

# VIAssist: Assistance for the Visually Impaired

Aparajita Marathe  
St.Francis Institute of  
Technology,  
Mt.Poinsur, Borivali(West),  
Mumbai

Denson George  
St.Francis Institute of  
Technology,  
Mt.Poinsur, Borivali(West),  
Mumbai

Rishi Malve  
St.Francis Institute of  
Technology,  
Mt.Poinsur, Borivali(West),  
Mumbai

## ABSTRACT

There has been a vast advancement in Computer Vision since its inception and it has been used for the ease and betterment of the visually impaired. In this paper, we propose a bus number recognition system which will assist the visually impaired to board the correct bus. The proposed work will be deployed on bus stops. It will consist of a camera, a processing unit and speakers. Once a bus is detected, the visually impaired will be notified that a bus is approaching through constant beeps. The video feed will be further processed and ultimately the bus number and the destination of the bus will be announced through the speakers. The buses in Maharashtra have bus numbers displayed in Devanagari script. Hence we store templates corresponding to the Devanagari bus numbers in the database. Various cities use different scripts, thus by changing the database of templates this system can adapt accordingly.

## General Terms

Computer Vision, Bus Number Recognition System (BNRS)

## Keywords

VIAssist, Speeded Up Robust Features (SURF) Algorithm, Template Matching.

## 1. INTRODUCTION

India has a total of 8 million blind people approximately <sup>[1]</sup> and one out of every five blind people in the world lives in India. When it comes to metropolitan cities like Mumbai, commuting is one of the biggest concerns that visually impaired individuals face. Many visually impaired people heavily rely on public transport facilities like the local railway and bus services. The local railways have periodic announcements and beepers to inform the blind individuals about the incoming train and the location of the compartment respectively. However when it comes to the Brihanmumbai Electricity Supply and Transport (BEST) system, there are very few facilities to help blind people commute with ease. A visually impaired person has to rely on fellow normal sighted persons for information about bus arriving and its destination. The proposed system “VIAssist” is a standalone system that will provide bus numbers and destination of BEST buses in the form of audio output. This will help the visually impaired users become aware about the buses that arrive at the bus stop. The system will be installed at a strategic position on the bus stop. This position of the camera is crucial in order to acquire a proper image of the bus. This method employs a combination of a robust bus detection module and a light weight template matching module to give satisfactory recognition results.

## 2. LITERATURE SURVEY

Computer vision is one of the most important and useful technologies for the visually impaired (blind and low-sighted)

people both in outdoor and indoor scenarios. There have been several works in this area. The following are the different approaches that have been used to detect a bus and recognize its bus number.

G.Lavanya et Al.<sup>[2]</sup>, says their methodology is based on helping the blind people board the desired bus with the help of ZigBee units. ZigBee is used to create personal area networks built from small, low-power digital radios. ZigBee units are present in the bus as well as with the blind user. Primarily, the blind person in the bus station is identified with RF(Radio Frequency) communication. The blind informs the location he needs through the microphone. An output of bus numbers which can take the blind user to the destination is given as the output. The ZigBee transceiver located in the bus sends the bus numbers in the form of signals to the transceiver located with the blind person. Finally the bus number of the bus situated in the bus stand is announced to the blind through the headphones. This would aid the blind person to choose the right bus in which he/she has to travel. Inside the bus, when the destination is reached, GPS-634R along with the controller and voice synthesizer is used to produce an audio output. The work done in this paper involves overhead of several additional devices (ZigBee unit, microphone, headphones) to be carried by the blind user. The cost of installation and maintenance of ZigBee units and GPS in all buses is also quite high.

Claudio Guida et Al.<sup>[3]</sup>, used machine learning and geometric computer vision for automatically reading bus numbers with a smart phone. This method allowed a visually impaired person to stand at the bus stop with his mobile phone in hand to acquire the bus number of the oncoming bus. As the sound of bus is heard by him in close proximity, the user directs the phone camera towards it. This starts the process of data acquisition from the mobile camera. To attain better results several consecutive live images of the acquired sequence are processed separately and the most frequent result is returned to the user via voice synthesis. First of all, the oncoming bus is localized inside the image and only the image containing the region of interest (bus ROI) is considered for further processing. This is the Localization module and it uses trained classifiers. Once the bus is localized, a further image cropping is performed in order to isolate the number region. This is done by a template description of the bus facade which is matched with the image content of the Bus ROI. The last stage of their method included reading the number of the bus using binarization. Experimental results have shown that bus detection rate was quite high in most of the sequences. Detection rates showed that with the exception of the number 8, every bus number is correctly recognized with 100% rate. Although this method is very accurate the problem of correct image acquisition by the blind user arises. Also the memory and computational power available in smart phones is quite

limited. This can be a setback taking into consideration such a heavy application.

Hangrong Pan et Al. [4], have designed a computer vision-based system to detect and recognize bus information from images captured by a camera at a bus stop. In bus detection, Histogram of the Oriented Gradient (HOG) descriptor was employed to extract the image based features of bus facade. Cascade SVM model was applied to train a bus classifier to detect the existence of bus facade. In bus route number recognition, they designed a text detection algorithm on the basis of layout analysis and text feature learning. This was used to recognize the text codes from detected text regions for audio notification. Two datasets were involved in their experiments, the bus dataset and text dataset respectively. The bus dataset was self-collected. This dataset was used for training bus classifier and evaluating the accuracy of bus facade detection. It consisted of camera-based natural scene images, captured in bus stations by a Samsung Galaxy III cell phone. The overall performance is calculated to be 80.93%.

The above designed system is for sliding window with digital display. However, if such a system has to be designed for Indian roads, for example the BEST bus route, a different approach has to be adopted. The BEST buses have fixed number plates with predefined font. For this particular application template matching technique is adequate to detect and identify the bus number of the approaching bus.

### 3. PROPOSED WORK

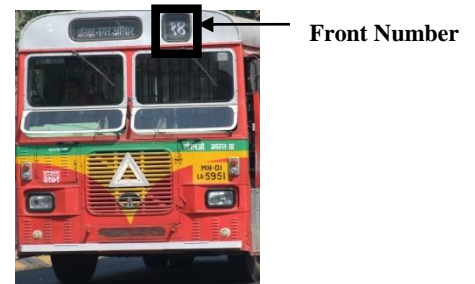
The proposed system consists of a camera placed on the bus stand. This camera takes as input a continuous video feed and supplies the same to the system. The system extracts the frame which contains the BEST bus. This frame acquired is then sent for noise removal and further processing. The processed frame is then matched with the templates in the database and the corresponding audio output is given through the speakers.

A camera is placed in the top corner of the bus stop to capture the front façade of the approaching bus as shown in Fig.3.1. For the scenario when two buses come one after the other with very less distance between them and where the façade of the second bus is hidden due to the first bus, the following solution is considered:

An additional camera is installed in the back to capture the number displayed in the rear end of the bus as shown in Fig.3.2. The rear camera is triggered when the front camera detects a BEST bus in the video. Template matching takes place for both frontal view and side rear view of the bus. The results of front camera are temporarily stored. The rear camera keeps capturing frames and processing them. Once the first bus leaves and the second bus arrives, the rear camera finds the right template match. The system compares the results of the front and rear. AND logic is applied as given in table 3.1. and respective outcome is announced through the speakers.

**Table 1**

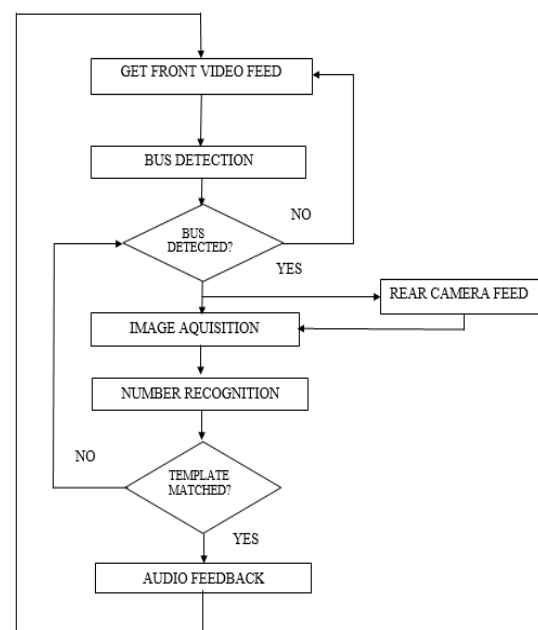
FRONT CAMERA	REAR CAMERA	OUTPUT
Bus 1	Bus 1	Bus 1
Bus 1	Bus 2	Bus 1 and Bus 2
Bus 2	Bus 1	Bus 2 and Bus1
Bus 2	Bus 2	Bus 2



**Figure 1 Front View of BEST bus**



**Figure 2 Side View of BEST bus. Rear Number**



**Figure 3 Working of VIAssist**

Figure 3. illustrates the working of VIAssist and the flow of information in the system.

## a. Front and Rear Camera Video Feed

The front and rear camera video feed is captured continuously from the strategically positioned cameras on the bus stand. The front camera video stream is given as input to the Bus Detection module and the rear camera video stream fed to the template matching module.

## b. Bus Detection

This module uses the Speeded Up Robust Features (SURF)<sup>[5][6]</sup> detector algorithm (proposed in 2006) for detection of the bus. SURF algorithm includes two main processes which are as follows:

### 1. Interest Point Detection

#### a) Use of Fast Hessian Detector<sup>[6]</sup> for blob detection:

Hessian matrix is used because of its good performance in accuracy. More precisely, we detect blob-like structures at locations where the determinant of the matrix is maximum. Given a point  $x = (x, y)$  in an image  $I$ , the Hessian matrix  $H(x, \sigma)$  in  $x$  at scale  $\sigma$  is defined as follows:

$$H(x, \sigma) = \begin{matrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{matrix} \quad \text{.. (Eq. 1)}$$

Where,

- $L_{xx}(x, \sigma)$  is the convolution of the Gaussian second order derivative  $\frac{\delta^2}{\delta x^2} g(\sigma)$  with the image  $I$  in point  $x$ .
- Similarly,  $L_{xy}(x, \sigma)$  and  $L_{yy}(x, \sigma)$  are Gaussian second order derivative with the Image  $I$  in point  $x$ .

To further increase the accuracy of Hessian matrix it is approximated using box filters. The approximated equation is

$$\det(H_{approx}) = D_{xx}D_{yy} - (\omega D_{xy})^2 \quad \text{.. (Eq. 2)}$$

Where,

- $D_{xx}$  is the approximation of  $L_{xx}(x, \sigma)$ .
- $D_{xy}$  is the approximation of  $L_{xy}(x, \sigma)$ .
- $D_{yy}$  is the approximation of  $L_{yy}(x, \sigma)$ .
- The approximated determinant of the Hessian  $\det(H_{approx})$  represents the blob response in the image at location  $x$ .

#### b) Use of Integral Images

In order to make object more self-contained integral images are used. Integral images allow the use of box type convolution filters for fast computation.

$$s(x, y) = i(x, y) + s(x - 1, y) + s(x, y - 1) - s(x - 1, y - 1) \quad \text{.. (Eq. 3)}$$

Where,

$s(x, y)$  is value at any point  $(x, y)$  in the summed area table.

## 2. Detection of feature vector descriptor based on the neighborhood

The purpose of a descriptor is to provide a unique and robust description of a feature, a descriptor can be generated based on the area surrounding the interest point. The region is split up regularly into smaller  $4 \times 4$  square sub-regions. This keeps important spatial information in. For each sub-region, we compute a few simple features at  $5 \times 5$  regularly spaced sample points. A typical feature vector is given by

$$V = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|) \quad \text{.. (Eq. 4)}$$

Where,

- $d_x$  is haar wavelet responses in  $x$  direction.
- $d_y$  is haar wavelet responses in  $y$  direction.
- $|d_x|$  is modulus of haar wavelet responses in  $x$  direction.
- $|d_y|$  is modulus of haar wavelet responses in  $y$  direction.

## c. Number Recognition

Number plate recognition module also uses template matching technique<sup>[7]</sup> using parameter based correlation. Once the bus number is localized, recognition of the bus number is achieved using template matching by following equation:

$$R(x', y') = \frac{\sum_{\{x' y'\}} [T'(x', y') I(x + x', y + y')]}{\sqrt{\{\sum_{\{x' y'\}} T(x', y')^2 \sum_{\{x' y'\}} I(x + x', y + y')^2\}}} \quad \text{.. (Eq. 5)}$$

Where,

- $T$  is a template image.
- $I$  is an original image.
- $x$  and  $y$  are the pixel coordinates.
- $R$  is a correlation coefficient.

## d. Audio Feedback

The corresponding bus number and destination are retrieved from the database and announced using Text-To Speech.

## 4. IMPLEMENTATION

The implemented simulation is a software prototype developed in JAVA programming language using JavaCV library. The prototype has provisions for two video inputs (front and rear). VIAssist is bus stop specific. The templates of the buses that come to that particular stop are stored in the database. This helps in reducing the computation time and also less storage space is required.

The working of the simulation is as follows:

The front and rear video are given as inputs. The front video is processed and frames are generated. The SURF algorithm is applied on every frame for bus detection. A beeping sound is triggered when the first frame containing the bus is detected. This frame is processed for number recognition using correlation based template matching. If a suitable result is not found in this frame all the consequent frames are also processed using template matching and the best result is selected. Appropriate thresholding of the correlation coefficient is needed to minimize false positives. Once the template is matched the corresponding bus number and destination is retrieved from the database and announced. The similar process is followed for the rear video input. If the front and the rear outputs are same the system recognizes it as the same bus. If the outputs are different the bus number and the destination of the second bus is announced. The screenshots of the simulated work is shown below in figures 3 and 4. The respective results are announced through the speakers.

## 5. RESULTS AND DISCUSSIONS

### a) Front camera results:

Front camera video samples: 5

Table 2

Actual Bus Number	Bus Detected	Bus number Recognized	Recognized Number	Result
296	Yes	Yes	296	Correct
296*	Yes	Yes	206	Incorrect
207	No	No	N/A	Incorrect
203	Yes	Yes	203	Correct
206	Yes	Yes	206	Correct

\*Different sample of 296 bus to detect variation in results.

#### Accuracy:

From the above results for front camera,

Detection: 80%

Correct Recognition: 60%

### b) Rear Camera Results:

Rear camera video samples: 5

Table 3

Actual Bus Number	Bus Number Recognized	Recognized Number	Result
296	Yes	296	Correct
296	Yes	296	Correct
207	Yes	207	Correct
203	No	N/A	Incorrect
206	Yes	296	Incorrect

#### Accuracy:

From the above results for rear camera,

Correct Recognition: 60%

## 6. CONCLUSION AND FUTURE SCOPE

VIAssist is simulated using SURF algorithm and template matching technique. Image Processing has the potential to

help overcome the difficulties faced by visually impaired people. This system can be used for assisting the blind and can help them commute using the public bus transportation with ease.

In the future, this simulated work can be implemented on actual bus stops using the required hardware. The number of buses in the database can be increased. This system can be deployed on several bus stops. The scope of the project can be further refined by adding RFID tags to help the blind users traverse inside the bus. Also the information gathered at the current bus stop can be stored and passed to the following bus stops thus giving commuters a clear idea about the current location of the bus. When the system is idle, the camera can be used for surveillance purposes.

## 7. ACKNOWLEDGMENTS

We would like to thank our college for providing us the opportunity to embark on this project. We are extremely grateful to **Ms. Anuradha Raghvan** who provided the most valuable insight and helped in shaping the idea of this paper. We are thankful to all the family members and friends for their cooperation, motivation and help throughout the documentation of this paper.

## 8. REFERENCES

- [1] Deccan Herald: India Accounts For 20 percent of global blind population; New Delhi, April 6<sup>th</sup>, 2012, (PTI): [HTTP://WWW.DECCANHERALD.COM/CONTENT/240119/INDIA-ACCOUNTS-20-PERCENT.HTML](http://www.deccanherald.com/content/240119/india-accounts-20-percent.html)
- [2] G.Lavanya, 'Passenger BUS Alert System for Easy Navigation of Blind', IEEE 2013.
- [3] Claudio Guida, Dario Comanducci, and Carlo Colombo, 'Automatic bus line number localization and recognition on mobile phones—a computer vision aid for the visually impaired', Springer 2011.
- [4] Hangrong Pan, Chucai Yi and Yingli Tian, 'A Primary Travelling assistant system of Bus Detection and Recognition for Visually Impaired People', IEEE 2013.
- [5] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, 'SURF: Speeded Up Robust Features', Springer 2006.
- [6] OpenCV SURF Algorithm: [http://docs.opencv.org/trunk/doc/py\\_tutorials/py\\_feature/py\\_surf\\_intro/py\\_surf\\_intro.html](http://docs.opencv.org/trunk/doc/py_tutorials/py_feature/py_surf_intro/py_surf_intro.html)
- [7] OpenCV Template Matching Algorithm: [http://docs.opencv.org/doc/tutorials/imgproc/histograms/template\\_matching/template\\_matching.html](http://docs.opencv.org/doc/tutorials/imgproc/histograms/template_matching/template_matching.html)

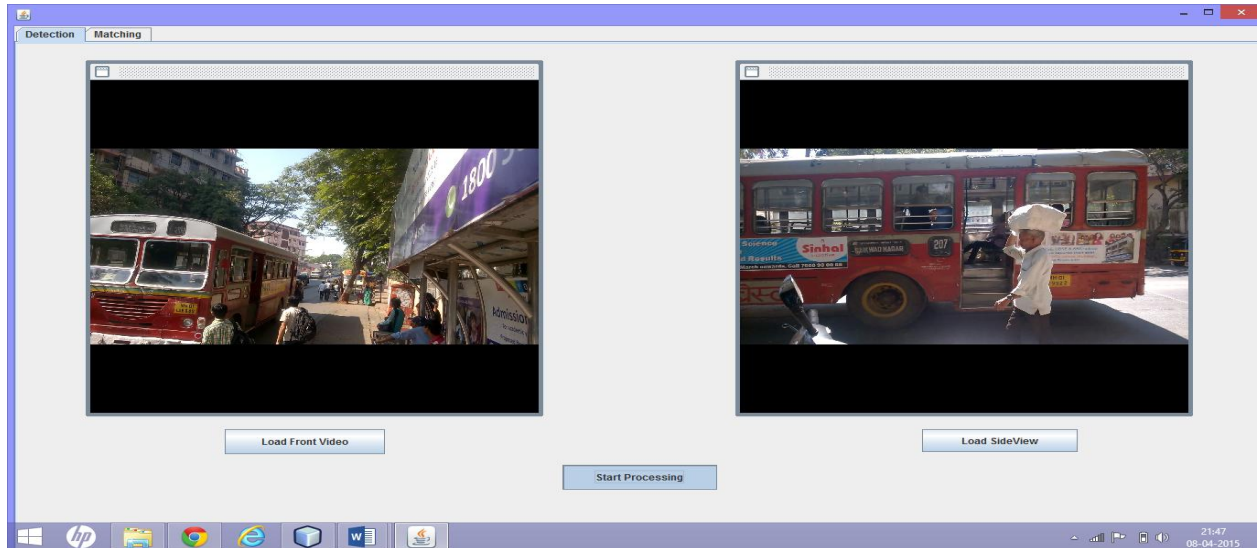


Figure 4 Simulation for Bus Detection

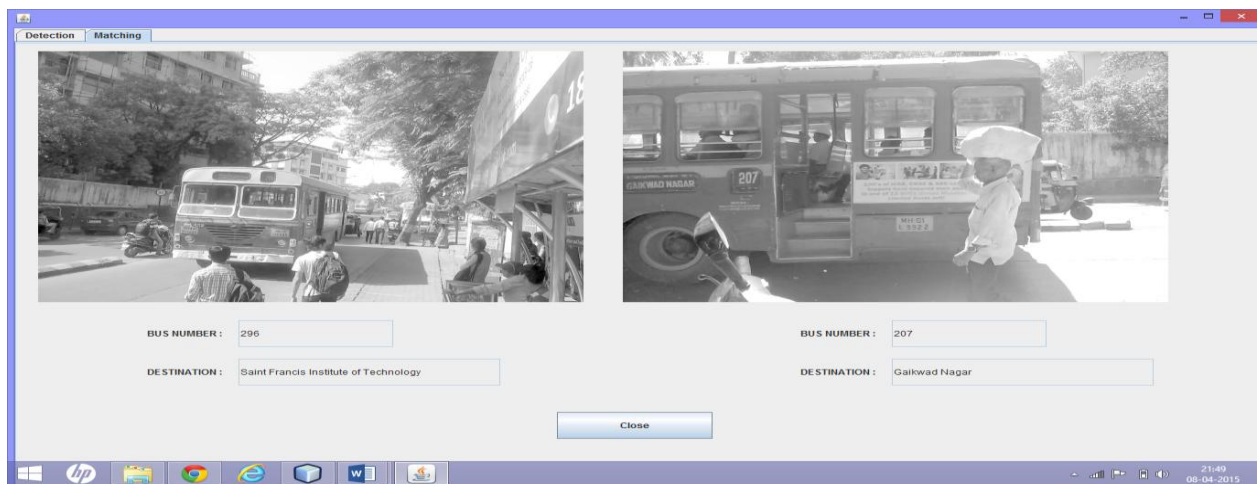


Figure 5 Simulation for Bus Number Recognition