ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 2, Issue 3, March 2013*

# Haptic Robotic Arm Using Voice & Gesture Recognition

Aishwarya Patil[1], Anilkumar Daharwal[2], Ankur Harshe[3], Mohnish Gakare[4], Monika Sajjanwar[5]

Student, Dept. of Computer Technology YCCE Nagpur, Maharashtra, India[1,2,3,4,5]

**Abstract:** Our project presents a wireless interface to control an arm using voice & gesture commands through a computer. Voice recognition is the process of taking the spoken word as an input to a computer program. This process is important to virtual reality because it provides a fairly natural and intuitive way of controlling the simulation while allowing the user's hands to remain free. Arm with Voice Recognition is to create a wireless voice controlled arm which can be operated through a range of 10 to 50 meters using transmitter and receiver. This vehicle is equipped with a wireless camera which will transmit the live pictures and videos remotely. Voice recognition is "the technology by which sounds, words or phrases spoken by humans are converted into electrical signals, and these signals are transformed into coding patterns to which meaning has been assigned". Robotic arms are generally heavy rigid devices. They operate in controlled environments and perform highly repetitive tasks under pre-programmed control. The main goal is to present a programming robotic arm system for carrying out flexible pick and place behaviour and control it using voice recognition and gesture recognition. This is the most advance version of "pick n place robot" perhaps and most popular and widely used in recent industries. Arm is equipped with the camera which will transmit video in the database.

**keywords:** Voice Recognition, Gesture Recognition, Morphing Methods, Contour and Hull.

## I. INTRODUCTION

Our project consists of Haptic robotic arm where Haptic is the science of applying touch sensation and control to interaction with computer applications. Arms are types of jointed robot manipulator that allow robots to interact with their environment. Many have onboard controllers or translators to simplify communication, though they may be controlled directly or in any number of ways. Here we control our robot by voice and gesture commands. Our robotic arm is an automatic manipulator, which is easily programmable and controlled by basic voice Command. As the robot is wirelessly connected to user so this robot can be used in non-human friendly situation. Haptic robotic arm is a robot whose motions can be controlled by the user by giving specific voice and gesture commands.

It acquired its information using a speech recognizer then by comparing and converting the speech to text it manipulates its grip. The graphical user interface running along with the software provides a very convenient method for the users to train. It also provides many other facilities in operating the robot. After processing the speech, the necessary motion instructions are given to the robotic platform via a RF link. The speech recognition software running on a PC is capable of identifying the 6 voice commands issued by a authenticated user.

Hand gesture has been one of the most common and natural communication media among human being. Hand gesture recognition research has gained a lot of attentions because of its applications for interactive human-machine interface and virtual environments. The gesture recognition software on PC is capable of identifying 4 gesture commands. Hand gestures provide a separate complementary modality to speech for expressing ones ideas. So, a natural interaction between humans and computing devices can be achieved by using hand gestures for communication between them.

## II. LITERATURE SURVEY

The literature and survey in video-surveillance systems is large. Traditionally, these systems were based on static sensor devices such as CCTV cameras and later on, smart cameras were used, as computer vision algorithms were able to be embedded on these sensors. The image processing in video-surveillance systems mainly consists in video analysis of the monitoring area. The analysis may be used to interpret stationary objects and people of a scene or to interpret dynamic scenes. The analysis of dynamic scenes are based on the motion estimation or tracking people to perform behavioural analysis or activity recognition presents a survey of the state of-the-art on learning and understanding scene activity.

Automatic speech recognition (ASR) systems may prove useful in situations where the eyes and hands are occupied with control and monitoring functions. Speech

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 3, March 2013

recognition technology may allow the driver of a vehicle to concurrently perform certain in-vehicle secondary tasks without adversely affecting performance of the primary driving task. Currently, in-vehicle secondary tasks are accomplished through manual input from the driver. Automotive systems that are designed to be controlled via manual input may affect driving performance by forcing the driver to look away from the forward road scene and remove one hand from the steering wheel to complete the task.

The fact that the driver must divert visual, manual, and/or cognitive attention away from the primary driving task to successfully perform an in-vehicle task raises concerns over how these in-vehicle systems should be designed and what modality of input should be used to control these systems. The main tasks for a team of surveillance robots are: *patrolling*, i.e., the task of continuously visiting relevant location of the environment where information have to be gathered *threat response*, that is any specific set of actions needed as a consequence of the detection of a threat. Patrolling can be either passive or active [2]. Passive patrolling is executed without any information from other components of the system, while active patrolling is driven by some specific request for information gathering.

In both cases, robots must be able to move in the environment safely and effectively (so standard robotics modules, like mapping, localization, navigation, obstacle avoidance, etc. are provided to the robots), and to act in a coordinated way, by taking into account dynamic task assignment (for example, Which location has to be visited by each robot), as well as action synchronization Coordination techniques for multi-robot patrolling are described in that include also an extensive experimental analysis showing that on-line coordinated behaviour, in contrast with predefined off-line strategies, are fundamental for actual deployment of surveillance robots. As mentioned, the second issue of surveillance robots is their way of interacting with human operators.

## III. PROPOSED MULTIMODEL SYSTEM

- VOICE RECOGNITION

In Voice recognition system, Speech technologies allow computers equipped with a source of sound input, such as a microphone, to interpret human speech. This input can be classified as: uninterrupted recognition and interpreted recognition. Interpreted recognition has transcription, control, mix of both. Uninterrupted recognition is basically just recording of the sound .Sinc it does not involve trying to understand what's in the waveform, there's no need for further processing. Transcription is the dictation of words into a text editor.

The speech recognition library used is *'SpeechLib'* and interface used for speech recognition is *"SpSharedRecoContext"*. The grammar is defined by using the ISpeechRecoGrammar interface in the SpeechLib. The interface between GUI and Robot is wireless. Commands implemented are: *'Start', 'Break', 'Left', 'Right', 'Pick', 'Release'*. The COM port for receiving the voice commands is detected and we need to select the appropriate port.

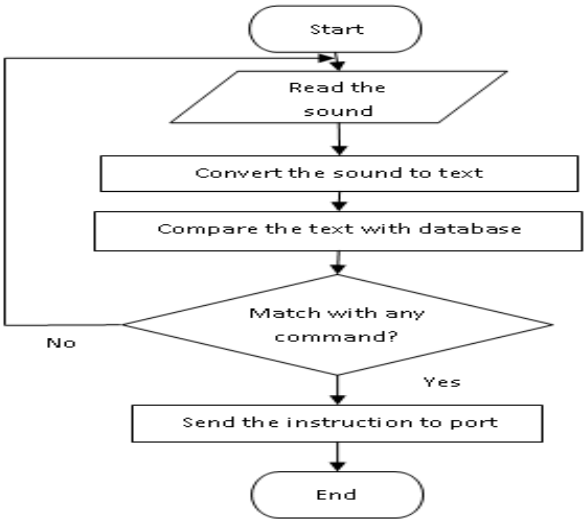



Fig. Flow diagram of the robot controlling program

- GESTURE RECOGNITION

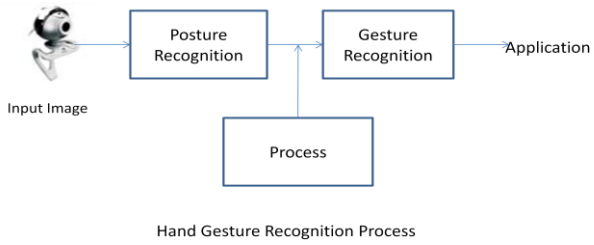
Gesture recognition is the process of recognizing and interpreting a stream continuous sequential gesture from the given set of input data. Hand gesture recognition system can be used for interfacing between computer and human using hand gesture. Total image level is divided into two classes one is hand and other is background. A dynamic gesture is intended to change over a period of time whereas a static

gesture is observed at the spurt of time. A waving hand means goodbye is an example of dynamic gesture and the stop sign is an example of static gesture. In vision based hand gesture recognition system, the movement of the hand is recorded by video camera.

This input video is decomposed into a set of features taking individual frames into account. Some form of filtering may also be performed on the frames to remove the unnecessary data, and highlight necessary components. For example, the hands are isolated from other body parts as well as other background objects. The isolated hands are recognized for different postures. Since, gestures are nothing but a sequence of hand postures connected by continuous motions, a recognizer can be trained against a possible grammar. With this, hand gestures can be specified as building up out of a group of hand postures in various ways of composition, just as phrases are building up by words.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 2, Issue 3, March 2013*




Hand Gesture Recognition Process

Consider a robot navigation problem, in which a robot responds to the hand pose signs given by a human, visually observed by the robot through a camera. We are interested in an algorithm that enables the robot to identify a hand pose sign in the input image, as one of five possible commands (or counts). The identified command will then be used as a control input for the robot to perform a certain action or execute a certain task. For examples see Figure 1. The signs could be associated with various meanings depending on the function of the robot. For example, a *"one"* count could mean *"Start"*, a *"two"* count could mean *"stop"*. Furthermore, *"three"* and *"four"* counts could be interpreted as, *"turn right"*, and *"turn left"*.

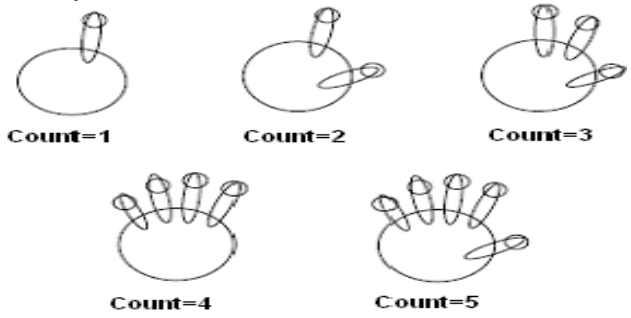



*Figure 1*: Set of hand gestures, or "counts" considered in our work [6]. Image acquisition setup:

It consists of a video camera, web camera, or an analogue camera with suitable interface for connecting it to processor. Digital camera captures the motion of human arm and sends these images to processor. It generate image in various format like RGB, GRAY, YUY.

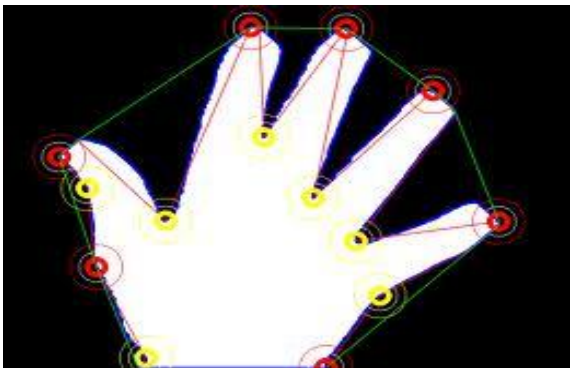
Processor: It consists of either a personal computer or a dedicated image processing unit. It convert image into binary form which allow it to process further. As RGB image have many parameters it is convenient to process image in gray format.

Skin colour detection - Skin can be easily detected by using the colour information. First, we use the constraint, i.e. R.G.B, to find the skin colour regions which may include a wide range of colours, such as red, pink, brown, and orange colour. Therefore, we will find many regions other than the skin regions. However, those non-skin regions satisfy our constraint will be excluded due to there is no motion information. Second, we may obtain some sample colours from the hand region. To find the skin regions, we compare the colours in the regions with the pre-stored sample colour. If they are similar, then the region must be skin region. The hand region is obtained by the hand tracking process in the previous frame. Fig. 3 shows our skin detection results. The rectangular region is the hand region in the previous frame. Finally, we may eliminate some skin similar colours, e.g. the orange colour.


Fig.3. (a) The origin frame, (b) extracted skin regions satisfying R.G.B, and (c) compare the colours of the extracted skin regions with the sample skin colour [4].

A morphological filtering method is used to effectively remove background and object noise in the segmented image. Morphological filtering techniques are used to remove noises from images so that we can get a smooth contour. A contour tracking algorithm is applied to track the contour in clockwise direction. Contour of a gesture is represented by a Localized Contour Sequence (L.C.S) whose samples are the perpendicular distances between the contour pixels and the chord connecting the end-points of a window centred on the contour pixels. After edge detection we get a boundary of hand in image that is our contour of hand image .now a algorithm is applied on the contour to track it in clockwise direction and the contour pixel are numbered sequentially [1],

[5].first we ran a search in image to find a topmost nonzero i.e., contour pixel then numbered the contour in sequential order in clockwise direction from that point.



Machine control: After making the conclusion, mechanical action is to be taken e.g. using serial or

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 2, Issue 3, March 2013*

parallel port of a PC to control left and right motors of a robot to direct it.

Motor Driver (L293D): The L293D is a high voltage, high current four channel driver designed to accept standard TTL logic levels and drive inductive loads and switching power transistors. Motor driver IC gives desired voltage to the robotic arm to ensure the proper motion.

## IV. APPLICATION AREAS

In this project, the focus is identify monitoring the objects in the industries, companies, schools, colleges, public sector and mostly used in defence for spying enemy and protect from obstacles.

- The main objective for developing this application is that, it can provide the user with security of data.
- Only the authorized user and administrator can access the application.
- A person from a remote place can comfortably control the motion of robotic arm by using voice & gesture recognition
- In an Industrial Area where the worker can't handle the harmful equipments.
- In an Industrial Area to spy on the workers through the camera.
- In Houses for paralysis persons to identify the object and handle them.
- To identify the unknown object.
- In mining industries.

## VI. CONCLUSION

In this paper, we have implemented the robotic arm that could be controlled by voice and gesture commands. It can understand any human voice; it is not single speaker dependent. . But it is highly sensitive to the surrounding noises. We also proposed a simple algorithm for hand gesture recognition. Given observed images of the hand, the algorithm segments the hand region, and then makes an inference on the activity of the fingers involved in the gesture. Based on our motivating robot control application, we have only considered a limited number of gestures. Robotic arm is wirelessly controlled. So that user doesn't need to physically appear in the working area, it can be used in. In future we have a plan to build it commercially for mankind. It is helpful for all of the robotic research and project based work.

## REFERENCES

[1] V.I. Pavlovic, R. Sharma, T.S. Huang, Visual interpretation of hand gestures for human-computer interaction, A Review, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 677–695.

[2] T. Baudel, M. Baudouin-Lafon, Charade: remote control of objects using free-hand gestures, Comm. ACM 36 (7) (1993) 28–35.

[3] D.J. Sturman, D. Zeltzer, A survey of glove-based input, IEEE Computer Graphics and Applications 14 (1994) .

[4] T. Takahashi, F. Kishino, A hand gesture recognition method and its application, Systems and Computers in Japan 23 (3) (1992) 38–48.

[5] A.F. Bobick, A.D. Wilson, A state-based technique for the summarization and recognition of gesture, Proceedings fifth international conference on computer vision (1995) 382–388.

[6] L. Bretzner, I. Laptev, and T. Lindberg, "Hand Gesture Recognition using Multi-Scale Colour Features, Hierarchical Models and Particle Filtering", IEEE International Conf. on Automatic Face and Gesture Recognition, 2002.

[7] "Navigation for an Intelligent Mobile Robot", IEEE Journal of Robotics and Automation. Vol1 (1) March 1998.

[8] Nilsson, N. J. Principles of Artificial Intelligence, Tioga Press, 2005.

[9] Survey on Contemporary Remote Surveillance Systems for public Safety. IEEE Transactions on Systems, Man and Cybernetics, 2007.

[10] Integrated person tracking and activity recognition with plan-view templates from a single stereo camera. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2009.

[11] Collaborative Multi-Robot Exploration. In Proc. IEEE International Conference on Robotics and Automation. 476-481. San Francisco. USA. California. IEEE International Conference on Robotics and Automation 2010.

[12] Joris de Ruiter, Natural Language Interaction - the understanding computer, 2010 [Online], Available: www.few.vu.nl/~jdruiter/published_work/Natural_language_interaction.pdf

## BIOGRAPHY

**Aishwarya Patil** is final year student of Department of Computer Technology, Yeshwantrao chavan college of engineering Nagpur, NagpurUniversity (Maharashtra) India.

**Anilkumar Daharwal** is final year student of Department of Computer Technology, Yeshwantrao chavan college of engineering Nagpur, Nagpur University (Maharashtra) India.

**Ankur Harshe** is final year student of Department of Computer Technology, Yeshwantrao chavan college of engineering Nagpur, Nagpur University (Maharashtra) India.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 2, Issue 3, March 2013*

**Mohnish Gakare** is final year student of Department of Computer Technology, Yeshwantrao chavan college of engineering Nagpur, Nagpur University (Maharashtra) India.

**Monika Sajjanwar** is final year student of Department of Computer Technology, Yeshwantrao chavan college of engineering Nagpur ,Nagpur University (Maharashtra) India.