Telecommunication Engineering, Vishwakarma Institute of Technology, Pune, India

***Abstract* - In this paper, a camera based real time signboard reading system has been proposed, in order to assist the visually impaired people. The system is a portable backpack, which consists of a monocular camera, mounted on a spectacle and earphones for providing audio feedback. The aim of the project is to develop a prototype which will aid the visually impaired to read signs and help them in their day-to-day life.**

Keywords – Signboard Reader, Raspberry Pi, Python, OpenCV, Oriented FAST and Rotated BRIEF, Spectacles.

I. INTRODUCTION

Visually impaired people face many challenges in their daily routine. Most of these problems can be solved by utilizing modern technologies being produced and researched. Many social awareness programs are conducted worldwide associating the social work with science that can help developing the society. Technologies like – assistive cane, vibration stick, smartphone based navigation system, etc. are being researched and manufactured to help blind or visually impaired. Out of the 314 million visually impaired people worldwide, 45 million people are blind. Even in the developed countries these numbers are increasing rapidly. Recent developments in computer vision, digital cameras, and portable computers make it feasible to assist these individuals by developing camera based products that combine such optical character recognition systems. The proposed project in this paper introduces an assistive system which will aid the visually impaired and blind by identifying the signboards they come across in their daily life. This will enable the blind to travel without any human assistance.

II. LITERATURE SURVEY

The project aims in assisting the visually impaired in their daily routine. The study included following at the initial stage:

I. Literature Review
II. Interviews of blind/visually impaired

The objective of the interviews was to understand the needs of the visually impaired. This was done by forming a set of questions relevant to the topic which would help in understanding their needs and opinions regarding the proposed system. The data was collected by visiting the blind people and taking their responses. The literature review stage included a detailed review and study of journal and conference papers based on signboard reading and text recognition. This helps in understanding the techniques used in different pre-implemented system to meet the needs of visually impaired and the limitations in the recent technologies.

III. FRAMEWORK

Based on the collected data, needs and limitations, a system was proposed which is presented in this paper. This system includes a spectacle with a monocular camera (Raspberry Pi V2.1) to capture the signboards. The features from the captured set of real time images will be detected, matched and recognized using OpenCV in Python using a microcontroller – Raspberry Pi 3B+. The text output will be converted to audio/speech in Marathi language and will be transmitted to the earpiece worn by the visually impaired.

A monocular camera is mounted on the goggles of the user. This camera is connected to the processor which is fed with the required algorithm and kept in a backpack..This processor works on the real life frames which the camera captures. These frames are worked upon with the algorithm required to detect features from the images. After the process of feature matching, the obtained text representing the signboard is then converted into audio format using a piece of code. This audio signal generated in the processor is then transferred to a headset/earphone which is connected to the processor. Audio output will be in Marathi. This prototype reads the text to the user, helping the user to identify the signboards. The system is powered by a power bank.

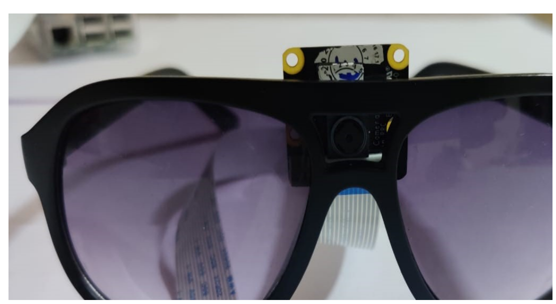

Figure 1 : Spectacles with mounted camera

System Prototype

Figure 2 : System prototype – Backpack, spectacles, camera, power bank, RPi, Earphones

A. *Feature extraction :*

The technique used for extracting features from the images in the first stage is Oriented FAST and Rotated BRIEF(ORB).ORB is basically a fusion of FAST keypoint detector and BRIEF descriptor with many modifications to enhance the performance. First it uses FAST to find keypoints, then apply Harris corner measure to find top N points among them. It also uses pyramid to produce multiscale-features.
It computes the intensity weighted centroid of the patch with located corner at center. The direction of the vector from this corner point to centroid gives the orientation. To improve the rotation invariance, moments are computed with x and y which should be in a circular region of radius r, where r is the size of the patch.
Now for descriptors, ORB use BRIEF descriptors. But BRIEF performs poorly with rotation. So what ORB does is to steer BRIEF according to the orientation of key points. For descriptor matching, multi-probe Locality-Sensitive Hashing which improves on the traditional LSH, is used.

The algorithm is written in Python programming language. Python provides the ease of programming in simpler way (syntactically) with letting many options for the manipulating the image in many different ways suitable to this project. The major library used for the purpose is Open Computer Vision i.e. OpenCV. OpenCV is an open source module which allows various built-in functions to detect the exact portion of an image (a video frame in this case) for extraction of text of a signboard and read it using the different machine learning algorithms.

B. *Real time feature matching :*

ORB(Oriented FAST and Rotated Brief) algorithm has been used. FAST is used for detecting the key points and BRIEF for the descriptors. The descriptors for all the templates are stored. The number of features taken into consideration for each template is 500 while 2500 features have been considered for the real-time images. Template matching is done in real time at the rate of 5 FPS. The image is resized to 800x600 for faster results. On extracting a frame, we calculate descriptors of the current frame and compare them with the descriptors of all templates. The recognized sign is the one which has maximum number of matching descriptors. Two threshold values have been set for comparison of the real-time image with each template. Finally, output is converted into more user-friendly audio format.

C. *Text to Speech :*

The signboard is read to the user after converting the predefined text associated with the signboard into audio using the aplay() command. aplay() is a command-line sound player for Advanced Linux Sound Architecture(ALSA) sound card drivers. ALSA is the lowest level of the Linux sound stack.
The audio output is in Marathi language.

D. *Dataset :*

Dataset consists of 11 signboards which are most commonly encountered in our daily lives. Most of the signs are traffic signs while other emergency warning signs are also considered. The images were converted to grayscale images of 100*100 pixels. The features of all images were extracted using ORB.
The prepared dataset is a robust dataset and can be modified as per specific requirements of application. The 11 signs which have been considered include:

1. Stop
2. Fire exit
3. Pedestrian crossing
4. Speed breaker
5. Washroom
6. Hospital
7. First Aid
8. Speed limit
9. Construction ahead
10. Right turn
11. Left turn

Figure 3 : Stop sign, Zebra Crossing, Fire Exit signboards

Figure 4 : Grayscale images of signboards

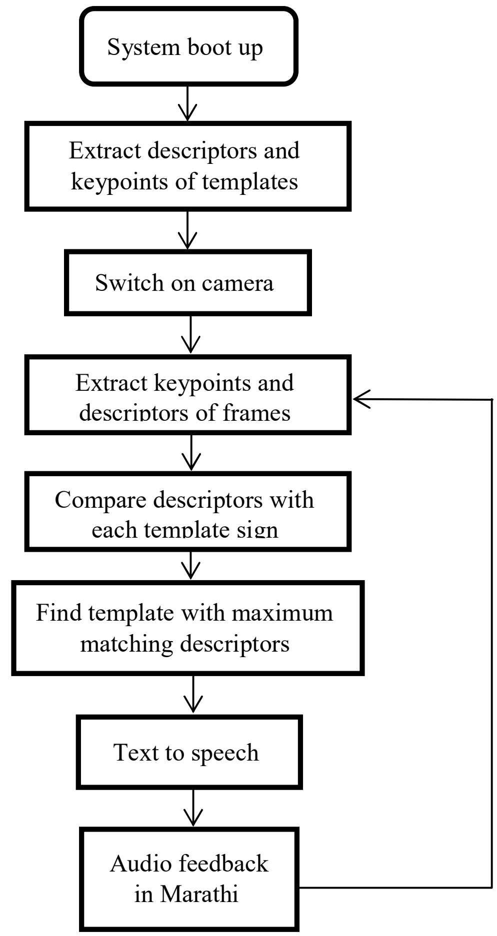

Figure 5: Flowchart

E. Microcontroller :

Microcontroller used in the system is Raspberry Pi 3B+. Raspberry Pi 3 Model B+ has a fast 1.4 GHz processor and a three-times faster gigabit Ethernet (throughput limited to ca. 300 Mbps by the internal USB 2.0 connection) or 2.4 / 5 GHz dual-band 802.11ac Wi-Fi (100 Mbps). Other features are Power over Ethernet (PoE), USB boot and network boot (an SD card is no longer required).

The camera and earphones are connected to the microcontroller and the entire system is powered using a power bank.

IV. RESULTS

The signboard images from the dataset were recognized using the system introduced in the paper. Whenever a signboard or the image of signboard captured by the camera is successfully matched with any of the templates, the user is provided with an audio output in Marathi, mentioning the name of signboard. The set of images were first tested on webcam of a computer. The features were matched for all signboards, given that the threshold values were set. The audio output was obtained.

The same process was carried out using Raspberry Pi 3B+ microcontroller. All predefined signboards were successfully matched and recognized on a real time basis. An audio feedback in Marathi was obtained and could be listened through earphones connected with RPi.

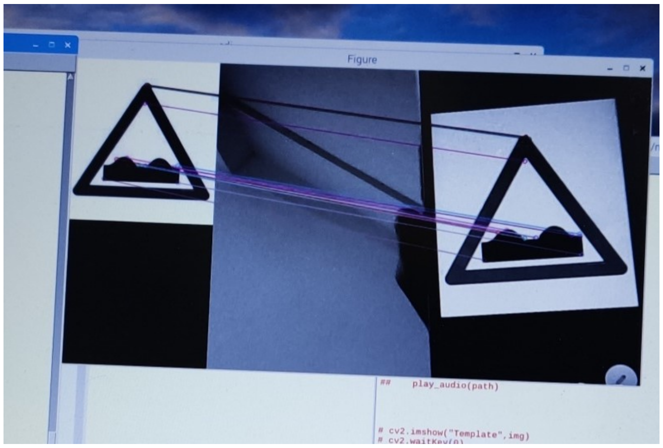

Figure 6 : Feature matching for a Speedbreaker sign

V. CONCLUSION

In this paper, we have presented an approach for automatic recognition of signboards. The proposed approach can efficiently read signboards in real time and provides audio feedback, thereby assisting the visually impaired in their daily lives. The system proposed in this paper is based on ORB (Oriented FAST and Rotated BRIEF), which is an efficient and accurate feature extraction technique. The dataset consists of traffic signs and other emergency warning signs which are encountered most commonly in our day-to-day life. The dataset can be easily expanded / modified as per problem specific applications. The audio output for our system is in Marathi language but output can be provided in other regional languages as well. The system is cost effective and portable. The feature extraction might get complicated for complicated signboards like shop signs, etc. Some modifications in the algorithm can solve this problem. The range of detection covered by the camera is low but it can be increased by using a better camera. This system will enable the visually impaired or blind people to travel without any human assistance and independently identify the signboards, making their routine easier.

Acknowledgement

This work was supported and guided by Prof. (Dr.) Shripad Bhatlawande, Head of Department, Department of Electronics and Telecommunication Engineering, Vishwakarma Institute of Technology, Pune.

References

[1] X. Chen, J. Yang, J. Zhang, and A. Waibel, "Automatic Detection and Recognition of Signs From Natural Scenes" IEEE Transactions on Image Processing, vol. 13, no. 1, pp. 87–99, 2004.

[2] H. Li, D. Doermann, and O. Kia, "Automatic Text Detection and Tracking in Digital Video", Jan. 1998.

[3] X. Chen and A. Yuille, "Detecting and reading text in natural scenes", Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.

[4] C. Yao, X. Bai, W. Liu, Y. Ma, and Z. Tu, "Detecting texts of arbitrary orientations in natural images", 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012.

[5] U. Bhattacharya, S. K. Parui, and S. Mondal, "Devanagari and Bangla Text Extraction from Natural Scene Images", 2009 10th International Conference on Document Analysis and Recognition, 2009.

[6] Z. Ni, S. Fu, B. Tang, H. He, and X. Huang, "Experimental studies on indoor sign recognition and classification," 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), 2014.

[7] Z. Liu, Y. Luo, J. Cordero, N. Zhao, and Y. Shen, "Finger-eye: A wearable text reading assistive system for the blind and visually impaired," 2016 IEEE International Conference on Real-time Computing and Robotics (RCAR), 2016.

[8] S. K. Nagavi, M. S. Gothe, and P. S. Totiger, "Optical Character Recognition based Auto Navigation of Robot by Reading Signboard," International Journal of Engineering Research and, vol. V4, no. 07, Aug. 2015.

[9] A. Ganz, J. M. Schafer, Y. Tao, C. Wilson, and M. Robertson, "PERCEPT-II: Smartphone based indoor navigation system for the blind," 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2014.

[10] Jun Ohya, Akio Shio, and Shigeru Akamatsu ,"Recognizing Characters in Scene Images" IEEE Transactions On Pattern Analysis And Machine Intelligence.Vol.16, No. 2, February 1994.

[11] TrungQuy Phan, Palaiahnakote Shivakumara, Shangxuan Tian and Chew Lim Tan, "Recognizing text with perspective distortion in natural scenes"2013 IEEE International Conference on Computer Vision.1550-5499/13 $31.00 © 2013 IEEE DOI 10.1109/ICCV.2013.76

[12] Naveen Sankaran,C.VJawahar,International Institute of Information Technology - Hyderabad, India. "Recognition of Printed Devanagari Text Using BLSTM Neural Network",Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)

[13] Alessandro Bissacco, Mark Cummins, Yuval Netzer, Hartmut Neven Google Inc.,"Photo OCR: Reading Text in Uncontrolled Conditions",2013 IEEE International Conference on Computer Vision,1550-5499/13 $31.00 © 2013 IEEE DOI 10.1109/ICCV.2013.102

[14] Chucai Yi, Student Member, IEEE, Yingli Tian, Senior Member, IEEE, and Aries Arditi, “Portable Camera-Based Assistive Text and Product Label Reading From Hand-Held Objects for Blind Persons”,IEEE/ASME Transactions On Mechatronics, Vol. 19, No. 3, June 2014

[15] R. Panahi and I. Gholampour, “Accurate Detection and Recognition of Dirty Vehicle Plate Numbers for High-Speed Applications”, IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 4, pp. 767–779, 2017.

[16] R. Wijnhoven and P. D. With, “Identity verification using computer vision for automatic garage door opening,” IEEE Transactions on Consumer Electronics, vol. 57, no. 2, pp. 906–914, 2011.

[17] Z.-X. Chen, C.-Y. Liu, F.-L. Chang, and G.-Y. Wang, “Automatic License-Plate Location and Recognition Based on Feature Salience,” IEEE Transactions on Vehicular Technology, vol. 58, no. 7, pp. 3781–3785, 2009.

[18] Oliveira-Neto, F., Han, L. and Jeong, M. (2013). An Online Self-Learning Algorithm for License Plate Matching. IEEE Transactions on Intelligent Transportation Systems, 14(4), pp.1806-1816

[19] A. Ashtari, M. Nordin and M. Fathy, "An Iranian License Plate Recognition System Based on Color Features", IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 4, pp.1690-1705, 2014

[20] M. Mathias, R. Timofte, R. Benenson, and L. V. Gool,“Traffic sign recognition — How far are we from the solution?,” The 2013 International Joint Conference on Neural Networks (IJCNN), 2013.

[21] A. Mishra, K. Alahari, and C. V. Jawahar, “Top-down and bottom-up cues for scene text recognition,” 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012.

[22] N. Ezaki, M. Bulacu, and L. Schomaker, “Text detection from natural scene images: towards a system for visually impaired persons,” Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., 2004.

[23] P.Comelli,P.Ferragina,M.Granieri,andF.Stabile,“Opticalrecognitionofmotorvehiclelicenseplates,”IEEETransactionsonVehicularTechnology,vol.44,no.4,pp.790–799,1995.

[24] T. Naito, T. Tsukada, K. Yamada, K. Kozuka, and S. Yamamoto, “Robust license-plate recognition method for passing vehicles under outside environment,” IEEE Transactions on Vehicular Technology, vol. 49, no. 6, pp. 2309–2319, 2000.

[25] C. Shi, C. Wang, B. Xiao, Y. Zhang, S. Gao, and Z. Zhang,“Scene Text Recognition Using Part-Based Tree-Structured Character Detection,” 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013.