International Journal of Advanced Technology in Engineering and Science www.ijates.com
Volume No.03, Issue No. 06, June 2015 SN (online): 2348 – 7550

# IMPLEMENTATION OF REAL TIME EMBEDDED AUDIO-VISUAL TRACKING USING DUEL PROCESSOR

**P.G.Shinde[1], Dr.D.M.Yadav[2]**

*[1,2]Electronics & Telecommunication Engineering Department, Savitribai Phule Pune University, (India)*

**ABSTRACT**

*We examined the benefits of embedded systems on audio-visual tracking. The Aim of this paper is to plan an installed audio-visual tracking. Reconciliation of sound-related and visual observations has turn into a pattern in robots or smart human machine interface. The framework has the capacity perform human face tracking, voice action identification, sound source heading estimation continuously. Evaluating the sound source heading serves to introduce the human face tracking when target alter course. The implementation architecture is based on an embedded ARM11 (Raspberry pi) and AT mega 16 processor. For speech signal processing an eight channel digital microphone array is developed and the associated pre-processing and interface feature are designed. By using Viola-Jones object detection framework we can detect human face which is used in many application such as security, robotics, Surveillance. All the experiments are conducted in real time environment and the experimental results show that this system can execute all the audition and vision function in real time. For visual tracking face detection algorithm is used for face detection.*

***Keywords ARM 11, Audio-Visual Tracking , AT MEGA 16 Processor.Face Detection Algorithm.***

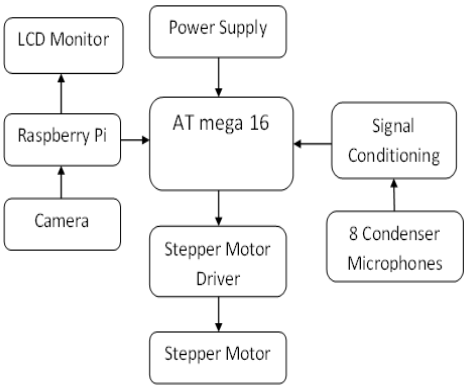
## I. INTRODUCTION

Integration of auditory and visual perceptions has become a trend in robots or intelligent human–machine interfaces .For intelligent machines to interact naturally with human, the abilities to understand spoken language and respond to auditory events are necessary since the auditory system can provide useful information about the environment, such as sound source location and the interpretation of the content of the sound sources. However, in real environment, the speech signal is easily corrupted by interference signal such as other talkers or room reverberation. Hence, sound source tracking and speech enhancement are two basic functions of the auditory system. Secondly, sound source tracking is usually followed by visual tracking when the target is within the view scope of the camera. The integrated interface greatly enhances the robustness of human or target tracking. However, the integration also poses a great challenge in technology development. This paper describes the auditory–visual system architecture, algorithms and implementation on an embedded platform that can be applied to robots or other human machine interfaces.

Face detection in image sequence has been an active research area in the computer vision field in recent years due to its potential applications such as monitoring and surveillance, human computer interfaces, smart rooms , intelligent robots , and biomedical image analysis . Face detection is based on identifying and locating a human

face in images regardless of size, position, and condition Simple features such as color, motion, and texture are used for the face detection in early researches. However, these methods break down easily because of the complexity of the real world. Face detection proposed by Viola and Jones is most popular among the face detection approaches based on statistic methods. Although real-time face detection is possible using high performance computers, the resources of the system tend to be monopolized by face detection.

## II. IMPLIMENTATION METHODOLOGY:




**Fig(1): Block Diagram for Audio-Visual Tracking Section**

Figure (1) shows the complete block diagram for audio-visual tracking .In this work, we propose and implement the human face tracking, voice activity detection, and sound source direction estimation on an embedded duel platform. It includes an Advanced RISC Machine (ARM 11) processor also called Raspberry Pi and AT mega 16 microcontroller. All of the audio algorithm are executed on AT mega16 and ARM is used to execute the human face tracking algorithm. The audio interface includes an eight-channel microphone array.8 condensor microphones are used for audio tracking. Here unipolar, permanent magnet 12 volt stepper motar is used for movement of camera. Each microphone circuit consists of amplifiers, filters, and comparators.

A stepper motar is an electromechanically device which converts electrical pulses into discrete mechanical movements. Here MAX232 required for PC. The MAX232 is a dual driver/receiver that includes a capacitive voltage generator to supply EIA-232 voltage levels from a single 5-V supply.MAX232 work as voltage level converter or it is also called as charge pump capacitor.

4 bit LCD is used for debugging purpose.

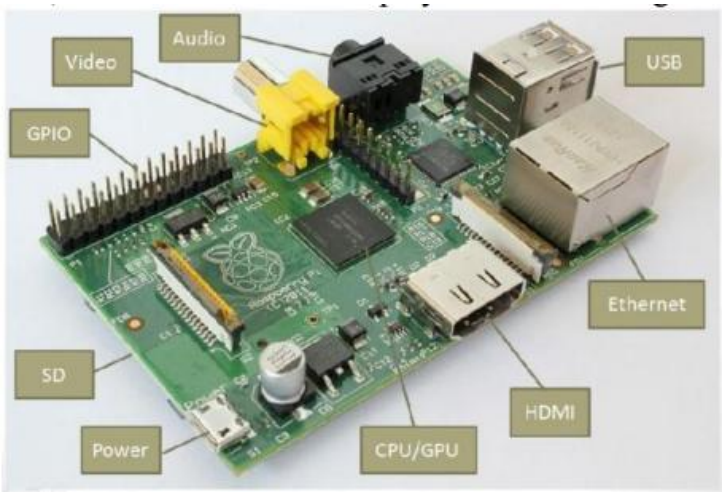
In this paper for audio-visual tracking different software are used AVR Studio, Proteus, and Express PCB and for visual face detection open cv is used on ARM11.

Proteus is one of the most famous simulators. It can be uses to simulate almost every circuit on electrical fields. It is easy to use because of the GUI interface that is very similar to the real Prototype board. Moreover, it can be used to design Print Circuit Board (PCB). For image capturing and image processing we have used OpenCV on Raspberry Pi. OpenCV i.e. Open Source Computer Vision is a library of programming functions that helps in real-time image processing. OpenCV is written in C++ and its primary interface is in C++. The library is cross-platform. It focuses mainly on real-time image processing.

OpenCV (Open Source Computer Vision) is a library of programming functions mainly aimed at real-time computer vision, developed by Intel Russia research center in Nizhny Novgorod, and now supported by Willow Garage and Itseez. It is free for utilization under the open-source BSD permit. The library is cross-stage. It concentrates basically on constant image handling. On the off chance that the library finds Intel's Integrated

Performance Primitives on the framework, it will utilize these restrictive enhanced schedules to quicken itself.

## III. OVERVIEW OF THE RASPBERRY PI




**Fig(2): Raspberry Pi Model.**

The Raspberry Pi is a credit-card sized computer. We widely used ARM-based processors include: low cost, low-to-very-low power consumption, decent processing power, and open development environment. Raspberry Pi has an in number preparing limit on the grounds that of utilizing the ARM11 construction modeling and Linux-based framework.

The image of Raspberry Pi is indicated in Fig. . A typical web camera is utilized for live feature catching. The associations are additionally demonstrated in the image. This framework proposed here utilizations Raspberry Pi Model B associated with a intex web camera all inclusive serial transport (USB) webcam. Power supply needed by the gadget is 3V DC which is being changed over first from 5V DC. Outer hard commute is needed for information stockpiling. Here 16 GB Secure Digital High Limit rise above card is utilized. Web association is being given through CAT6 Ethernet link. The camera module is being joined to the stepper engine which is being controlled through the gadget. This is done to guarantee that camera catches pictures in all bearings. Raspberry Pi model is an ARM 11 processor with portrayal and further determinations are indicated in the table underneath.

Raspberry Pi is used for video processing and sending the processed video to user PC The images captured by the camera should be processed very fast to provide real time visualization of environment to the user.

## IV. ANALYSIS OF FACE DETECTION

The human face poses much a larger number of issues than different items since the human face is a dynamic question that comes in numerous structures and hues . On the other hand, facial detection and following gives numerous advantages. Facial acknowledgment is unrealistic if the face is not disconnected from the foundation. Human Computer Interaction (HCI) could incredibly be enhanced by utilizing feeling, posture, and motion acknowledgment, all of which oblige face and facial element detection and following . Albeit a wide range of calculations exist to perform face detection, each has its own particular shortcomings and qualities. Some utilization substance tones, some utilization forms, and other are significantly more intricate including layouts, neural systems, or channels. These calculations experience the ill effects of the same issue; they are

computationally lavish . A picture is just an accumulation of shading and/or light force values. Investigating these pixels for face detection is time intensive and hard to finish on account of the wide varieties of shape and pigmentation inside of a human face. Pixels frequently require reanalysis for scaling and exactness.

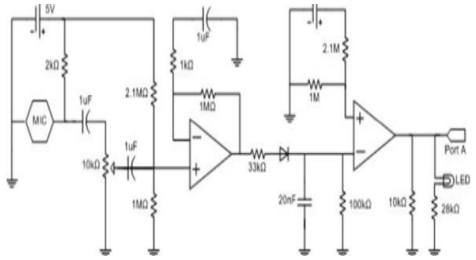
Viola and Jones devised an algorithm, called Haar Classifiers, to rapidly detect any object, including human faces, using AdaBoost classifier cascades that are based on Haar-like features and not pixels.

Face detection methods have been investigated for a considerable length of time and much advance has been proposed in writing. The majority of the face detection systems concentrate on recognizing frontal faces with great lighting conditions. These techniques can be ordered into four sorts: information based, element invariant, format coordinating and appearance-based.

A methodology used to identify protests when all is said in done, pertinent to human faces too was exhibited by Viola-Jones. This strategy demonstrated to distinguish protests greatly quickly and is equivalent to the best ongoing face detection frameworks. Viola and Jones (2004) displayed in their exploration another picture representation called Integral Image which permits quick computation of picture components to be utilized by their detection calculation. The second step is a calculation in view of AdaBoost which is prepared against the pertinent item class to choose a negligible arrangement of elements to speak to the article. Viola and Jones utilized components separated from the preparation set and AdaBoost calculation to choose the best list of capabilities and building the last classifier which includes few stages. Every stage comprises of couple of basic powerless classifiers that cooperate to frame a more grounded classifier. Each stage consists of few simple weak classifiers that work together to form a stronger classifier filtering out the majority of false detections at early stages and producing an adequate final face detector.

## V. ANALYSIS OF AUDIO TRACKING

Condenser microphone (mic) is a transducer that uses a capacitor (or condenser) to convert acoustical energy into electrical energy. The sound waves result into electric signals that can be used to drive a circuit or device. The pin connected to the case of mic is connected to ground and the other pin gives the output. here 8 condensor microphone array are used to detect the sound source and sound direction .To control the level of the microphone output, resistors are used to center the signal around 1.5V. This level-shifted output is amplified via an operational amplifier and then passed through a passive low pass filter. It is then put through a half-wave rectifier with a capacitor to bridge the gaps between positive swings caused by the half-wave rectifier. That output is then passed though an analog comparator to discretize the signal for reading into the port pin on the MCU. The discrete output of the microphone circuit is approximately 4V when sound is detected and 0V when no sound is detected.



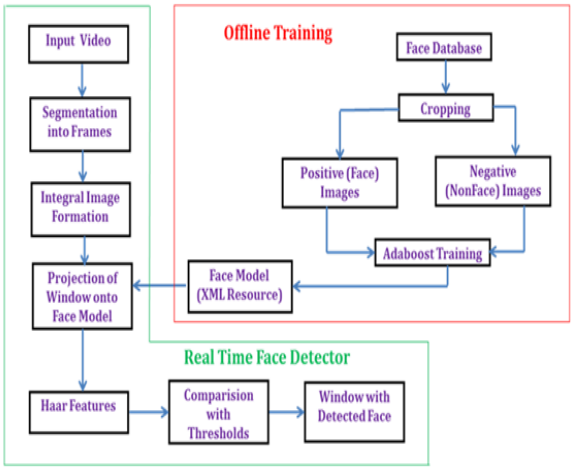

**Fig(2)Microphone Circuit**

All 8 microphone circuit outputs are read in parallel using PORTA on the microcontroller. The output is also passed through an LED circuit to ground in order to debug when a particular microphone circuit detects sound. The audio tracking is work on fastest finger first algorithm.8 microphone array are placed in circular manner at 45 degree angle ,sound source transmit signal in all direction propagate in all direction and sound is detected on which microphone sound come first and then send command to stepper motar to rotate the angle and take a picture. visual tracking is then perform on ARM 11processor by using Viola and Jones face detection algorithm.

## VI. FACE DETECTION ALGORITHM

Here for face tracking we use Viola-Jones object detection algorithm Paul Viola and Michael Jones presented a fast and robust method for face detection which is 15 times quicker than any technique at the time of release with 95% accuracy at around 17 fps. This work has three key contributions:

1. "**Haar-Like**" feature representation
2 "**Integral image**"
3. **AdaBoos**t learning algorithm.




**Fig(4):Face Detection Flowchart**

## VII. CONCLUSION

An embedded audio-visual tracking is implemented and we can detect sound source and face detection based on voice activity detection speaker direction estimation, speech and face detection algorithm by using eight microphone array and face detection using ARM11.By using Viola-Jones object detection framework we can successfully detect human face which is used in many application such as security, robotics, Surveillance.

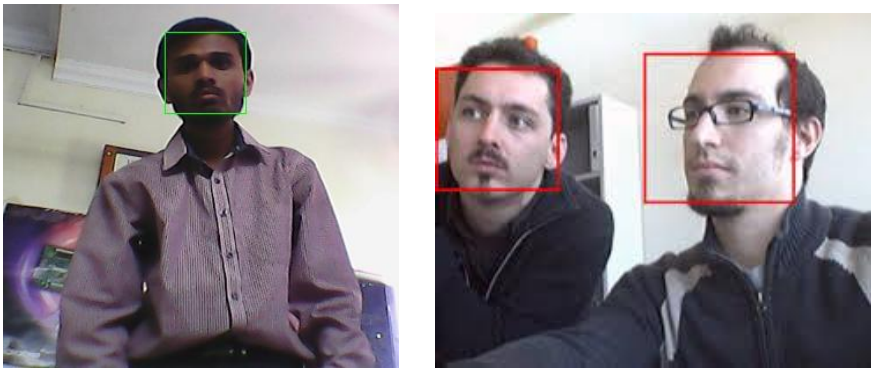
## VIII. ACKNOWLEDGMENT

I would like to express my gratitude & sincere thanks to my guide Dr.D.M.Yadav, of Electronics and Telecommunication Engineering, for his constant motivation and support during the course of my work.

We would like to thank to Dr. Y.S.Angal, HOD, Electronics and Telecommunication Engineering, Pune for his support and providing us the facilities.

## IX. RESULT

In this work, we design a high integrated audio–visual tracking system based on Voice activity detection, speaker direction estimation, Face detection algorithm is implemented on ARM11 , andATmega16. Eight-channel digital microphone array data acquisition system is also detect the sound source and sound direction .The visual tracking is performed on ARM11 using Viola-Jones object detection framework . some output images are shown below



**Fig(5): Experimental Result Images**

## REFERENCES

[1] J. Hu, C.C. Cheng, W.H. Liu, Robust speaker's location detection in a vehicle environment using GMM models, IEEE Transactions on System, Man and Cybernetics, Part B 36 (2) (2006) 403–412.

[2] K. Nakadai, T. Lourens, H.G. Okuno, H. Kitano, Active audition for humanoid, in: Proceedings National Conference on Artificial Intelligence, 2000, pp. 832–839.

[3] ] H.G. Okuno, K. Nakadai, T. Lourens, H. Kitano, Sound and visual tracking for humanoid robot, Springer Applied Intelligence 20 (2004) 253–266.

[4]Paul Viola, Michael J. Jones, "Robust Real-Time Face Detection", International Journal of Computer Vision 57(2), 137–154, 2004.

[5] Jwu-sheng Hu, Ming-Tang Lee, Chia-Hsing Yang, "An embedded audio-visual tracking and speech purification system on a dual core processor platform "Microprocessor & Microsystem Journal Volume 34,Issue7-8November 2010.page no 274-278.

[5] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in IEEE ICIP 2002, Vol. 1, pp 900-903, 2002.

[6]."Human and machine recognition of faces: A survey", R. Chellappa, C. L. Wilson, and S. Sirohey, Proc. of IEEE, volume 83, pages 705-740, 1995

[7] M.Gopi Krishna, A. Srinivasulu, , "Face Detection System On AdaBoost Algorithm Using Haar Classifiers". International Journal of Modern Engineering Research (IJMER) Vol. 2, Issue. 5, Sep.-Oct. 2012 pp-3556-3560

[8] Dr. Shantanu K. Dixit, Mr. S. B. Dhayagond, "Design and Implementation of e-Surveillance Robot for Video Monitoring and Living Body Detection" International Journal of Scientific and Research Publications, Volume 4, Issue 4, April 2014 1ISSN 2250-3153.