

SPEECH RECOGNITION BASED MOTION TRACKER

Project submitted in partial fulfillment of requirements
For the Degree of

BACHELOR OF ENGINEERING

by

**PRASHANTH NAYAK
DIPAI MODI
RAJU PATEL**

Under the guidance of
Internal Guide

Prof. Kiran Talele



Department of Electronics Engineering
Sardar Patel Institute of Technology
University of Mumbai
2010-2011

BHARTIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
MUNSHI NAGAR, ANDHERI (W),
MUMBAI - 400 058.
2010-11

CERTIFICATE OF APPROVAL

This is to certify that the following students

PRASHANTH NAYAK
DIPAL MODI
RAJU PATEL

have satisfactorily carried out the project work entitled

“SPEECH RECOGNITION BASED MOTION TRACKER”

towards the fulfillment of Bachelor of Engineering course in Electronics of
the Mumbai University

Prof. K.T. Talele
INTERNAL GUIDE

Prof. K. T. Talele
H. O. D. ELECTRONICS

Dr. Prachi Gharpure
PRINCIPAL

**UNIVERSITY OF MUMBAI
2010-2011**

PROJECT APPROVAL CERTIFICATE

This is to certify that the project entitled
“SPEECH RECOGNITION BASED MOTION TRACKER”
has been duly completed by the following students

PRASHANTH NAYAK
DIPAL MODI
RAJU PATEL

Under the guidance of Prof. Kiran Talele in the partial fulfillment of the requirement for the award of degree of Bachelors of Engineering (Electronics) of the university of Mumbai

INTERNAL EXAMINER

EXTERNAL EXAMINER

**DEPARTMENT OF ELECTRONICS ENGINEERING
SARDAR PATEL INSTITUTE OF TECHNOLOGY
MUNSHI NAGAR, ANDHERI (W), MUMBAI-400058**

Acknowledgements

It gives us immense pleasure to present this report of our project conducted as a part of the Bachelor of Engineering course as per the guidelines laid down by the University of Mumbai. We would like to take this opportunity to share a few words of admiration and gratitude for all those who have contributed towards making this project a successful venture.

We express our sincere gratitude to our Head Prof. K.T. Talele for his advice, and guidance from the very early stage of this work as well as giving us extraordinary experiences throughout the work which made her a backbone of this project . Above all and the most needed he always gives encouragement and motivation to do this type of work. Without his moral support it was not possible to complete work.

We gratefully acknowledge Mr Nirmal for his advice, crucial contribution. and his actual involvement with his originality has triggered and nourished my intellectual maturity that we will benefit for a long time , we are grateful in every possible way and hope to keep up our collaboration in the future.

Their invaluable guidance and encouragement motivated us to work harder towards accomplishing this project.

Contents

1	Project Objective	1
2	Introduction	3
3	Literature Survey	5
3.1	Speech Recognition	5
3.1.1	Weighted Distortion Measure on Standard Deviation for VQ-Based Speaker Identification	5
3.1.2	New Feature Extraction Methods Using DWT and LPC for Isolated Word Recognition	6
3.2	Object tracking	7
3.2.1	Design and implementation of Real-Time Object Tracking System Using the Gaussian Motion Model and the Otsu Algorithm	7
3.2.2	An occlusion Tolerant Method For Multi Object Tracking	8
4	Speech Recognition	12
4.1	Speech Processing	12
4.1.1	What is Speech?	12
4.1.2	Technical Characteristics of the Speech Signal:	12
4.1.3	Bandwidth	12
4.2	Speech Recognition by Linear Prediction Model	12
4.3	Applications	15
4.4	Issues and Challenges of Speech Recognition	15
5	Motion Detection	16
5.1	Introduction	16
5.2	Object Tracking	16
5.2.1	Image Acquisition	16
5.2.2	Background accumulation	17
5.2.3	Frame capturing	17
5.2.4	Isolating The Image	17
5.2.5	Detecting Moving Object	17
5.2.6	Counting number of Pixel	18
6	Vehicle Control System Design and Overview	19
6.1	Introduction	19
6.2	Vehicle Control Systems	19
6.2.1	Problem definition stage	19

6.2.2	Designing block diagram	19
6.2.3	Implementing circuits and components	19
6.2.4	Developing Flowchart for software	20
6.2.5	Implementing software code	20
6.2.6	Implementing software code	21
6.3	Burning the hex file into microcontroller with Programmer:-	21
6.3.1	Testing	21
6.3.2	Running	21
7	Experimental Results	22
7.1	Speech Recognition	22
7.2	Motion Detection and tracking	23
8	Applications	24
8.1	Application	24
8.2	Advantages	24
8.3	Neccesity	24
8.3.1	Surveillance	24
8.3.2	Military	25
8.3.3	Search and Rescue	25
9	Conclusion and Future Scope	26
9.1	Conclusion	26
9.2	Future Work	26
9.3	Issues	26

List of Figures

2.1	Block Diagram of Speech operated Motion Tracker	4
3.1	VQ-based speech recognition system	5
3.2	DWLPC (Dyadic Wavelet decomposed LPC) feature extraction system . . .	6
3.3	UWLPC (Uniform Wavelet decomposed LPC)	7
4.1	Speech Recognition System	13
4.2	Linear Predictive coding	14
5.1	Discrete Wavelet Transform	17
5.2	Isolation	17
5.3	Background Substraction	18
5.4	Background Substraction	18
6.1	Motion Tracker	20
7.1	LPC Test Results	22

Abstract

The objective of the project is to design a speech operated motion tracker for detection of motion. Our tracker is an essential requirement of any surveillance system. Nowadays CCTV cameras are used for security purposes. These cameras can even track the object that is in motion but as they are fixed at a particular place. In this work we present an operational computer vision system for real-time detection and tracking of object in motion. The system captures video of a scene and identifies moving objects using wavelet segmentations, operated by Speaker through speech. We have implemented speech recognition using Wavelet Decomposed LPC which gives 80 percent accuracy. Using Wavelet segmentation we were able to detect motion even in varying brightness, this gives an advantage of tracking motion in the regions of low brightness. Motion tracker can be used in surveillance areas where the CCTV cameras fail, Motion tracker will be able to follow the object until a clear image of the object is retrieved.

Chapter 1

Project Objective

Motion tracking is an important task within the field of computer vision. The proliferation of high-powered computers, the availability of high quality and inexpensive video cameras, and the increasing need for automated video analysis has generated a great deal of interest in Motion tracking algorithms. There are three key steps in video analysis: detection of interesting moving objects, tracking of such objects from frame to frame, and analysis of object tracks to recognize their Motion behavior. Therefore, the use of object tracking is pertinent in the tasks of motion-based recognition, that is, human identification based on gait, automatic object detection, etc, Automated surveillance, that is, monitoring a scene to detect suspicious activities or unlikely events, video indexing, that is, automatic annotation and retrieval of the videos in multimedia databases, humancomputer interaction, that is, gesture recognition, eye gaze tracking for data input to computers, etc, traffic monitoring, that is, real-time gathering of traffic statistics to direct traffic flow. Vehicle navigation that is, video-based path planning and obstacle avoidance capabilities.

In its simplest form, tracking can be defined as the problem of estimating the trajectory of an object in the image plane as it moves around a scene. In other words, a tracker assigns consistent labels to the tracked objects in different frames of a video additionally, depending on the tracking domain, a tracker can also provide object-centric Information, such as orientation, area, or shape of an object. Tracking objects can be complex due to: loss of information caused by projection of the 3D world on a 2D image, noise in images, complex object motion, nonrigid or articulated nature of objects, partial and full object occlusions, complex object shapes, scene illumination changes, and Realtime processing requirements. One can simplify tracking by imposing constraints on the motion and/or appearance of objects. For example, almost all tracking algorithms assume that the object motion is smooth with no abrupt changes. One can further constrain the object motion to be of constant velocity or constant acceleration based on a priori information. Prior knowledge about the number and the size of objects, or the object appearance and shape, can also be used to simplify the problem. Numerous approaches for object tracking have been proposed. These primarily differ from each other based on the way they approach the following questions: Which object representation is suitable for tracking? Which image features should be used? How should the motion, appearance, and shape of the object be modeled? The answers to these questions depend on the context/environment in which the tracking is performed and the end use for which the tracking information is being sought. A large number of tracking methods have been proposed which attempt to answer these questions for a variety of sce-

narios. Our project emphasizes on measurement of different body parameters. The objective of our project is to assemble a Biomedical System to measure the following parameters:

Chapter 2

Introduction

As the security threats are increasing day by day. The need for smarter and more advanced surveillance machines has increased. Today CCTV cameras are used for security. Many CCTV cameras provide motion detection facility. These cameras can even track the object that is in motion but as they are fixed at a particular place they cannot track an object or target after he crosses the cameras range. This is a serious problem, many a times we are unable to identify the intruder using CCTV cameras. So a need has risen in the surveillance industry for a better surveillance system. Which can detect motion or intruder, should be able to follow the intruder where ever he goes and moreover to recognize the intruder making the security system more reliable and robust.

The problem can be solved by our Motion Tracker .The robot is equipped with a wireless camera. The robot is semi autonomous and is able to detect motion and to track the target in motion. The robot is wireless, battery powered so it can follow the target where ever it goes. With the help of Motion Tracker we can detect if there is an intrusion or not then track the intruder and we can find out the intruders hide out, so without wasting time we can catch the intruder.

We have classified our Motion Tracker in four stages Speech Recognition, Motion Detection, and Vehicle Control. The robot responds to voice commands hence Speech Recognition stage is required, Motion Detection stage is required to detect the motion extract the target and give commands to the vehicle, Target Recognition stage is used to recognize the extracted target and thus inform the user about it, Vehicle control stage is used to control the robots motion (i.e. motors).

Motion Tracker can be widely used in surveillance systems to detect motion and sound alarm and to follow the thief (Motion) till the appropriate person is alarmed. Used in military or defense for spying on enemies and to search for motion on border. The Motion Tracker could be used to find any injured persons in calamity regions. This is possible in Motion Tracker because it uses Target Recognition stage, so that it could be programmed to search for the particular image.

Motion Tracker is a vehicle or a semi-autonomous robot. The robot must be able to detect motion and should be able to track the target which is in motion. The robot is equipped with a wireless camera through which it scans the area. If the motion is detected the robot

should recognize the target and inform the user about the target. The robot then should track the target as commanded by the user. As per figure 1 the Motion Tracker can be divided in three stages as Speech Recognition, Motion Detection and vehicle control.

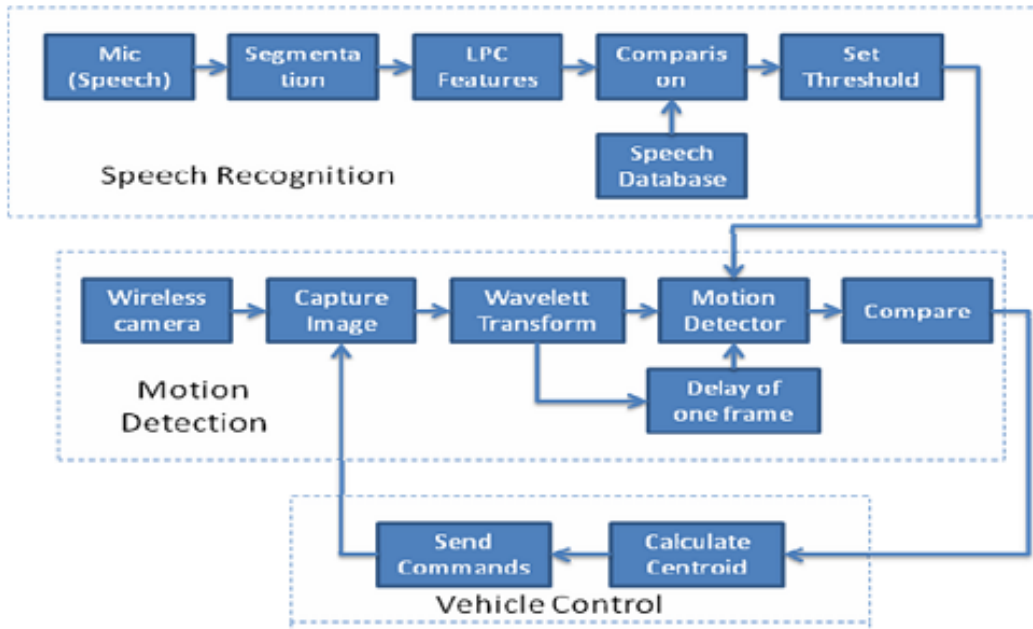


Figure 2.1: Block Diagram of Speech operated Motion Tracker

Chapter 3

Literature Survey

3.1 Speech Recognition

3.1.1 Weighted Distortion Measure on Standard Deviation for VQ-Based Speaker Identification

Speaker recognition is the technology of identification and authentication based on Speech.VQ algorithm is preferred over speech recognition because of its real time performance.

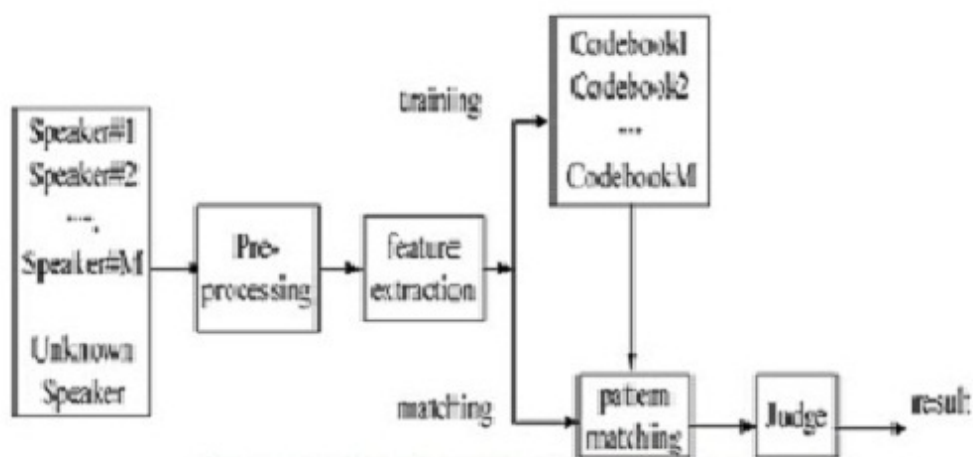


Fig. 1 VQ-based speaker identification system

Figure 3.1: VQ-based speech recognition system

In VQ-based speaker recognition, two roles are played in VQ algorithm . which is used both in the training and matching phases. In the training phase, the speaker models code books are constructed by clustering the feature vectors in K separate clusters, Each cluster is represented by a code vector, which is the center (average vector) of the cluster. The resulting sets of code vectors are called the speaker codebook. The choice of the original code book

has a great effect on the optimum codebook design. In order to avoid slowly convergence or non-convergence to optimal results we select diakoptic algorithm. The capacity of code book should be set reasonably since a lot of experiments have been proven that as the capacity of codebook increases, the identification performance are also improved. For the clustering of speech feature vectors, we takes the classic LBG.

3.1.2 New Feature Extraction Methods Using DWT and LPC for Isolated Word Recognition

In this paper a new feature extraction methods which utilize reduce order linear predictive coding (LPC) coefficients for speech recognition has been proposed .The coefficient are derived using DWT. The signal is decomposed into subbands by using DWT .LPC features have been estimated from this DWT.LPC coefficient obtained from each subband are concatenated to form a final feature vector .For modeling HMM recognizer is used. In speech signal, high frequencies are present very briefly at the onset of a sound while lower frequencies are present latter for long period . DWT resolve all these frequencies well. The DWT parameters contain the information of different frequency scales. This helps in getting the speech information of corresponding frequency band. We thus introduce a DWT to decompose speech signal into the frequency bands. In order to parameterize the speech signal, we first decompose the signal into four frequency bands uniformly or in dyadic fashion. Most of the information in the human speech is only contained in a few scales of Wavelet Transform (WT) decomposition. Hence a few (2-3) number of WT scales help to

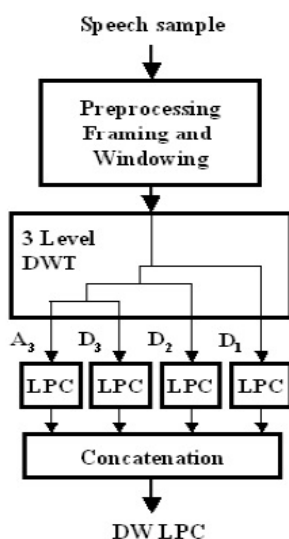


Figure 3.2: DWLPC (Dyadic Wavelet decomposed LPC) feature extraction system

reduce the final feature vector dimensions. Figure shows the block diagram of proposed DWLPC (Dyadic Wavelet decomposed LPC) feature extraction system. Three level DWT decomposition of preprocessed and windowed speech has been done using Daubechies's wavelet filters. Actual wavelet coefficients retain the time information hence LPC features have been estimated from the DWT coefficients in time domain.

$$f_i = [a_{A_3} \ a_{D_3} \ a_{D_2} \ a_{D_1}]^t .$$

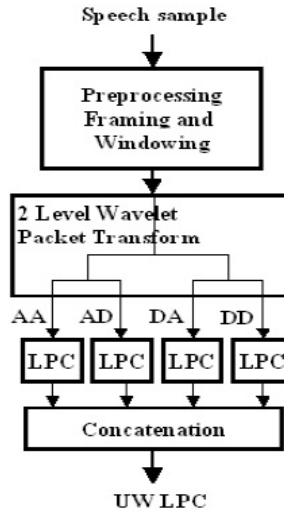


Figure 3.3: UWLPC (Uniform Wavelet decomposed LPC)

LPC features of p th order have been extracted from each subband of wavelet decomposed speech signal. The LPC coefficients obtained from each subband are concatenated to form a final feature vector. Thus the feature vector f derived from frame i can be expressed as where, A_3 is a row vector formed using prediction coefficients obtained from 3 A and D_j is row vector formed using prediction coefficients obtained from j D ($j = 1,2,3$). T indicates a vector transpose. UWLPC (Uniform Wavelet decomposed LPC) features are derived from subbands of uniform bandwidth. The subbands are obtained by two level wavelet packet decomposition as shown in Figure . The feature vector is formed similar to DWLPC.

3.2 Object tracking

3.2.1 Design and implementation of Real-Time Object Tracking System Using the Gaussian Motion Model and the Otsu Algorithm

In this paper a moving object tracking algorithm has been proposed by using background subtraction ,guass motion,and ostu algorithm.

The steps are as follows

- Preprocessing - This is done to reduce the impact of dark and bright environment for fast moving objects.The algorithm use is 'White balance algorithm by LAM'.
- YCbCr colour space conversion-The target object is red switching to YCbCr color space, the red part will be highlighted.
- Detecting moving object-background is updated at regular interval. Adaptive threshold is calculated using threshold using Guassian model.
- Binarization-OSTU algorithm is used for segmentation.Mobile object is separated out.

3.2.2 An occlusion Tolerant Method For Multi Object Tracking

This paper mainly given an approach for tracking occluded vehicles in real road traffic or any other such real time applications. They have used third level DWT. They performed following steps.

1. Used 2-dimentional bounding box to represent vehicle.
2. Used Kalman filter and weighted scale parameter to predict center position and motion.
3. Then gave feedback, the center of new vehicle region to Kalman filter to modify Kalman gain and estimate the optimum position.

In this paper they have gone through certain steps which are as follows:

Motion Detection

1. By Discrete Wavelet Transform
 2. By Background Difference
 3. Bounding Box Of Moving Objects
1. Discrete Wavelet Transform Here an image is 1st decomposed into $3J+1$ subbands,

$$\{LL_K, [LH_K, HL_K, HH_K]_{K=1,2,3,\dots,j}\}$$

Where, J is level of compression

LL is the k-th low frequency sub image. Low frequency component in vertical horizontal direction.

LH is the k-th Low frequency component in horizontal Direction high frequency component in vertical direction.

HL is the k-th High frequency component in horizontal direction Low frequency component in vertical direction.

HH is the k-th High frequency component in horizontal direction high frequency component in vertical direction.

By using this we can detect mobile objects using them with decreased comuting cost image noise. Here in this paper have adopted three-level decompition and LL3(Low frequency part).

2. Background differences

This method can be used for detecting motion and can gain more intact mobile objects than other methods. It subtracts background from current frame assuming that background difference will be less than a predefined threshold when some pixel belongs to a background.

Here in this paper they have reconstructed background image calculating maximum probability gray scale at each pixel and selective backgroundupdate is then adopted. Here 1st some threshold value like T is defined and after it each pixel's value is thresholded by that threshold value T using following formula,

3. Bounding Box Of Moving Objects

$$f(x,y,t) = \begin{cases} 1 & \text{if } |LL_1(x,y,t) - LL_{\text{back}}(x,y,t)| \geq T \\ 0 & \text{otherwise} \end{cases}$$

Where, $f(x,y,t)$ is difference image.
 x,y are image co-ordinates.

$LL_1(x,y,t)$ is low frequency sub-image of third level
DWT of current frame.

$LL_{\text{back}}(x,y,t)$ is low frequency sub-image of third level
DWT of background image.

Here A 2-dimentional boundary box is used which is used to represent vehicle, which is determined computing the maximum and minimum value of x and y co-ordinates of connected components in $f(x,y,t)$ and then it is transformed to original resolution according to interband special relationship of DWT.

If $\text{Object}^i(t)$ denotes set of co-ordinates of points in moving object i,
 $O_{\text{min}}^i(t)$ is the left – top corner co – ordinates of moving object i, and
 $O_{\text{max}}^i(t)$ is the right – bottom corner co – ordinates of moving object i.
 $(x^i(t), y^i(t))$ be center position of boundary box,
Then,

$$O_{\text{min}}^i(t) = \{x_{\text{min}}^i(t), y_{\text{min}}^i(t) | x, y \text{Object}^i(t)\}$$

$$O_{\text{max}}^i(t) = \{x_{\text{max}}^i(t), y_{\text{max}}^i(t) | x, y \text{Object}^i(t)\}$$

$$x^i(t) = [x_{\text{min}}^i(t) + x_{\text{max}}^i(t)] * 0.5$$

$$y^i(t) = [y_{\text{min}}^i(t) + y_{\text{max}}^i(t)] * 0.5$$

II. OCCLUSION REASONING AND TRACKING ALGORITHM

1. Occlusion
 2. Scale parameter $S(t)$
 3. Estimating and Adjusting boundary edges of Box.
1. Occlusion

When multiple objects exist and they are very close to each other or very close next to each other, occlusion or grouping of those objects is routine event. This causes main problem in tracking the object. This occlusion reasoning is as follows,

- Estimating the center position for each boundary box in next frame by,

$$X(t|t-1) = AX(t-1|t-1).$$

- : If more than one center shoot same red dashed box, the grouping or occlusion occurs

2. Scale parameter $S(t)$

It is used to estimate box change in dimentions which is caused by motion component along optical axis away or towards camera. If C is a detected area for frame t , where binary image $f(x,y,z)$ and $p+q$ order geometrical quadrature $M_{pq}(t)$ can be expressed respectively by fol

$$M_{pq}(t) = \iint x^p y^q f(x, y, t) dx dy = \int_{(x,y) \in A} x^p y^q dx dy$$

3. Estimating and Adjusting boundary edges of Box.

- : Sort the estimated box centers in tracking list by their x-co-ordinate and y-co-ordinate
- Judge observable of boundaries edges and observable boundaries edges

If

$$x1 < x2$$

Left boundary of car1 and right boundary of car2 are observable. So the green box of car1 is moved left to overlap the left boundaries edge of red dashed box and the green box of car2 is moved left to overlap the right bounding edge of red dashed box, and vice versa.

If

$$y1 < y2$$

The top bounding edge of car1 and the bottom bounding edge of car2 are observable, so the green box of car1 is moved up to overlap the top bounding edge of red dashed box and green box of car2 is moved up to overlap the bottom bounding edge of red dashed box, and vice versa. The blue boxes are the adjusted results

- Judge unobservability of bounding edges(i.e. the right and bottom bounding edges of car1 and the left and top bounding edges of car2).

If the number of moving pixels n on an unobservable bounding edge is more than a threshold, this bounding edge is moved to increase the area of box.

If the number of moving pixels is less than a threshold, this bounding edge is moved to decrease the area of box, or else stop. The peachy boxes are the finally adjusted results, which will be regarded as $Z(t)$ to input into estimate.

Chapter 4

Speech Recognition

4.1 Speech Processing

Speech processing is always looked upon as a fascinating field in human computer interaction. It is one of the fundamental steps towards understanding human cognition and their behavior. This report explicates the theory and implementation of Automatic speech reorganization, which is a speaker-dependent real time isolated word recognizer.

4.1.1 What is Speech?

Speech is the vocalized form of human communication. A person making a speech is called a speaker. . Each spoken word is created out of the phonetic combination of a limited set of vowel and consonant speech sound units. . The vocal abilities that enable humans to produce speech also provide humans with the ability to sing, cry, laugh, shout etc.

4.1.2 Technical Characteristics of the Speech Signal:

The bandwidth of the signal is 4 kHz. The signal is periodic with a fundamental frequency between 80 Hz and 350 Hz There are peaks in the spectral distribution of energy at $(2n - 1) * 500$ Hz ; $n = 1, 2, 3, . . .$ The envelope of the power spectrum of the signal shows a decrease with increasing frequency (-6dB per octave)

4.1.3 Bandwidth

The bandwidth of the speech signal is much higher than the 4 kHz stated above. In fact, for the fricatives, there is still a significant amount of energy in the spectrum for high and even ultrasonic frequencies. However, as we all know from using the (analog) phone, it seems that within a bandwidth of 4 kHz the speech signal contains all the information necessary to understand a human voice.

4.2 Speech Recognition by Linear Prediction Model

Speech recognition (also known as automatic speech recognition or computer speech recognition) converts spoken words to machine-readable input (for example, to key presses, using the binary code for a string of character codes). The term "voice recognition" is sometimes

used to refer to speech recognition where the recognition system is trained to a particular speaker - as is the case for most desktop recognition software; hence there is an aspect of speaker recognition, which attempts to identify the person speaking, to better recognize what is being said. Speech recognition is a broad term which means it can recognize almost anybody's speech - such as a call centre system designed to recognize many voices. Voice recognition is a system trained to a particular user, where it recognizes their speech based on their unique vocal sound.

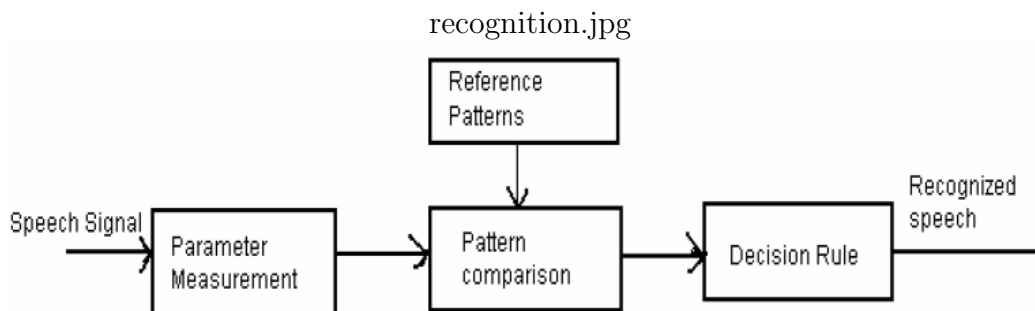


Figure 4.1: Speech Recognition System

The above figure shows a schematic diagram of a speech recognition system. First the speech signal is sampled and various parameters, which can be assumed to be characteristics of the speech, are measured. Then they are compared with reference pattern and the output is given to decision rule which decides whether the required phoneme is present or not.

To identify words, we use LPC (Linear Predictive Coding) which is a popular method of extracting speech characteristics from sample values. The linear predictive coding (LPC) method of speech analysis and synthesis is based on modeling the vocal tract a linear all-pole (IIR) filter having the system function

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - a_1z^{-1} - a_2z^{-2} + \dots + a_pz^{-p}}$$

where p is the number of poles, $a(k)$ are the parameters that determine the poles.

Word Capturing The signals coming from the microphone is processed only when you speak something. The program waits until the sample value exceeds some threshold value (which can be adjusted by the user). When the program is triggered by a significant sample, a number of following samples are captured to process. After that to determine the actual boundaries of the word spoken, edge detection is performed. Here the center of gravity of the energy distribution of the signal is calculated and then from that point intervals where the amplitude level lies below a threshold level are removed. Finally we can have a set of voice samples corresponding to a particular word free of silent periods

LPC Processing The steps we followed for the extraction of speech characteristics from captured samples using LPC is described in the following block diagram.

1. Pre-emphasis:

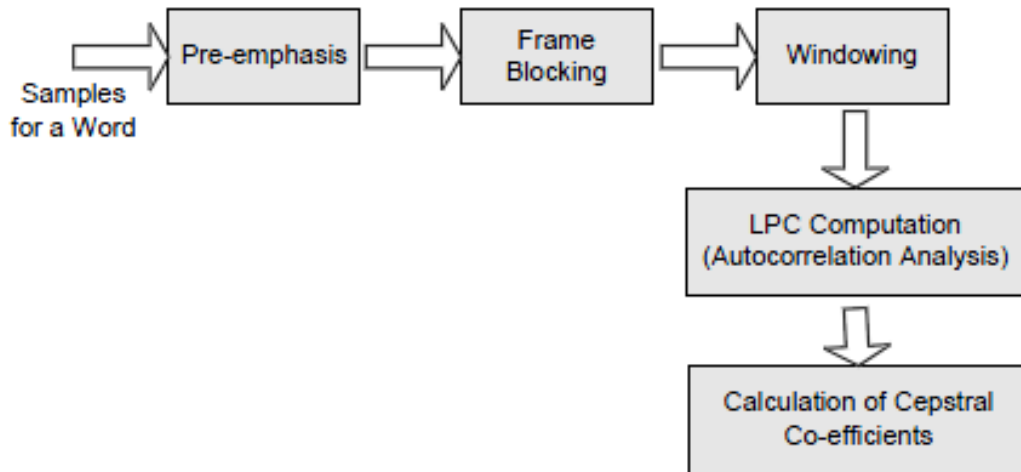


Figure 4.2: Linear Predictive coding

This operation is necessary for removing DC and low frequency components of the incoming speech signal. It also makes the signal spectrum flatter. Pre-emphasis is done using a first order FIR filter which can be described by the transfer function,

$$H(z) = 1 - az^{-1}$$

Here we used $a = 0.9$. FIR filtering was applied to the signal in the time domain

2. Frame Blocking:

Each signal is now converted into a set of fixed length frames, with some number of samples in each frame is overlapping. If the frame length is L and each frame is shifted by M samples away from the adjacent frame, then n th frame can be denoted by,

$$x[i] = s[(n - 1) * M + i] \text{ when } n = 1, 2, \dots, N \text{ and } i = 1, 2, \dots, L.$$

3. Windowing:

Each individual frame is windowed to minimize the signal discontinuities at the borders of each frame. We used the Hamming Window for this purpose. The set of samples for each frame is multiplied by the time domain version of the Hamming window with size equal to the frame length.

4. LPC Calculation:

First step of calculating LPC parameters is to get the autocorrelation vector for each frame. If the order of the autocorrelation is P , then the autocorrelation vector, r can be given by,

$$r(m) = \sum_{n=0}^{N-1-m} x(n) * x(n + m)$$

where $m = 0,1,2,..,P$ and $x(i)$ s ($i = 1,2,..,L$) are sample values in the windowed frame. Then Hermitian Toeplitz matrix of r is computed as shown below:

$$\begin{pmatrix} r[0] & \dots & r[P-1] \\ \vdots & \ddots & \vdots \\ r[P-1] & \dots & r[0] \end{pmatrix}$$

Finally the LPC parameter matrix, a is calculated by matrix multiplication of inverse of R and r .

$$a = R^{-1} * r$$

In our software we used order(P) as 10.

4.3 Applications

Speech recognition has many applications In Medical Fields In Military Highperformance fighter aircraft Telephony and other domains Automatic translation Automotive speech recognition (e.g., Ford Sync) Telematics (e.g. vehicle Navigation Systems) Court reporting (Real-time Voice Writing) Hands-free computing: voice command recognition computer user interface Home automation Interactive voice response Mobile telephony, including mobile email Multimodal interaction Pronunciation evaluation in computer-aided language learning applications Robotics Video Games, digital speech-to-text Speech-to-text (transcription of speech into mobile text messages)

4.4 Issues and Challenges of Speech Recognition

First, the obvious: speech is a complex audio signal, made up of a large number of component sound waves. Speech can easily be captured in wave form, transmitted and reproduced by common equipment; this is how the telephone has worked for a century, however, once we move up the complexity scale and try to make a computer understand the message encoded in speech, the actual wave form is unreliable. Vastly different sounds can produce similar wave forms, while a subtle change in inflection can transform a phoneme's wave form into something completely alien. In fact, much of the speech signal is of no value to the recognition process. Worse still: any reasonably accurate mathematical representation of the entire signal would be far too large to manipulate in real time. Therefore, a manageable number of discriminating features must somehow be extracted from the wave before recognition can take place. A common scheme involves "cepstral coefficients" (cepstral is a mangled form of spectral); the recognizer collects 8,000 speech samples per second and extracts a "feature vector" of at most a few dozen numbers from each one, through a mathematical analysis process that is far beyond the scope of this article.

Chapter 5

Motion Detection

5.1 Introduction

In recent years, motion detection has attracted a great interest from computer vision researchers due to its promising applications in many areas, such as video surveillance, traffic monitoring or sign language recognition. However, it is still in its early developmental stage and needs to improve its robustness when applied in a complex environment. Several techniques for moving object detection have been proposed in [1], among them the three representative approaches are temporal differencing, background subtraction and optical flow. Temporal differencing based on frame difference, attempts to detect moving regions by making use of the difference of consecutive frames (two or three) in a video sequence. This method is highly adaptive to dynamic environments, but generally does a poor job of extracting the complete shapes of certain types of moving objects. Background subtraction is the most commonly used approach in presence of still cameras. The principle of this method is to use a model of the background and compare the current image with a reference. In this way the foreground objects present in the scene are detected. The method of statistical model based on the background subtraction is flexible and fast, but the background scene and the camera are required to be stationary when this method is applied. Optical flow is an approximation of the local image motion and specifies how much each image pixel moves between adjacent images. It can achieve success of motion detection in the presence of camera motion or background changing. According to the smoothness constraint, the corresponding points in the two successive frames should not move more than a few pixels. For an uncertain environment, this means that the camera motion or background changing should be relatively small. The method based on optical flow is complex, but it can detect the motion accurately even without knowing the background.

In our work we have applied wavelet transform to current frames to compress the image. Then this frame is subtracted from the background. Number of pixel are calculated and command is sent to the robot.

5.2 Object Tracking

5.2.1 Image Acquisition

For Real time video we have interfaced camera to our system

Install Your Image Acquisition Device

Installing the frame grabber board in your computer. Installing any software drivers required by the device. These are supplied by the device vendor connecting a camera to a connector on the frame grabber board. Verifying that the camera is working properly by running the application software that came with the camera and viewing a live video stream.

5.2.2 Background accumulation

First the background is set by averaging 30 to 300 frame. The background is compressed using discrete wavelet transform. This background is then subtracted from every frame to detect moving object.

5.2.3 Frame capturing

Current frame is captured. DWT is computed. As shown.

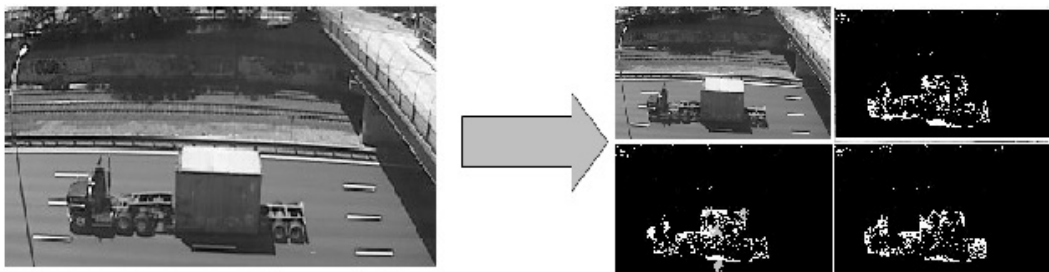


Figure 5.1: Discrete Wavelet Transform

5.2.4 Isolating The Image

Compressed image is isolated from the DWT.

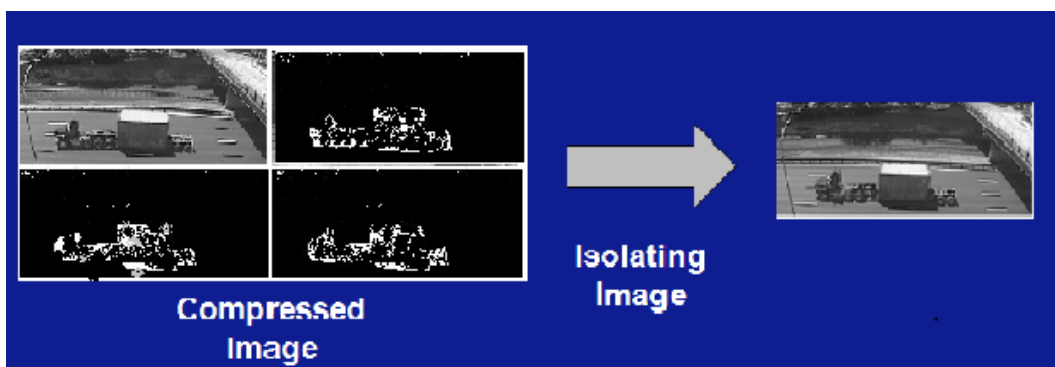


Figure 5.2: Isolation

5.2.5 Detecting Moving Object

Moving object is detected using background subtraction.

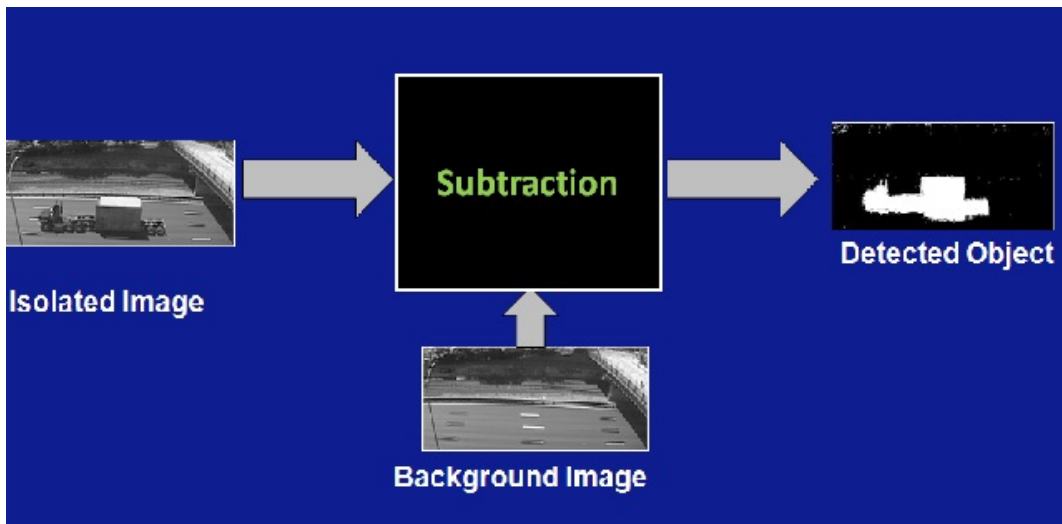


Figure 5.3: Background Substraction

5.2.6 Counting number of Pixel

Number of pixel is counted in different region of the image and command is passed to robot.

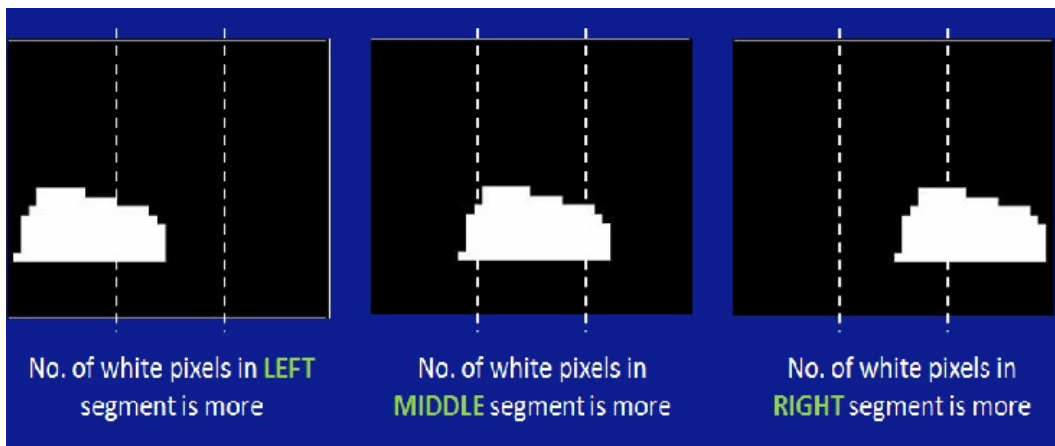


Figure 5.4: Background Substraction

Chapter 6

Vehicle Control System Design and Overview

6.1 Introduction

This is the last and the most important stage. This stage comprises of all the physical components i.e. Motors, Wireless transmitter receiver, Camera. Our Motion Tracker robot is shown below microcontroller board to control the physical parts mentioned above. The Microcontroller is used to take the commands that are received by the receiver and accordingly control the motion of the motors and the motion of the camera. The microcontroller will also be programmed to control the speed of the motors. The vehicle will be differential drive. The vehicle will also be able to climb over obstacles.

6.2 Vehicle Control Systems

For the complete development of above system can be divided into the following stages

6.2.1 Problem definition stage

This is the very first stage to develop any project. It actually defines the aim and the concept of the project. In this section the purpose i.e. why this particular project is required is explained.

6.2.2 Designing block diagram

At this stage we have categorized the whole system into different individual modules. These modules (block diagrams) will be helpful in understanding the concept and working of the integrated system.

6.2.3 Implementing circuits and components

This is the actual implementation of circuit of each block. At this stage we have actually designed each block separately and finally integrated them into the complete working system.

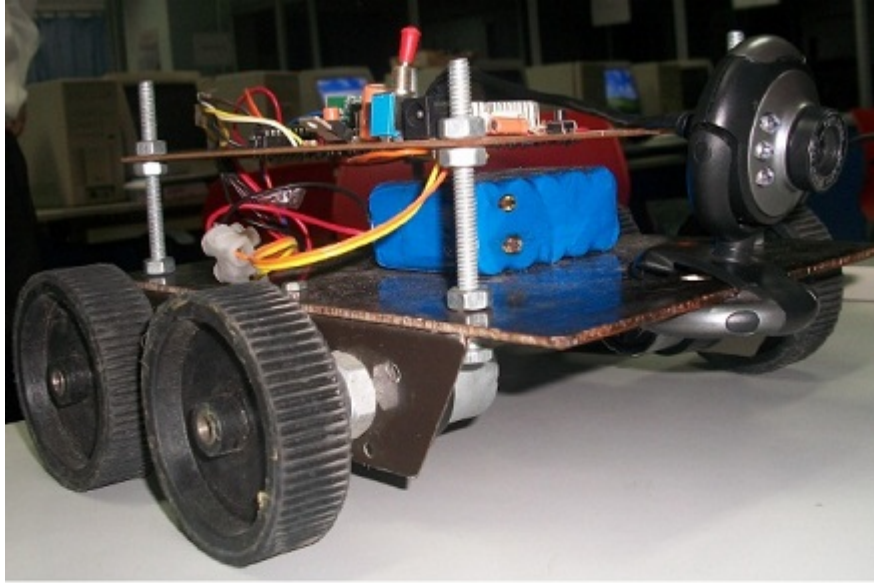


Figure 6.1: Motion Tracker

6.2.4 Developing Flowchart for software

To get the logical flow of the software, the development of flowchart is having a prominent role. So we have to analyze the complete system and organized the flowchart in such a manner that one can understand the complete working of the software.

6.2.5 Implementing software code

After the development of the algorithm and flow chart we implemented it in OpenCV Writing actual code for Microcontroller:- After the development of the algorithm and flowchart we have actually translated them in C language for ATMEGA8535 Microcontroller so that it can understand the instructions and run as per our requirement.

6.2.6 Implementing software code

The hand written code on paper was then transferred into computer. For that we have used Keil pre-install on PC. The Keil is a Computer Aided Program to simulate the working of Microcontroller in real time without burning the software into actual IC. We simulated and compiled our program for error checking. After removing several compiling errors the program was converted into machine language i.e. Intel hex format.

6.3 Burning the hex file into microcontroller with Programme

In this stage the compiled hex format file was downloaded or burned into ATMEL 89s52 Microcontroller.

6.3.1 Testing

This time we tested our project for actual working, after loading the software into the microcontroller. Any errors found were removed successfully.

6.3.2 Running

This is the last and final stage of development of our project. In this stage a user flowchart was made so that anyone can use this system without any difficulty.

Chapter 7

Experimental Results

The speech operated motion tracker is a robot which uses speech to decide the sensitivity of the robot towards motion. There are basically two ranges defined that is minimum and maximum. The minimum range increases the sensitivity of the motion tracker so as to detect the smallest possible motion. The maximum will decrease the sensitivity hence to detect only large motion. The motion tracker is shown below.

7.1 Speech Recognition

The speech recognition was initially done using LPC (Linear Predictive Coding).

Linear Prediction Coefficients: -

Test 1		Test 2		Test 3		Test 4	
Word Spoken	Word Recognized	Word Spoken	Word Recognized	Word Spoken	Word Recognized	Word Spoken	Word Recognized
One	Hello	One	Stop	One	One	One	Hello
Two	Stop	Two	Two	Two	One	Two	One
Start	Hello	Start	Two	Start	Hello	Start	Hello
Stop	Hello	Stop	Two	Stop	Hello	Stop	One
Hello	Hello	Hello	Hello	Hello	One	Hello	Stop

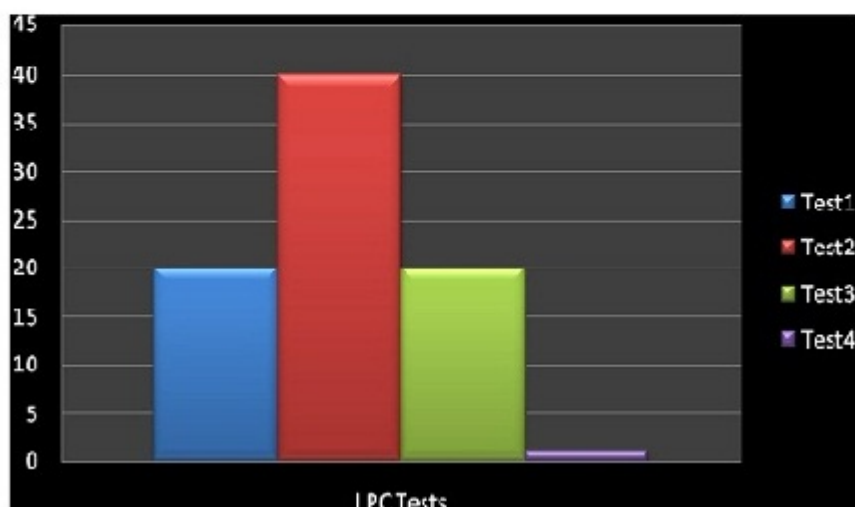


Figure 7.1: LPC Test Results

The results so obtained were less accurate hence we adopted the algorithm of WDLPC (Wavelet Decomposed LPC). The table below shows the comparison between them.

7.2 Motion Detection and tracking

The motion detection part is done using Wavelet compression for faster and better results. The advantage of using Wavelet transform was to remove the brightness variations in the images which caused errors while computing. The wavelet transform is used to compute the edge detection of the frames. The differences of the edge detected frames were taken and then morphological operations were made to improve the results.

Rate of correct detection, the number of objects detected in motion divided by the total number of objects in motion.

Rate of false detection, PFA, the number non motion object detected to the total number of non motion object detected.

Chapter 8

Applications

8.1 Application

- video surveillance
- visual navigation and monitoring
- content-based indexing and retrieval
- object-based coding
- traffic monitoring
- sports analysis for enhanced TV broadcasting
- video post-production
- Video object tracking techniques vary according

8.2 Advantages

- The user knows which object the robot is tracking
- Voice activation, i.e. Security is provided
- The user has the command over the tracking of the robot
- Pre-configurable, i.e. the robot can search and follow a given image (Target recognition stage).
- Can be used as object follower
- Can be used for search purposes

8.3 Neccesity

8.3.1 Surveillance

In surveillance systems to detect motion and sound alarm and to follow the thief (Motion) till the appropriate person is alarmed.

8.3.2 Military

Military or defense for spying on enemies. To search for motion on borders.

8.3.3 Search and Rescue

The Motion Tracker could be used to find any injured persons in calamity regions.

Chapter 9

Conclusion and Future Scope

9.1 Conclusion

Among many different algorithms that are available for speech processing and image processing that we studied and learned, we found few algorithms that suited our project they are Wavelet Transform and LPC for Speech Recognition and Wavelet Transform and Morphological Operators for Image Processing.

The use of LPC alone in speech Recognition gives a less accurate result so we used Wavelet Decomposed LPC to increase our accuracy.

By using wavelet Decomposition we observed that brightness effects on images were reduced. We compute an edge detected image using Wavelet transform which helped us in improving our results

9.2 Future Work

The project Motion Tracker can be further continued to recognize the particular object in motion. The software part can be compiled into the hardware and the robot can be made fully autonomous by using integrated camera and microphone.

9.3 Issues

Significant progress has been made in object tracking during the last few years. Several robust trackers have been developed which can track objects in real time in simple scenarios. However, the tracking problem is tractable, for example, smoothness of motion, minimal amount of occlusion, illumination constancy, high contrast with respect to background, etc., are violated in many realistic scenarios and therefore limit a trackers usefulness in applications like automated surveillance, human computer interaction, video retrieval, traffic monitoring, and vehicle navigation. Thus, tracking and associated problems of feature selection, object representation, dynamic shape, and motion estimation are very active areas of research and new solutions are continuously being proposed. One challenge in tracking is to develop algorithms for tracking objects in unconstrained videos, for example, videos obtained from broadcast news networks or home videos. These videos are noisy, compressed,

unstructured, and typically contain edited clips acquired by moving cameras from multiple views. Another related video domain is of formal and informal meetings. These videos usually contain multiple people in a small field of view. Thus, there is severe occlusion, and people are only partially visible.

One interesting solution in this context we employed audio in addition to video for object tracking. There are some methods being developed for estimating the point of location of audio source, This audio-based localization of the speaker provides additional information which then can be used in conjunction with a video-based tracker to solve problems like severe occlusion

The following issues were faced during our testing of the project.

1. Occlusion
2. Processing speed
3. The robot movement speed
4. Multiple moving objects
5. Sensitivity towards motion
6. The distance from the object

Bibliography

- [1] F. Itakura,, *Minimum prediction residual principle applied to speech recognition.* IEEE Trans. Acoust. Speech, Signal Proces., vol. ASSP-23, pp. 67-72, February 1975.
- [2] S. B. Devis and P. Mermelstein. *Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences.* IEEE Trans. Acoust. Speech, and Signal Proces., vol. ASSP-28, no. 4, August 1980.
- [3] Y. Hao and X. Zhu,. *A new feature in speech recognition based on wavelet transform.* in Proc. IEEE 5th Inter. Conf. on Signal Processing (WCCCICSP 2000), vol 3, pp. 1526-1529, 21-25 August 2000..
- [4] J. Mahkoul,*Linear prediction : A tutorila review.* in Proc. IEEE., vol. 63, no. 4, pp. 561-568, April 1975.
- [5] C. S. Burrus, R. A. Gopinath, and H. Guo *Introduction to Wavelet and Wavelet Transform.* Englewood Cliffs, NJ: Printice-Hall Inc., 1998
- [6] L. R. Rabiner *TI 46-Word Speaker-Dependent Isolated Word Corpus.* NIST, NIST Speech Disc 7-1.1, 1991.
- [7] Z. Tufekci and J. N. Gowdy, *Feature extraction using discrete wavelet transform for speech recognition.* IEEE Inter.Conf. Southeastcon2000, pp. 116-123, April 2000.
- [8] C. S. Burrus, R. A. Gopinath, and H. Guo,. *Introduction to Wavelet and Wavelet Transform,*. Englewood Cliffs, NJ: Printice-Hall Inc., 1998.