ROBOT NAVIGATION USING IMAGE PROCESSING AND ISOLATED WORD RECOGNITION

Shashank, Gaurav, Manish Sharma, Anupam Shukla
Department of Information and Communication Technology
Indian Institute of Information Technology and management, ABV IIITM
Gwalior, India
shashank.iiitm08@gmail.com,

Sameer Shastri
Department of Electronics & Telecommunication Technology
Government Engineering College, GEC
Raipur, India

Abstract – In this paper, a practical implementation of image processing and isolated word recognition has been explained. A modeled vehicle has been driven autonomously on the road and can navigate on a road on its own, stop at zebra-crossing, follow traffic lights and reach an end point or can be controlled through voice commands. The vehicle has been developed using an old scrap body of a Kinetic Sunny and its engine, Servo motors, microphone, PCB and cameras. The image processing is done with the help of a laptop using MATLAB image processing toolbox and speech processing is done using HTK (hidden markov model toolkit). The robot navigated through roads with marked lanes and traffic light using image processing after taking inputs from the environment with the help of CCD camera and webcam. However, on the roads with no standardized lane marking and traffic lights, the robot was controlled using voice commands only and processed by HTK toolkit.

Keywords - AGV navigation, Image processing, isolated word recognition, MATLAB, HTKtoolkit.

I. INTRODUCTION

This paper explains about an automated guided vehicle (robot vehicle) which was developed for operating on the roads with real life like environment. The robot was designed as an outdoor robot [1] which can detect and follow lanes, detect zebra crossing, traffic red light and stop at an end point. The specifications of robot, arena and rules are discussed in details in the third section[2]. This also involves traffic light detection, lane following and zebra crossing detection which requires a significant amount of image processing. Isolated word recognition using HTK toolkitwas also implemented [3] later on for the purpose of practical study. We developed a speaker independent isolated word recognition [4] application for commanding the vehicle manually. As speech is one of the most efficient methods of interaction with a robot [5]. The robot was controlled using both image processing and isolated word commands.

II. RELATED WORK

In the past two decades, much work has been done in robotics, image processing and speech recognition.Jin-HyungPark, and Chang-sung Jeong of the Korea University worked on an algorithm for real time signal light detection for intelligent transport system and unmanned ground vehicle [6]. Xiaohong Cong, HuiNing ,Zhibin Miao of harbinenguniversity worked on an intelligent approach for obstacle detection for self-navigation which controls the wheels speed of the robot based on the distance from the obstacles. as per the geometry of the environment the optimal speed for the wheels were also calculated [7].

Traffic signs carry a lot of information about the traffic and environmentetc. which provides great assistance in driving. Hence, traffic sign identification is also very important for intelligent transport system or unmanned ground vehicle. AuranuchLorsakul and JackritSuthakorn worked on Traffic Sign Recognition Using Neural Networks and implemented it on OpenCV which can be used for Driver Assistance Systems [8]. Raoul de CharetteandFawziNashashibi proposed a generic real time traffic light recognition algorithm to detect both supported and suspended traffic lights which works in both rural and urban environments [9]. Seonyoung Lee, Haengseon Son and Kyungwon Min of Convergent SoC Research Center, Korea Electronics Technology

Institute worked on lane detection implementation using Hough transform which is a famous approach for lane detection in driver assistance system [10]. They propose a high performance and optimized Line Hough Transform circuit architecture for minimizing the logic and the number of cycle time.

In very rough terrain a robot can be controlled using speech commands. this method can be especially useful in the places where the terrain is uneven or the traffic signs and lanes are not present. Much work is done in isolated word recognition in English and many other languages. Kuldeep Kumar and R. K. Aggarwal of Nation Institute of Technology Kurukshetra developed a speech recognition system for Hindi language using HTK toolkit [11]. Sukhminder Singh Grewal and Dinesh Kumar worked on speaker independent isolated word recognition system for English language with 81.23 % success rate [12].

III. Problem Description

A. Image Processing Problem

1) *Specifications*:

a) *Road:*

- Black tar surface.
- The average width of the road would be 5 – 10 m.
- Non-continuous white divider lines along the length of the road.
- The length of a divider line would be about 1.5 m and distance between two consecutive white divider lines would be 2.5 to 4 m with a tolerance of 20%.
- The road shall have no explicit limiters on either side.
- The road has both straight and curved sections.

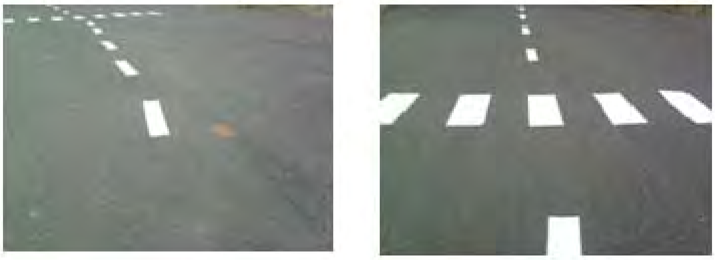

Figure 1.Road's image and Zebra crossing image via AGV's camera.

b) *Zebra Crossing*

- Many vertical white lines covering at least about half of the black area horizontally.
- It should be located at 3 to 5 m from the base of the traffic light.

c) *Traffic lights*

- Standard green and red lights with the following interpretations: RED-STOP and GREEN-MOVE.
- The lights shall be circular at a height of 2- 3m from the ground level.
- The box containing the traffic lights will be in horizontal orientation along the width of the road. The pole mounting the traffic lights can be on either side of the road.

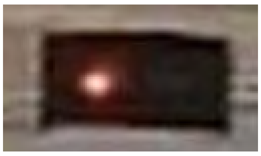

Figure 2.Red light used in the competition.

2) *The Robot should be able to perform the following tasks:*

- The robot sensors CCD spy camera and webcam] must be capable enough to recognize the divider strips, zebra crossing, traffic light signals and should be capable of catching them by taking pictures.
- It must be capable of following the white divider strip in straight line and turns or in short it should be able to follow the lanes on the road.
- It should be capable of recognizing the zebra crossing (i.e. a horizontal white strip on the road).
- It should be capable of following traffic signals (i.e. stopping on red lights and moving on green light).
- A person should be able to stop the vehicle by using an emergency kill switch.
- The vehicle will communicate with a laptop with the help of a digital circuit (Printed Circuit Board). The laptop will capture and store the image with the help of web cam and CCD spy cam for the required purposes. It will process the captured images and then it will pass the command to the PCB for the required action to perform the specific functions like breaking, steering and accelerating. These functions are performed by using DC servo motors which rotates at specific angle in specific direction for specific DC input.

B. SpeechRecognition Problem

In this case the robot should be controlled by a set of speech commands which are used for its navigation. This method is independent of the type of road and lanes, traffic signals etc. as the robot is controlled manually in real-time. This process is more useful in the places where the terrain is uneven and there is no proper maintenance of roadsand hence the image processing will fail. We used speaker independent isolated word recognition technique. This was done using hidden markov model which was implemented in HTK toolkit.

IV. WORKING OF SYSTEM

Two cameras (CCD camera & webcam) were mounted upon the robot along with a microphone on the laptop. The CCD camera detects lane and zebra crossing while the webcam is mounted at a height so, as to detect traffic light. The CCD camera provides input to an Analog to Digital converter which converts the signals and then sends them to the laptop mounted on the chassis of the vehicle. The microphone is directly attached to the laptop. The Laptop has MATLAB installed along with image processing toolbox and HTK toolkit.

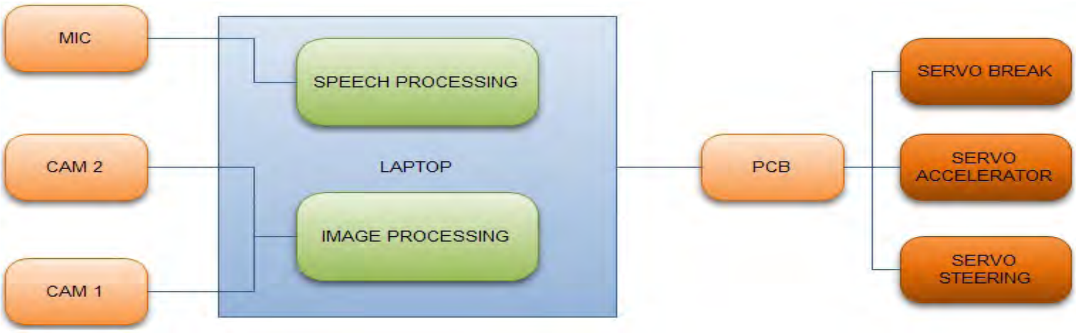

Figure 3.working system diagram.

The video feed from the CCD and webcam is being analyzed and processed by the MATLAB. The inputs from the microphone are analyzed by HTK toolkit. After processing the feeds the result is being directed to a PCB.

The Laptop is being connected to a printed circuit board (PCB) via serial port .The PCB acts as servo motor controller.

The PCB interprets the signals it receives from the laptop via serial port and then the microcontroller gives control signals to the servo motors. The servo motors act as a driver to the vehicle and they drive the vehicle according to the control signals given by the PCB. The steering servo motor changes the direction, while the acceleration and breaking servo motor accelerates or decelerates the vehicle.

V. CONSTRUCTED HARDWARE

The schematic and the actual image of the bot are shown in Figure4. Left hand side is the schematic diagram of side view of the bot and at the right hand side is the actual image of the side view of the bot when it was being constructed.

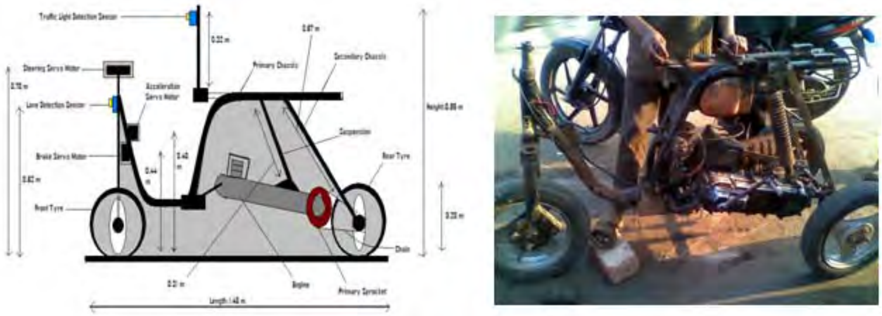

Figure 4. Side view schematic and Actual side view.

The rear view and its schematic are shown in Figure5. Left hand side is the schematic diagram of rear view of the bot and at the right hand side is the actual image of the rear view of the bot when it was being constructed. The fully constructed robot on which we have tested our image processing and isolated word recognition systems is shown in Figure 6.

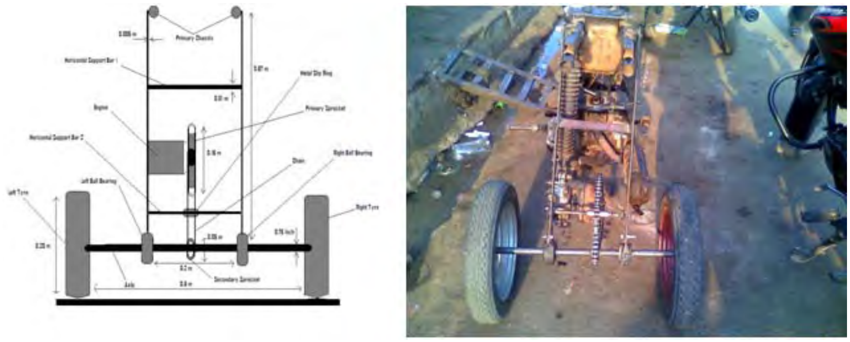

Figure 5. Rear view schematic and actual rearview.

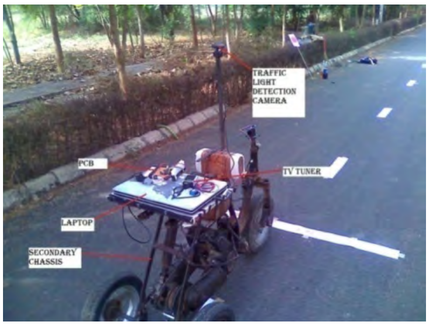

Figure 6. Complete Bot (running).

VI. ISOLATED WORD RECOGNITION SYSTEM ARCHITECTURE

The system developed by us is shown in the Fig. 7. The system is divided in two parts, testing part and training part. Training part is used for generating the input voice command tester. The model is discussed in details below:

A. Preprocessing

This is used to convert input analog signal into digital signal. Then the digital signal is spectrally flatten using first order filters.

B. Feature Extraction

This method is used to extract a set of parameters which differentiate the spoken words on the basis of these properties. These parameters are generated by processing the input digital signal. In this we process some special characteristics of signal (i.e. frequency response, energy of the signal etc.).

C. Training Model Generation

We have generated our training model using Hidden Markov Model [13]. There are other techniques available which could be used to generate training model like Artificial Neural Networks [14], Support Vector Machine [15] etc.

D. Word Classifier

This component is used to recognize the samples based on their unique parameters with the help of training model generated.

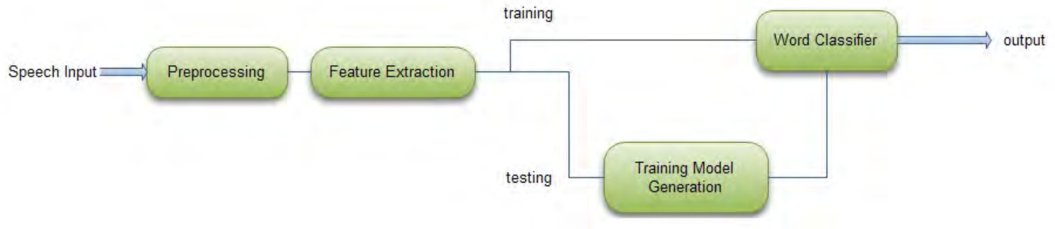

Figure 7. Word Recognition Model.

VII. ALGORITHMS AND COMPUTATION

We processed 4 frames per second for lane follower, 5 frames per second for traffic light detection and 2 frames per second for zebra crossing. The bot never speeded above 12 km/h (approx.). As the computation power was limited.

A. Algorithm used For Lane Following

```
• Capture image
• Crop it
• Turn Image to binary image
• Optimize image and remove noise
• Fill the holes
• Divide the image in two halves: left region and right region
• Check the ratio of lane strip area on the left side to the right side
• If the ratio of the left side of the image to the right side of the image is > 1.25
        ▪ Then    Turn left
• Else if the ratio of the left side of the image to the right side of the image is< 0.8
        ▪ Then    Turn right
• Else drive straight
```

B. Algorithm for Zebra Crossing Detection

```
• Capture image
• Crop it
• Turn it to binary image
• Fill the holes
• For every row of pixels
• If white pixel Then increase the counter;
• If counter is greater than 40% of row size of resolution
  CHECK (TRAFFIC LIGHT)
  If (! TRAFFIC LIGHT) Then FOLLOW LANE
```

C. Algorithm for Traffic Light Detection

- Capture image
- Crop it
- Process pixels diagonally (5 to 6 lines along the diagonal)
- Find the R G B values of pixel
- If (R-value>210) && (G-value<175) && (B-value<175)
- Count the pixel as red.
- If multiple red pixels are found
- Then Convert the image into binary and check whether it is circular by calculating the correlation coefficient between the current image and the capture binary images of other traffic red lights (we have used 8 different red light images for this).
- If (multiple correlation coefficient values (more than 4) above 0.250)
- Then it is a red light.

D. Procedure for creating speech database for training the isolated word recognition model

The five English words START, STOP, GO, LEFT and RIGHT were recorded by 8 speakers, each word uttered 5 times by every speaker. Thus giving a total no of 200(i.e. (5*5)*8) speech files. Voice inputs of 8 speakers were used to train the system. Each word being uttered 5 times by 8 speakers were used as test data. All the recordings have been done using a wave –surfer software with a good quality microphone (for training) [17].

E. MFCC parameterization

Mel-Frequency Cepstral Coefficients (MFCCs) are widely used features for automatic speech recognition systems to parameterize the speech waveform as sequence acoustic vectors.
To parameterize speech waveform into MFCC, HCopy tool of HTK toolkit is used. HCopy tool requires input as speech waveform (.wav format) and configuration file having set of parameters required for MFCC conversion .Configuration file used for MFCC parameterization contains the following settings [16].

```
SOURCEFORMAT = WAV
TARGETKIND = MFCC_0_D_A
TARGETRATE = 100000.0
SAVECOMPRESSED = T
SAVEWITHCRC = T
WINDOWSIZE = 250000.0
USEHAMMING = T
PREEMCOEF = 0.97
NUMCHANS = 26
CEPLIFTER = 22
NUMCEPS = 12
```

In brief here the target kind are MFCC using C_0 as the energy component, target rate specify frame period (HTK uses units of 100ns), the output should be saved in compressed form, and a CRC checksum should be added. The Fast Fourier Transform uses a Hamming window and pre-emphasis of speech signal will be performed using a coefficient of 0.97. The filter bank should have 26 channels and12 MFCC coefficients, 12 delta coefficient, 12 acceleration coefficient generated as output.

F. Training the HMM

Acoustic model used for speech recognition is HMM which generated using HTK toolkit.We have used phone based acoustic model in which we model only parts of wordsgenerally phones. And the word itself is modeled as sequence of phone.First we need to define a prototype model for HMM training, which are then re-estimated using the MFCC files and their associated transcription. Apart from the models of monophones, model for silent (sil) must be included.For prototype model, we have used a 3 state left-to-right HMM with no skips topology. The prototype models are initialized using the HTK tool HInit which initializes the HMM model based on one of the speech recordings. Then HERest is used to re-estimate the parameters of the HMM model based on the other speech recordings in the training set. As the dictionary contains multiple pronunciations for some words, The phone models created so far can be used to *realign* the training data and create new transcriptions, Using

HTK recognition tool HVite, After phone alignments, HMM set parameters are re-estimated using HTK tool HERest and we get monophone HMM set.

In final step of acoustic model building is to create context-dependent triphone HMMs. Firstly, the monophone transcriptions are converted to triphone transcriptions and a set of triphone models are created by copying the monophones and model re-estimated twice using HTK tool HERest. Secondlytriphones are converted into tied state triphones to ensure that all state distributions can be robustly estimated [16].

VIII. EXPERIMENTAL RESULTS

A. Results for Image Processing Part

The captured frames were processed in MATLAB image processing toolbox. The step by step processing of frame for Lane Following in shown in Figure 8. The correlation coefficientvalues for comparison of traffic light with various other traffic lights are shown in Table II. The time required by each process is shown in Table I.

TABLE I

PROCESSES	TIME REQUIRED
LANE FOLLOWING	0.15 seconds (approx.)
TRAFFIC LIGHT DETECTION	0.05 seconds (approx.)
ZEBRA CROSSING DETECTION	0.32 Seconds (approx.)

TABLE II

REFERENCE TRAFFIC LIGHT NUMBER	CORRELATION COEFFICIENT
Traffic light 2	0.2125
Traffic light 3	0.3218
Traffic light 4	0.2852
Traffic light 5	0.2759

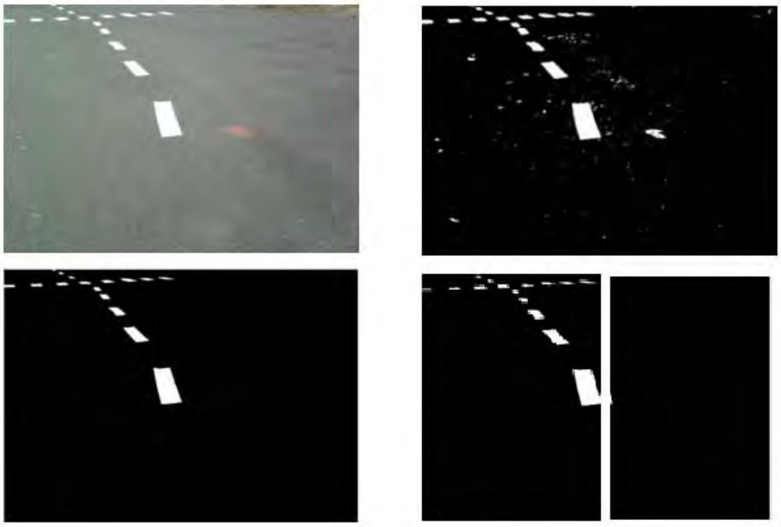

Figure 8. Frame processing for lane following.

B. Results for Isolated Word Recognition Part

Our isolated word recognition was trained by 8 people. Then the system was tested for 5 speakers, out of which 2 speakers were known and 3 speakers were unknown. The overall accuracy of the system was 91% and word error rate was 9%. Detailed results are mentioned in the Table III.

TABLE III

S. No.	Speaker number	Number of spoken commands	Correctly Recognized commands	% word accuracy	Word error rate
1	Speaker 1 (known)	20	19	95%	5%
2	Speaker 2 (known)	20	20	100%	0%
3	Speaker 3 (unknown)	20	18	90%	10%
4	Speaker 4 (unknown)	20	18	90%	10%
5	Speaker 5 (unknown)	20	16	80%	20%

The raw waveform and the corresponding MFCC plot for voice command ‘START’ are shown in Figure 9 and Figure 10.

IX. CONCLUSION AND FUTURE WORK

This robot was made as a prototype lane follower and can be used as a helper for drivers, people with disabilities, load carrier in plants and mines by road. And the persons who cannot drive can use voice commands to navigate the vehicle.Voice commands can also be used in uneven terrain, absence of proper marking on the road etc.

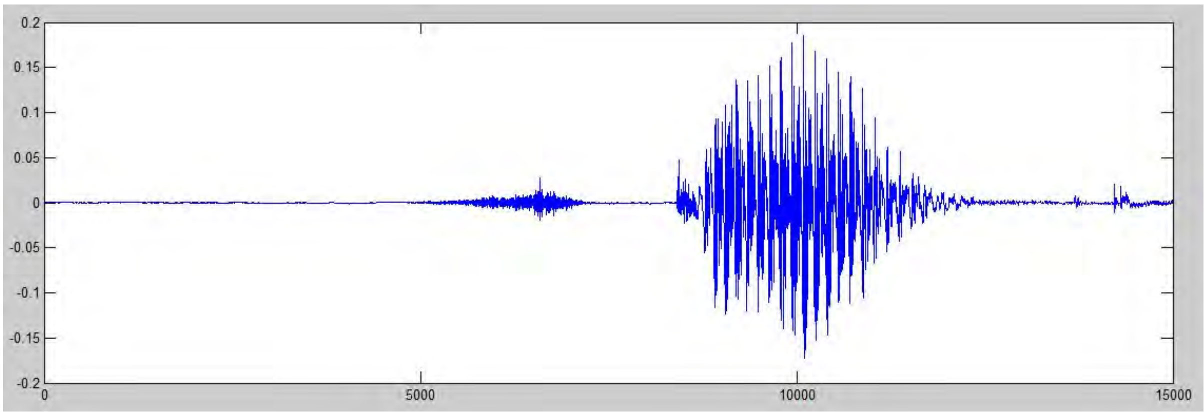

Figure 9. Speech Waveform for word START before preprocessing.

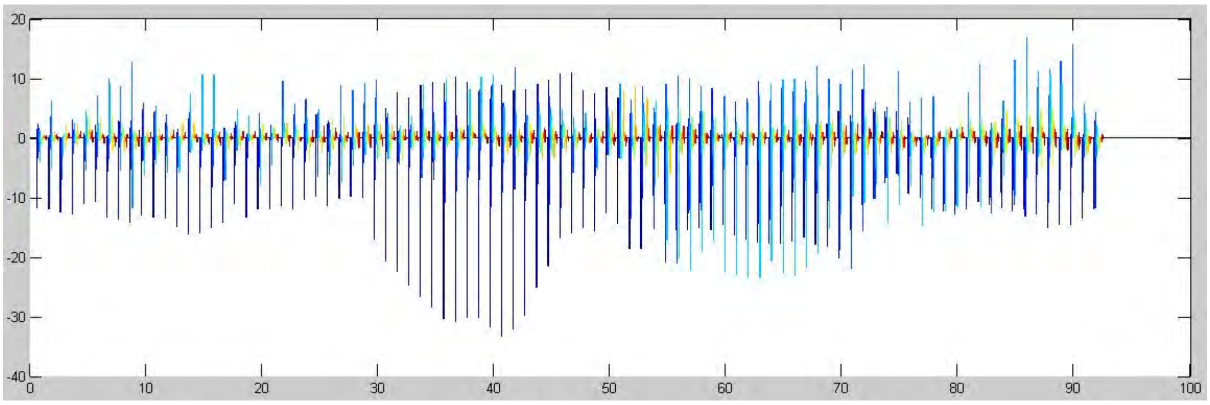

Figure 10. MFCC plot for the speech waveform input after pre processing

This work can be improved further by using better computing machine with more accurate algorithms for traffic light detection, lane following and Isolated word recognition system can be trained with very large number of people for better accuracy.

REFERENCES

[1] Ratner, D., McKerrow, P., “Navigating an outdoor robot along continuous landmarks with ultrasonic sensing”, (2003) Robotics and Autonomous Systems, 45 (2), pp. 73-82.
[2] http://www.robotix.in/rbtx09/acc_8milekshitij 8mile competition problem statement.
[3] http://htk.eng.cam.ac.uk/HTK (hidden markov model toolkit).
[4] Sukhminder Singh Grewal, Dinesh Kumar, "Isolated Word Recognition system for English language", International Journal of Information Technology and Knowledge Management, July-December 2010, Volume 2, No. 2, pp. 447-450.
[5] S. Furui “Speaker Independent Isolated Word Recognition using Dynamic features of Speech Spectrum”, IEEETransactions on acoustic, speech, and signal processing, Vol. ASP-34, NO.I,February 1986.
[6] Jin-Hyung Park, Chang-sung Jeong "Real-time Signal Light Detection" International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.2, No.2, June 2009.
[7] Xiaohong Cong, HuiNing ,Zhibin Miao, "A Fuzzy Logical Application in a Robot Self Navigation" Industrial Electronics and Applications, 2007. ICIEA 2007. 2nd IEEE Conference, may 2007.
[8] AuranuchLorsakul, JackritSuthakorn, "Traffic Sign Recognition Using Neural Network on OpenCV: Toward Intelligent Vehicle/Driver Assistance System", 4th International Conference on Ubiquitous Robots and Ambient Intelligence URAI 2007 (2007).
[9] Raoul de Charette, FawziNashashibi, "Traffic light recognition using image processing compared to learning process", IEEE/RSJ International Conference on Intelligent Robots and Systems, october 2009.
[10] Seonyoung Lee, Haengseon Son, Kyungwon Min, "Implementation of Lane Detection System using Optimized Hough Transform Circuit" Circuits and Systems (APCCAS), 2010 IEEE Asia Pacific Conference, december 2010.
[11] Kuldeep Kumar, R. K. Aggarwal, "hindi speech recognition system using HTK", International Journal of Computing and Business Research ISSN (Online), Volume 2 Issue 2 May 2011.
[12] Sukhminder Singh Grewal, Dinesh Kumar, "Isolated Word Recognition system for English language", International Journal of Information Technology and Knowledge Management, July-December 2010, Volume 2, No. 2, pp. 447-450.
[13] Mei-yuhHwang , Xuedong Huang, “Shared-Distribution Hidden Markov Models for Speech Recognition”, IEEE Transactionson Speech and Audio Processing, VOL. 1, NO. 4, OCTOBER 1993.
[14] PiotrWilinski, Basel Solaiman, A. Hillion, W. Czarnecki. Towards the border between neural and Markovian paradigms. IEEE Transactions on Systems, Man, and Cybernetics, Part B, 1998: 146~159
[15] G-D. Guo and S. Li, Content-based Audio Classification and Retrieval by Support Vector Machines, IEEE Trans. on Neural Networks, Vol. 14, No. 1, 209-215, January, 2003.
[16] The HTK Book for HTK Version 3.4, 2009 Cambridge University Engineering Department, ‘http://htk.eng.cam.ac.uk/docs/docs.shtml”.
[17] Jong-Seok Lee, Cheol Hoon Park, "Robust Audio-Visual Speech Recognition Based on Late Integration", Multimedia, IEEE Transactions on, On page(s): 767 - 779 Volume: 10, Issue: 5, Aug. 2008

AUTHORS PROFILE

Mr. Shashank was born in 1988. He is presently doing his Integrated post-graduation (B. Tech and M. Tech) in Information and Communication Technology form Indian Institute of Information Technology and Management, Gwalior. His research interest includes Image processing, Medical Image Processing, computer vision, Digital Signal Processing and Robotics.

Mr. Gaurav was born in 1985. He has done is B. Tech in Information Technology from Krishana Institute of Information Technology in 2009. He is presently doing his post-graduation (M. Tech) in Computer Science with specialization in Information Security form Indian Institute of Information Technology and Management, Gwalior. His research interest includes Signal Processing, Operating Systems,Image Processing Robotics and Information Security.

Dr. AnupamShukla is an Associate Professor in theICT Department of the Indian Institute of Information Technologyand Management Gwalior. He completed his PhD degree from NITRaipur, India in 2002. He did his post-graduation from JadavpurUniversity, India. He has 22 years of teaching experience. Hisresearch interest includes Speech processing, ArtificialIntelligence, Soft Computing and Bioinformatics. He has publishedaround 120 papers in various national and internationaljournals/conferences. He is referee for 4 international journals andin the Editorial board of International Journal of AI and SoftComputing. He received Young Scientist Award from MadhyaPradesh Government and Gold Medal from Jadavpur University.

Mr. Manish Sharma was born in 1989 and pursuing Integrated Post Graduation (B. Tech in Information Technology and M.B.A. in Finance) at ABV- Indian Institute of Information Technology and Management, Gwalior. He is currently in the fourth year of the five year program. His areas of interest are robotics, business analysis, and asset valuation.

Mr. Sameer Shastriis working as a lecturer in government engineering college, Raipur. He is working in area of autonomous vehicle guidance and control. He has 12 years of teaching and research experience and papers in International and National journals.