A Dissertation
on

'Multimodal Tracking System for Boarder Surveillance'

submitted in partial satisfaction of the requirements for the degree of

Master of Technology
in
Signal Processing & Digital Design

by

Chhavi Dhiman
(Roll no.:-2K12/SPD/05)

Under the guidance of

Prof. Dr. Rajiv Kapoor

Department of Electronics and Communication Engineering
Delhi Technological University
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042
(2013-2014)

Certificate

DELHI TECHNOLOGICAL UNIVERSITY,
(Formerly Delhi College of Engineering)
BAWANA ROAD, DELHI – 110042

Department of Electronics & Communication Engineering

This is to certify that the Dissertation entitled "Multimodal Tracking System for Boarder Surveillance" submitted by Chhavi Dhiman, Roll. No. 2K12/SPD/05, in partial fulfillment for the award of degree of Master of Technology in Signal Processing & Digital Design at Delhi Technological University, Delhi, is a bonafide record of student's own work carried out by him under my supervision and guidance in the academic session 2013-14. The matter embodied in dissertation has not been submitted for the award of any other degree or certificate in this or any other university or institute.

Date: Project Guide

Dr. Rajiv Kapoor
Professor & Head
Department of Electronics &
Communication Engineering
Delhi Technological University,
(Formerly Delhi College of Engineering)
Delhi -110042

Candidate Declaration

I hereby declare that the work presented entitled "Multimodal Tracking System for Boarder Surveillance" as Dissertation in requirement of partial fulfillment for the award of the degree of Master of Technology in Signal Processing and Digital Design and submitted to Delhi Technological University, Delhi is an authenticate record of my own work carried out under the supervision of Dr. Rajiv Kapoor, Professor & Head Department of Electronics & Communication Engineering .

The matter presented in this thesis has not been submitted by me for the award of any other degree of this university or any other university.

Date: Chhavi Dhiman

2k12/spd/05

Acknowledgement

First of all, I would like to express my gratitude to my advisor, Dr. Rajiv Kapoor, Professor & Head,Department of Electronics & Communication Engineering. He has been present throughout my entire journey at Delhi Technological University, as the professor in my very first class here, to the final signature on my thesis. I truly admire his depth of knowledge and strong dedication to students and research, that has made him one of the most successful professors ever. He has provided me tremendous insight and guidance throughout my research work, and I'm not sure where I would be without his encouragement and moral support. I am glad that I was given opportunity to work with him.

I am greatly thankful to entire faculty and staff of electronics & Communication Engineering Deptt. for their, continuous support, encouragement and inspiration in the execution of this "Dissertation" work.

I must also mention my friends : Sidhart Gautam ,Varun Sangwan and Ravi. It has been a pleasure working with them, and I wish them the best in all their endeavors.

Lastly, but perhaps most importantly, I would like to thank my family. None of this would have been possible without the love and encouragement of my father, Mr.Satya Narain, my mother, Mrs. Neelam Kumari , my two younger siblings, Khushboo and Krishh. They have always been there to help me navigate through life.

Chhavi Dhiman

2k12/spd/05

Abstract

Video surveillance activity is the activity of observing the behavior of object under motion, that object can be a human being or vehicle. The surveillance at boarders is a very sensitive issue of security of nation. But unfortunately till date no effective automated real time boarder surveillance system has been developed which can handle the real problems of the boarders and provide solutions to them. With the proposed method for tracking object on boarders, with the help of two features (*visual and audio/ seismic signal information*) a *real time multimodal system* has been developed which caters the solution to the problems at boarder i.e. large distance between the camera and the object. It is the major problem, due to the scenario of the boarder. This problem is overcome by adding seismic signal information, with the help of which we are able to identify without viewing the scene that how many and which object (tank/normal training vehicle) is there in the field. It is done by extracting the *unique pattern* available in the *mixed seismic signal*, if there is more than two vehicle/objects in the field. And the normal tracking is performed on the basis of visual information. For this specific task in this work we have used the simple and efficient *tracking algorithm* of optimization *(PSO) Particle Swarm Optimization.*

Table of Contents

List of Figures

List of Tables

Candidate Declaration

I hereby declare that the work presented entitled “Multimodal Tracking System for Boarder Surveillance” as Dissertation in requirement of partial fulfillment for the award of the degree of Master of Technology in Signal Processing and Digital Design and submitted to Delhi Technological University, Delhi is an authenticate record of my own work carried out under the supervision of Dr. Rajiv Kapoor, Professor & Head Department of Electronics & Communication Engineering .

The matter presented in this thesis has not been submitted by me for the award of any other degree of this university or any other university.

Date: Chhavi Dhiman

2k12/spd/05

Acknowledgement

First of all, I would like to express my gratitude to my advisor, Dr. Rajiv Kapoor, Professor & Head,Department of Electronics & Communication Engineering. He has been present throughout my entire journey at Delhi Technological University, as the professor in my very first class here, to the final signature on my thesis. I truly admire his depth of knowledge and strong dedication to students and research, that has made him one of the most successful professors ever. He has provided me tremendous insight and guidance throughout my research work, and I'm not sure where I would be without his encouragement and moral support. I am glad that I was given opportunity to work with him.

I am greatly thankful to entire faculty and staff of electronics & Communication Engineering Deptt. for their, continuous support, encouragement and inspiration in the execution of this "Dissertation" work.

I must also mention my friends : Sidhart Gautam ,Varun Sangwan and Ravi. It has been a pleasure working with them, and I wish them the best in all their endeavors.

Lastly, but perhaps most importantly, I would like to thank my family. None of this would have been possible without the love and encouragement of my father, Mr.Satya Narain, my mother, Mrs. Neelam Kumari , my two younger siblings, Khushboo and Krishh. They have always been there to help me navigate through life.

Chhavi Dhiman

2k12/spd/05

CHAPTER1

INTRODUCTION

1.1 OVERVIEW AND RESEARCH ISSUES

Visual surveillance is the action of monitoring the behaviors of people, vehicles or other objects using video cameras for security or social control purposes. The monitoring of behaviors is mainly based on the detection of any suspicious, illegal, abnormal or emergent activities. Currently, this process is mostly accomplished by human operators manually watching the video tapes recorded by surveillance cameras. Since the early 1980s, many researchers have put enormous effort into research to automate the process of visual surveillance while preserving or improving the reliability and accuracy of human operators. Some excellent automatic visual surveillance systems have been developed over the years, especially in the area of traffic monitoring, shopping malls, railway stations. A big challenge is to develop an efficient automatic surveillance system in the area ***boarder surveillance,*** where there is chance of occurrence of unwanted intruder tank or vehicle. On boarders there is large distance of hundreds of kilometers between camera and the object. Visual images sometimes may not be very clear to identify the object. So, single feature is not reliable for surveillance. We need to include more than one feature, other than visual feature, to track the object.

Tracking on the basis of visual information/ video frames which are not very clear at large distance does not give accurate and efficient results. But with the help of information received as seismic signals observed from the movement of any object (intruder tank) on the ground, and identifying different vibrations for different sized objects such as tank, truck, bicycle, along with visual information, it produces better result of tracking. It is also called as ***multimodal system of tracking,*** because two models – voice and image are being used to achieve the objective of tracking. This is perfect combination of information for boarder surveillance activity, to be performed in a better way.

Practically, seismic signals are being used for boarder surveillance but those signals are being analyzed manually. No automated boarder surveillance system exists, which makes use of both videos and seismic signals. In our work I have tried to put these two models together and develop an automated boarder surveillance system.

1.2 OBJECTIVES

The objective of my research is to develop a computer vision-based prototype and related methodology which can be used to develop a real-time automated boarder surveillance system that is capable of detecting and tracking of any sudden appearance of object on boarder area at a large distance.

To make the overall objective more attainable, this is broken down into the following tasks:

1. The seismic signal is received using a seismic sensor. The received signal may be due to the presence of one object on the ground or many. So, our first task is to separate every disturbance pattern for individual object using Repeating Pattern Separation Algorithm (RPSA) by breaking audio (mixed) signal into four signals using 4-channel Multirate filter Banks and processing each sub-signal via RPSA.
2. If in the broken/separated patterns, there is more than one pattern it means there are more than one object in the field with in the range of seismic sensor.
3. The visual tracking of the object in the video is performed using PSO (Particle Swarm Optimization Algorithm).
4. We need to identify one of the pattern or the only pattern which matches with any type of tank generated audio pattern. If the pattern exists in the audio, it means in the video/frame the intruder tank is present and our next task is to identify the tank by tracking it.

5. If there is no audio pattern existing in the audio patterns, it assures that there is no tank in the field within the range of sensor. This leads to no misleading of identifying any other object as tank.
6. In this way both the results obtained from visual tracking and audio separation are combined and a final decision is taken whether the object in the video is the intruder tank or not.

1.3 STRUCTURAL LAYOUT OF THE THESIS

In the following chapter 2, basics required for image processing and audio signal processing, are discussed in details. It provides an easy to understand the methodology used to implement the objective of the thesis.

In chapter 3, a literature review is provided of the related work in the field of tracking and the audio signal processing. This chapter deals with the methods followed to perform the defined task, the results obtained and the possible improvement that can be made to improve the results. It gives the insight of the possible solutions available to us for a given problem. And we can relate how the proposed method is better than the existing ones.

Chapter 4, named as Proposed Method, deals with the explanation of the proposed method in the thesis for the real time boarder surveillance problem statement. How the problem statement is handled, which algorithms are being used, are discussed in the form of flowcharts and algorithm format

Chapter 5, Methodology used in Seismic signal Processing, discusses about the explanation of the mathematics used in the Seismic signal Processing signals. To understand the algorithm and its functioning, it is important to understand the concept of new introduced terms, and how are they computed, what is their significance. All these questions are answered in this chapter.

In Chapter 6, all the simulations and results of the proposed algorithm are showcased and from them the conclusion is made that how the algorithm proves to be a better one.

In Chapter 7 Conclusion about the result, approach followed, drawbacks, and scope of improvements are discussed.

CHAPTER 2

BASICS OF IMAGE & AUDIO SIGNAL PROCESSING

Multimodal Video surveillance helps to monitor the dynamic scene for security control, for which it should be able to track a specific object throughout its appearance in the scene. The monitoring action is especially required now-a-days due to terrorism threats and activities around the world in recent years and the urgency to tighten security measures in crowded environments such as train stations, airport and shopping malls. The security at the boarders is equally important as it is in the country/ crowded/public places. As a result, video surveillance is currently one of the most active research topics in computer vision and pattern recognition, especially in human-based video surveillance. Video surveillance has a wide spectrum of promising applications, including access control in special areas, human identification at a distance, crowd flux statistics and congestion analysis, detection of anomalous behaviors, interactive surveillance using multiple cameras, and so on.

In general, the processing framework of video surveillance in dynamic scenes has been divided in three stages. The first stage deals with video based object tracking in the scene. The second stage deals with audio signals i.e. seismic signals received with help of seismic sensors with the capacity to sense the seismic vibrations of the ground within 25 meters. Third stage deals with combining the two results received from two stages via array processing.

STAGE 1:

My research visual tracking module mainly targets the three of these, background subtraction, motion detection and object tracking which will be discussed in the following sub-sections.

2.1 Reading the video

The captured video is read in terms of frames and the frames are ready to be processed.

2.2 Object Detection

To start tracking first task is to identify the object in the frame. The first step to identify the object is background subtraction.

2.2.1 Background Subtraction

Background Subtraction is preferably used for the videos, for object localization in foreground, which are captured by static cameras. Any significant change in an image background represents a moving object/ROI. The pixels of the regions undergoing change are marked for further processing. Usually, a connected component algorithm is used to collect connected regions representing the objects. This process is referred to as the *background subtraction.* This is performed by differencing two consecutive frames, which produce change occurred from one frame to another.
Throughout the frame there can be some small changes as well as some large changes. So, to obtain the ROI/ changed connected pixels only due to moving object post-processing steps i.e. morphology (dilation and erosion) are applied on the frame.

$$\text{frame\_diff} = \text{frame}^{i+1} - \text{frame}^{i} \quad i=1:N; \quad N= \text{total no. of frames in a video} \tag{1}$$

2.2.2 Morphology

Morphology is the technique of analysis of geometrical structure that enhances the vision of the image. ***Dilation and Erosion*** are the two basic methods used to perform image enhancement. With the different sequence of application of these two operations (dilation and erosion) two new methods are being defined under morphology called ***Opening*** (erosion and dilation) and ***Closing*** (dilation and erosion). Mathematically dilation can be represented as shown below:

Dilation ➔ $$A \oplus B = \bigcup_{b \in B} A_b \qquad (2)$$

Where A is an image and B is the ***Structuring Element (SE)***. It is similar to union operation. Hence dilation operation enlarges the connected components of the changed pixels obtained after background subtraction.

S.E. is shape/structure used to interact with the geometrical input matrix/structure. It can be a line, ring, ball. Its size can be a 3X3 square matrix or 21X21 square matrix large .

Erosion is the dual operation of the dilation operation. It can be defined mathematically as :

Erosion ➔ $$A \ominus B = \bigcap_{b \in B} A_{-b} \qquad (3)$$

Erosion is similar to intersection operation. Hence, it reduces the size of connected components of the changed pixels.

Mathematically opening is defined as :-

Opening ➔ $$A \circ B = (A \ominus B) \oplus B \qquad (4)$$

Closing ➔ $$A \bullet B = (A \oplus B) \ominus B \qquad (5)$$

We can conclude from fig. 1 that as a result of dilation operation input matrix expands to light Blue Square, with curved ends. Dark Blue Square shrinks to light Blue Square due erosion operation performed on it. Opening operation is equivalent to erosion followed by dilation operation. It results in a light blue square with the removal of sharp corners. Closing operation is equivalent to dilation followed by erosion operation.

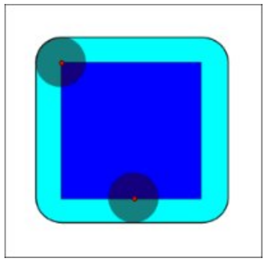


Fig: 1(a) Dilation

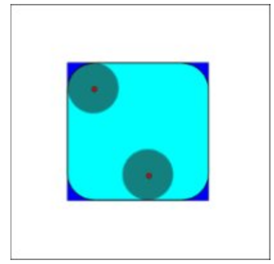


Fig: 1(b) Opening

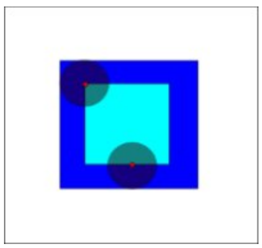


Fig: 1© Erosion

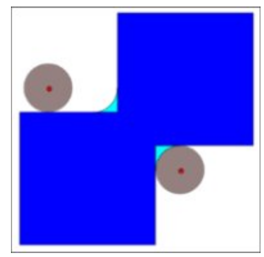


Fig: 1(d) Closing

fig. 2.1 A=dark blue square; B=circular disk (grey); Operation Output= Light blue colored square

It results in the input matrix along with the additional curved edges at the sharp corners.

2.2.3 Parameters calculation for object localization

After post-processing operation the largest connected components represent the object under motion. Now, the next task is to compute area and centroid of the obtained ROI. And further cover the ROI with a bounding box (a rectangular boundary around the object).

The obtained features about the desired object under ROI are used for tracking. We can apply the tracking algorithm i.e. Particle Swarm Optimization Algorithm (PSO). It is explained in the following chapters.

STAGE 2:

Audio tracking module deals with the algorithm which is based on the basic concept of mathematics and statistics according to which a long time series sequence is separated into small/ simple time series sub sequences. Individual sub sequence is being studied and unimportant subsequences are rejected.

To divide the long time series in sub sequences, M-channel Multi-rate Filter Banks are being used.

2.3 M-channel Filter Bank

A Filter Bank is an array of band-pass filters that separates the input signal into multiple sub signals. Each of the sub-signal carries a single frequency sub-band of the original signal. The process of decomposition of the input signal into multiple sub signals, performed by the filter bank is called *analysis*. The reconstruction process is called *synthesis.*

Major application of filter banks is the bank of receivers, where received signal are is required to be split into multiple sub signals. To reduce the data to be processed, save storage and lessen the complexity, multi-rate sampling techniques were introduced initially. On the basis of concept of multi-rate sampling techniques, Filter Banks are developed. Therefore they can be used for image coding, voice coding, and radar and so on.

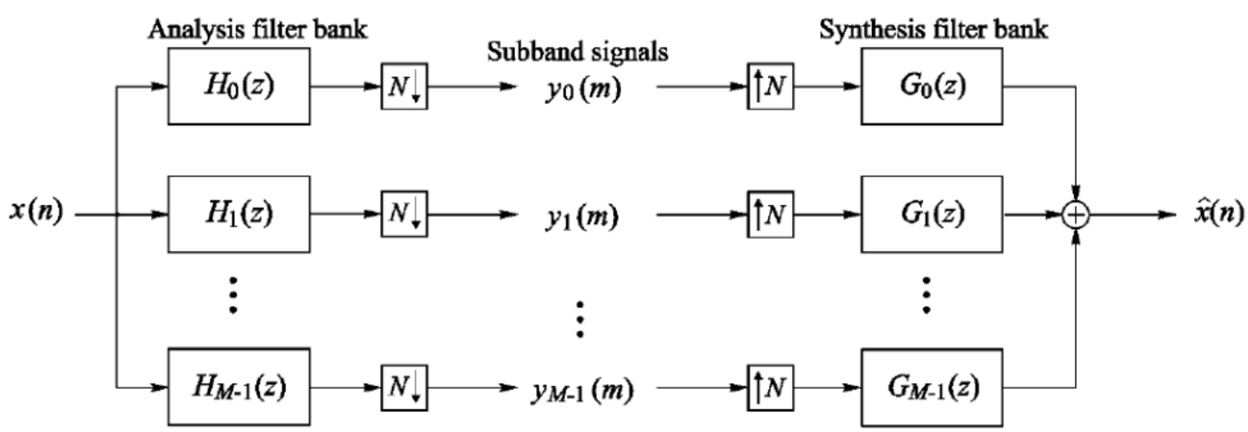


Fig 2.2 M-channel Filter Bank

2.4 Basic Multi-rate operations:

2.4.1 Decimation and Interpolation

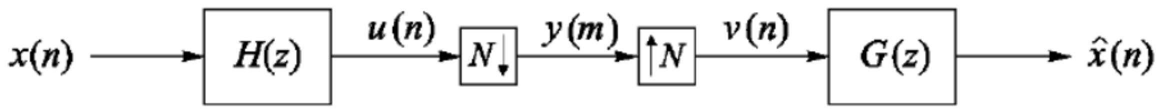


Fig 2.3 Typical components of Filter Bank

In above diagram, first x(n) an input signal is made to pass through H(z) a LPF and u(n) is received which is down-sampled / decimated by factor N. A decimated signal y(m) is obtained such that it contains every Nth element of the u(n) low pass signal. It is further up-sampled / interpolated by factor N and up-sampled signal v(n) is obtained. In interpolated signal N-1 number of zeros are inserted between two samples of the input signal y(m). The expressions of the intermediate signals of a filter bank, can be mathematically written as:

$$Y(Z^N)=V(Z) \tag{6}$$

After down-sampling and up-sampling by ***N*** the values w(nN) and u(nN) are still equal, while all other samples of v(n) are zero.

$$\frac{1}{N}\sum_{i=0}^{N-1} e^{j2\pi in/N} = \begin{cases} 1 & \text{for } n/N \in \mathbb{Z}, \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

Relationship between u(n) and v(n) can be written as:

$$v(n) = u(n)\frac{1}{N}\sum_{i=0}^{N-1} W_N^{-in} \tag{8}$$

Where $W_N = e^{-j2\pi/N}$

Z-transform is given by:

$$\begin{aligned} V(z) &= \sum_{n=-\infty}^{\infty} v(n) z^{-n} \\ &= \frac{1}{N}\sum_{i=0}^{N-1}\sum_{n=-\infty}^{\infty} u(n)\left[W_N^i z\right]^{-n} \\ &= \frac{1}{N}\sum_{i=0}^{N-1} U(W_N^i z). \end{aligned} \tag{9}$$

From eq. (6) and (9)

$$Y(z) = \frac{1}{N} \sum_{i=0}^{N-1} U(W_N^i z^{\frac{1}{N}}) \tag{10}$$

And V(z) = H (z) X (z)

$$Y(z) = \frac{1}{N} \sum_{i=0}^{N-1} \; H(W_N^i z^{\frac{1}{N}}) \; X(W_N^i z^{\frac{1}{N}}) \tag{11}$$

Finally we can conclude

$$\begin{aligned} \hat{X}(z) &= G(z) \; Y(z^N) \\ &= \frac{1}{N} \sum_{i=0}^{N-1} G(z) H(W_N^i z) X(W_N^i z) \end{aligned} \tag{12}$$

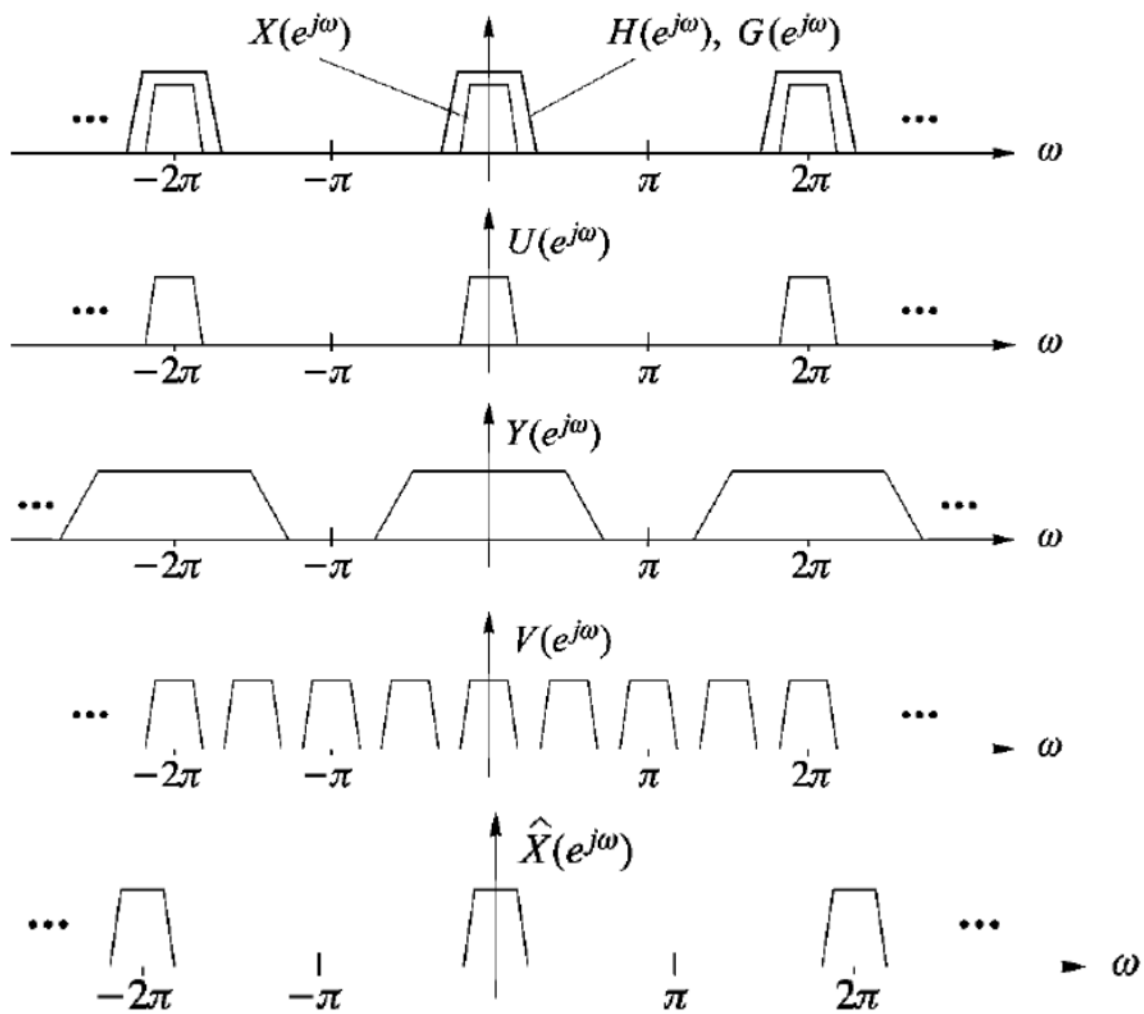


Fig 2.4 Signal spectra for decimation and interpolation for the structure in Fig 2.3 (non-aliased case).

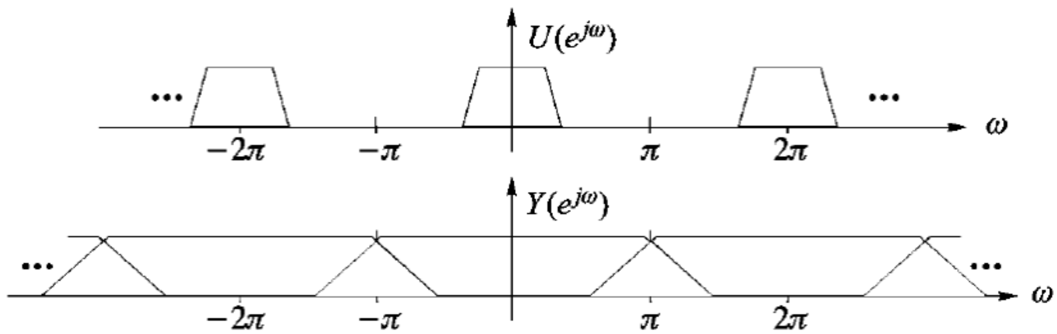


Fig 2.5 Signal spectra in the aliased case

The aliasing case occurs when spectra overlaps, as shown in fig 2.5. x(n) original signal can only be recovered from y(m) successfully only if no aliasing occurs. But aliasing is a normal operation in multi-rate filter banks. But still, the Multi-rate FBs allow perfect reconstruction, due the fact that they can be designed in such a way that the aliasing components from all parallel branches get compensated at the output.

Two channel filter bank, is called as Quadrature Mirror filter. It is a special case of M-band Filter bank. For QMF input-output relation is given as:

$$\begin{aligned}\hat{X}(z) \quad = \quad & \tfrac{1}{2} \quad [H_0(z)\,G_0(z) + H_1(z)\,G_1(z)]\ X(z) \\ & +\tfrac{1}{2} \quad [H_0(-z)\,G_0(z) + H_1(-z)\,G_1(z)]\ X(-z)\end{aligned} \tag{13}$$

First term of the equation (13) represents the transmission of the X(z) through the system, and second term represents the aliasing component at the output. For perfect reconstruction the first term or transfer function should be a delay function as shown below in equation (14).

$$S(z) = H_0(z)\,G_0(z) + H_1(z)\,G_1(z) = 2\ z^{-q} \tag{14}$$

And the second term, an aliasing component should be zero i.e.

$$F(z) = H_0(-z)\,G_0(z) + H_1(-z)\,G_1(z) = 0 \tag{15}$$

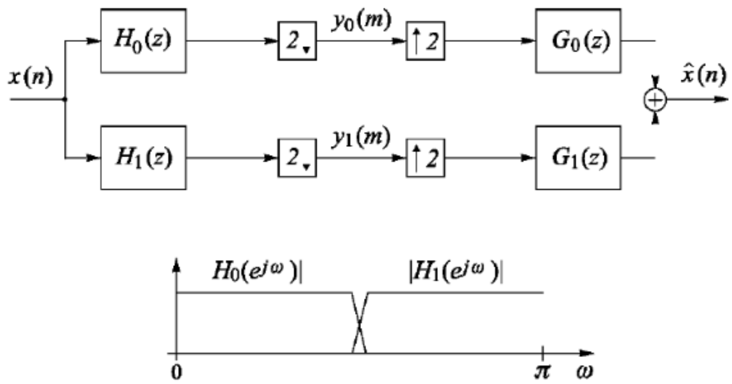


Fig 2.6 2-channel filter bank

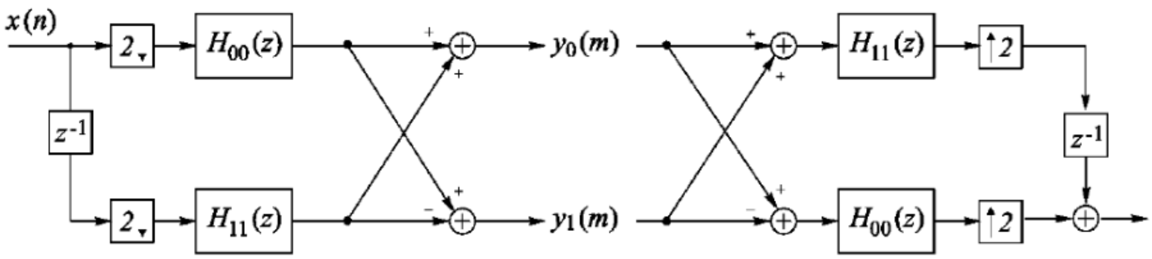


Fig 2.7 QMF in polyphase structure

The name QMF is due to the mirror image property with symmetry around π/2.

$$|H_1(e^{j\frac{\pi}{2}-\omega})| = |H_0(e^{j\frac{\pi}{2}+\omega})| \qquad (16)$$

2.5 Power Spectrum of Voice Signal :

The power spectrum of the signal represents the power of the signal at each frequency that it contains. The white noise contains all frequencies at the same power, therefore its power spectrum is a flat curve (horizontal straight line).

If we have a continuous time signal say y(t) of infinite length, the power spectrum can be expressed as:

$$P_y(w) \propto |Y'(w)\}^2 \qquad (17)$$

Where Y'(w) is the fourier transform of y(t)

$$\tilde{y}(\omega) = \int_{-\infty}^{\infty} y(t)e^{-i\omega t}dt \qquad (18)$$

2.6 Beat Spectrum :

A **beat** is defined as interference between two sounds of slightly different frequencies, perceived as periodic variations in volume, with the rate equal to the difference between the two frequencies.

The **beat spectrum** is defined as the measure of acoustic self-similarity as a function of time lag. A strong beat spectrum peaks occur at repetitive times in a highly structured music/ signal. Beat spectrum helps to detect the speed and the relative strength of particular beats. Therefore it helps to distinguish two rhythms of the same speed (tempo). **Beat spectrogram** graphically describes the rhythm variations over time.

The beat spectrum does not depend on particular attributes i.e. energy, frequency. Therefore beat spectrum can be calculated for a signal of any type of genre (frequency/energy).

2.7 Seismic Signals and Sensors:

Seismic waves are waves of energy generated generally as a result of earthquake, explosion or volcano that imparts low frequency acoustic energy, and travel through the Earth's layers. Other than natural activities seismic waves are generated due to movement of any abject on the ground. The propagation velocity of the waves through ground layers depends on density and elasticity of the medium. Velocity increases with the depth of the earth.

In geophysics the refraction or reflection of seismic waves in the structure of the Earth's interior is used for research. Man-made vibrations are often generated to investigate shallow, subsurface structure of the earth.

Heavier objects generate sufficiently good strength of seismic signals on the ground, which can be recorded by seismic sensors or accelerometers. In the market seismic sensors are available which can sense the ground vibrations in the range of 30 m, for study purpose.

Broadly we can classify the seismic waves in two types of the waves:-

1) Body waves
2) Surface waves

Body waves travel through the interior of the earth, but surface waves travel across the surface. Surface waves travel longer distance across the surface and can be sensed for longer duration than body waves. It is due to the fact that particle motion of the surface waves is larger than that of body waves. Therefore surface waves are responsible for more damage.

Body waves can be further classified in two types of waves:- Primary waves (P) and Secondary waves (S).

P-waves are compressional waves which are longitudinal in nature. They travel faster than other waves through the earth to arrive at seismograph stations first, hence the name, given to it, is primary wave. They move with twice the speed of S waves. In air they take the form of the sound waves, hence travel with the speed of sound in air.

S-waves are shear waves which are transverse in nature. It moves slower than P waves and produce displacement in the ground perpendicular to the direction of propagation.

Rayleigh wave is one of the types of surface waves. They are also called ground roll because they travel as ripples with motions that are similar to waves on the surface of water.

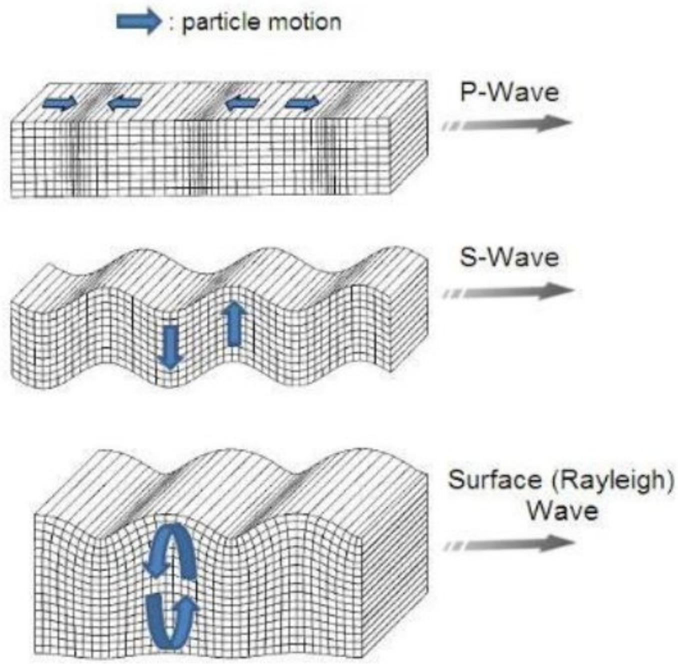


Fig 2.8 Types of Seismic waves

Free oscillations of Earth are generated as waves as a result of interference between two surface waves travelling in opposite directions. It is a surface standing wave. These waves are out of our scope of work.

Seismic sensor

Fig 2.9 Seismic sensor

Seismic sensor sense the vibrations generated by the motion of object on the ground. It senses the vibrations as analog / continuous-time signal. To read and analyze this signal, it needs to be converted into digital format. For this purpose we need an interfacing device between the sensor and the computer system. Arduino Duemilanove USB 328 Microcontroller Board is used, which is based on ATmega 328.

The board specifications, available in reference [3], are as follow :

1. Microcontroller	ATmega168
2. Operating Voltage	5V
3. Input Voltage (recommended)	7-12V
4. Input Voltage (limits)	6-20V
5. Digital I/O Pins	14 (of which 6 provide PWM output)
6. Analog Input Pins	6
7. DC Current per I/O Pin	40 mA
8. DC Current for 3.3V Pin	50 mA
9. Flash Memory	16 KB (ATmega168) or 32 KB (ATmega328) of which 2 KB used by bootloader
10. SRAM	1 KB (ATmega168) or 2 KB (ATmega328)
11. EEPROM	512 bytes (ATmega168) or 1 KB (ATmega328)
12. Clock Speed	16 MHz

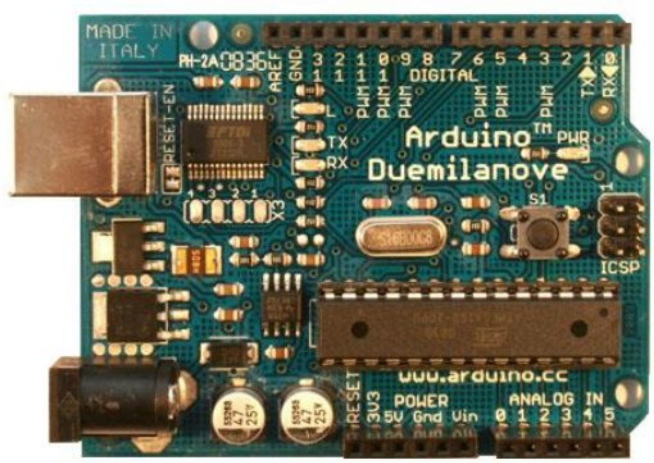


Fig 2.10 Aurdino Duemilanove microcontroller

It is very easy to operate as an interface. It can be directly connected to computer via USB connect with a AC-to-DC adapter/ battery to get started.

CHAPTER - 3
LITERATURE REVIEW

Object tracking based on particle filter with discriminative features

J Control Theory Appl 2013

In this work, the author has tried to improve the color based particle filter algorithm by adding discriminative features. In color based particle filter algorithm the weights of the particles plays an important role. In the basic algorithm weights of the particles do not represent the importance accurately and the particles generally get stuck in local regions. So, discriminative features are used to strengthen the weights of the particle filters. In the proposed visual tracking method, Bhattacharyya distance and local discrimination between the object and background are used to define the weights of the particles, which solves the local convergence problem of particles. This works for both single object tracing as well as multiple similar object tracking.

To define the object in the scene we are available with various features i.e. (CH) Color Histogram, HOG: Histogram of Oriented Graphs. Color Histogram feature is robust to scale variation, rotation and small changes in the background. But to improve the discrimination capability of color histogram of two objects of similar type in appearance other characters such as texture and contour are employed to represent the object. Recently, Histogram of Oriented Gradients method is employed to detect and track the object. HOG represents the edges and local contours. It is independent of color variations. But it is sensitive in orientation and does not describe the objects or background with large smooth regions. To overcome the drawbacks of HOG, HOG needs to be updated in time, when objects rotate.

The basic issue with a normal particle filter is particle degeneracy. To improve the robustness of color based particle filter the weights of the particles are always normalized Bhattacharyya distances between the color histogram of the object and the CH of regions marked by particles. If the CH of a region determined by a particle gives high similarity than the other regions, the weights of this particle are assigned as maximum.

Though there are certain limitations of this method also. If the color of the object is similar to the background, this algorithm fails to discriminate the two. It is also sensitive to illumination changes. In the future work, HOG can be used to calculate the local discriminations.

Observation and results based on this approach are as follows:-

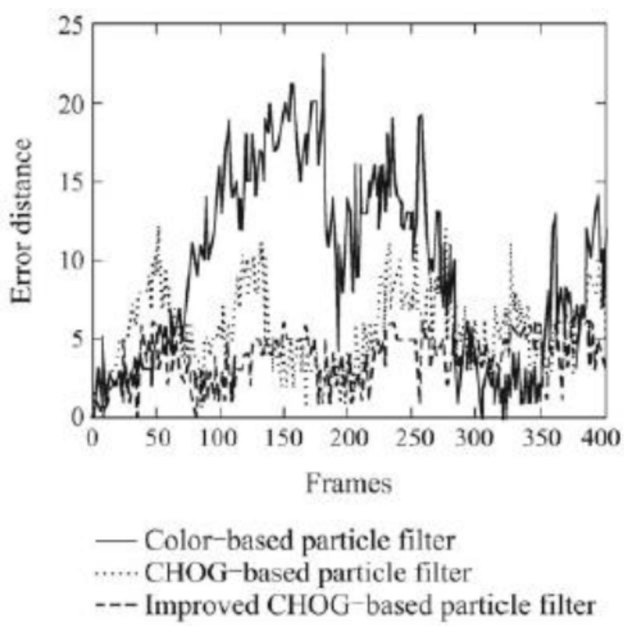


Fig 3.1 The error distance of three types of tracking algorithms
(C-PF,CHOG-PF, Improved CHOG-PF)

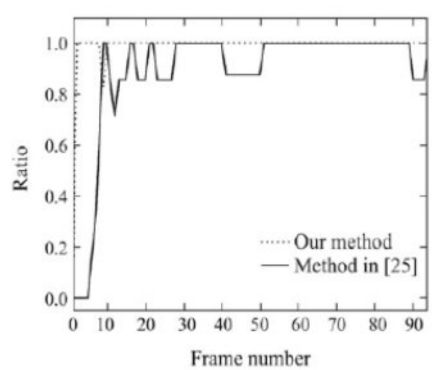


Fig 3.2 The ratio between the number of objects detected accurately and number of reality

Fig 3.3 Tracking effects of the improved method used in outdoor image sequences. The frame 124,137,150,172 and the are shown (left-right, top-bottom)

	Indoor sequence (640 × 480)	Car (576 × 768)	Football (640 × 480)
CH	10.5 f·s^{-1}	9.0 f·s^{-1}	14.0 f·s^{-1}
CHOG	8.3 f·s^{-1}	6.5 f·s^{-1}	11.2 f·s^{-1}
Our method	6.7 f·s^{-1}	4.7 f·s^{-1}	8.2 f·s^{-1}

Table 1 Comparison of frames per second with different methods

Object Tracking Using Maximum color Distance Under Illumination Change

International Journal of Advanced Robotic Systems 2012

This paper presents the new real and robust visual object tracking method, which can handle the large color variations in the frame due to illumination. The conventional Mean shift method performs well only for non-rigid moving objects. To handle the illumination variations maximum color distance on modified mean shift framework is being used to represent the object. The color distance is defined as the geometric distance between two color points in color space. Then the maximum color distance is the maximum value out of all calculated color distances. It is assumed that physical illumination model provides maximum color distance which is independent of ambient lighting and other illumination with the identical solid angle. The reflectance model based on physical theory is used to demonstrate that color distance is constant for ambient colored lights.

The proposed algorithm can be divided in two steps:-

1. Defining and modeling a target
2. Tracking the target

For modeling the target the current location and area of the tracked object with in the initial frame is initialized. Edges of the initialized object are detected. The color distances for the detected edge locations are calculated and probability map is generated for different shape template based on color distances.

For tracking in the next frame the candidate object is selected in the search region located at the center of the current location and its edges are detected. For each detected point on edges color distance is calculated and a probability map is produced. Now, the initialized target model (probability map) is compared with the candidate model (probability map) in the current frame. If the similarity with the target is less than the threshold, tracking is stopped and current location is updated or in the next frame new candidate is selected and corresponding probability map is generated and the process is continued further.

The similarity estimation between two models is measured using following equation:-

$$Similarity = w \mid f(T) - f(u_i) \mid + (1 - w) \cdot \frac{1}{\rho(u_i) \cdot \rho(T) + \varepsilon}$$
$$0 \le w \le 1, u_i \in U \qquad (19)$$

Where f : a histogram function for the maximum color distance calculation

P : a probability map,

T : the initialized target,

ε : a small positive number used to avoid dividing by zero and

w : the weight value of the colour distance histogram and the shape template. 0.5

This equation is applied to all candidates and is used to determine that the target in the next frame is a candidate with the minimum amount of similarity.

One of the results of the proposed robust object tracking algorithm is shown below, where the hand and face are successfully distinguished and tracked, being different faces:-

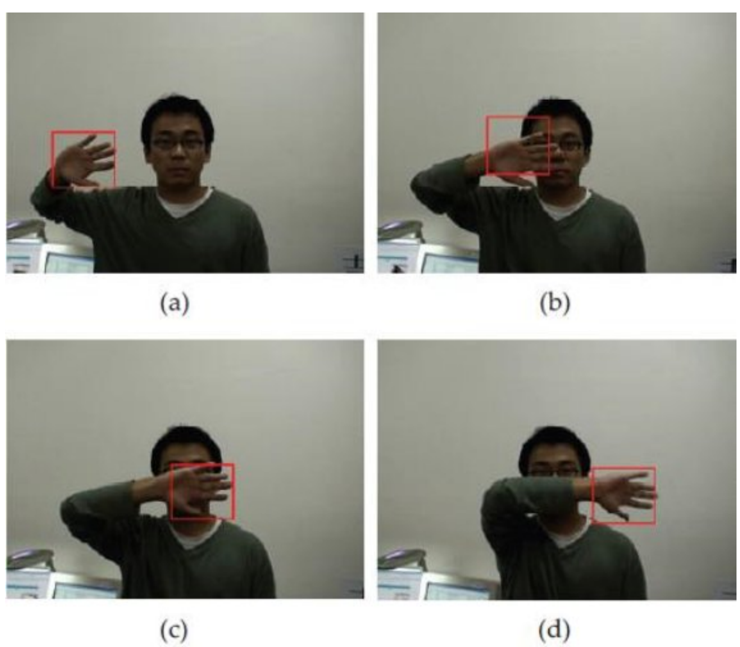


Fig 3.4 Experimental result for the crossover of a similar color distribution background by proposed algorithm. (a) initializedtarget (b) move into the face (c) intermediate crossover(d) move out of the face

The figure provides the comparison in the results of Normalized Gradient Matching, Mean Shift based method and the proposed method based on maximum color distance under illumination changes. It is observed that the proposed tracker tracks the face under illumination changes and gives better results.

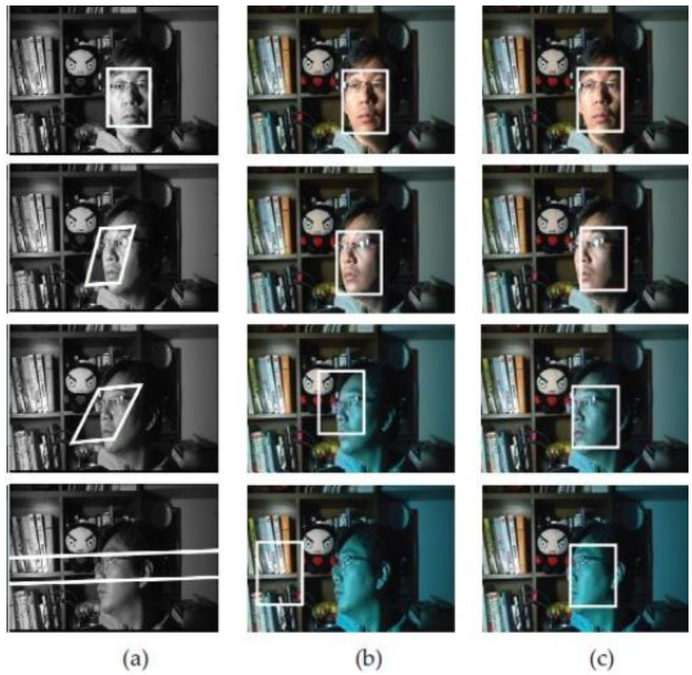


Fig 3.5 Experiment results for concurrent variation of appearance, colour and location under colour illumination change. (a) Normalized Gradient Matching (b) Mean shift based on color (c) The proposed algorithm

In future work this algorithm can be made robust by using invariant shape information with MCD.

Video Based Moving Object Tracking by Particle Filter

International Journal of Signal Processing,
Image Processing and Pattern Vol. 2, No.1, March, 2009

This research work is supported by MIC & IITA through IT leading R&D support project [2006-S-028-01].

In practical scenario, the video based object tracking requires to deal with non-stationary image stream that changes over time. The proposed algorithm is robust enough to handle non-linearity problem. In the paper, the proposed tracking system makes use of color information, distance transform (DT) based shape information and non-linearity. To handle non-gaussianity and nonlinearity particle filter has proven to very successful. Two approaches have been worked out to handle practical non stationary images. In the first approach a color based particle filter tracker is developed which makes use of deterministic search of the window, whose color content matches a reference histogram model. In the Second approach, particle filter uses the shape information, which is collected from edge features. The template and estimated regions in the video scene shape similarity is measured by their normalized cross-correlation of the distance transformed images. Where the template is created instantly by selecting any object from the scene by a rectangle. In this paper it is shown, information collected through these two approaches when fed to particle filter gives improved result from non-linearity.

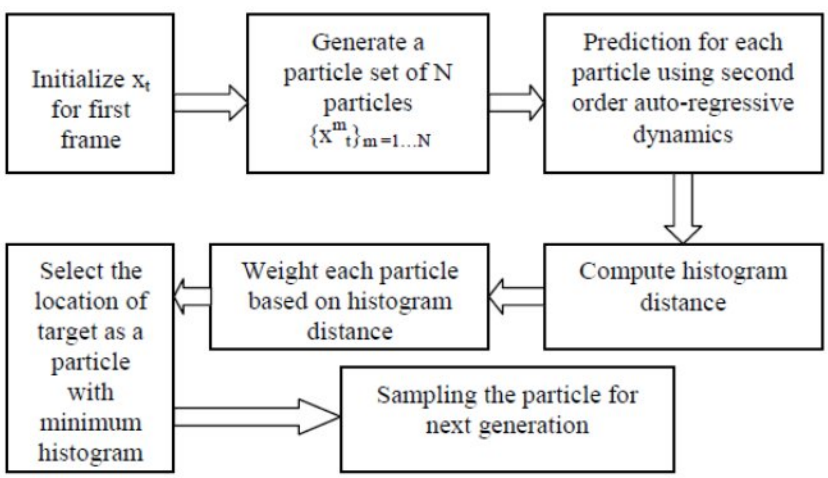


Fig 3.5 Color based Particle Filter implementation

Fig 3.6 (a) Original image (b) DT image map

The cross-correlation of template t(x,y) with a subimage f(x,y) is given by the following equation:

$$N_{f,t} = \sum_{x,y} \frac{(f(x,y) - \bar{f})(t(x,y) - \bar{t})}{\sigma_f \sigma_t} \tag{20}$$

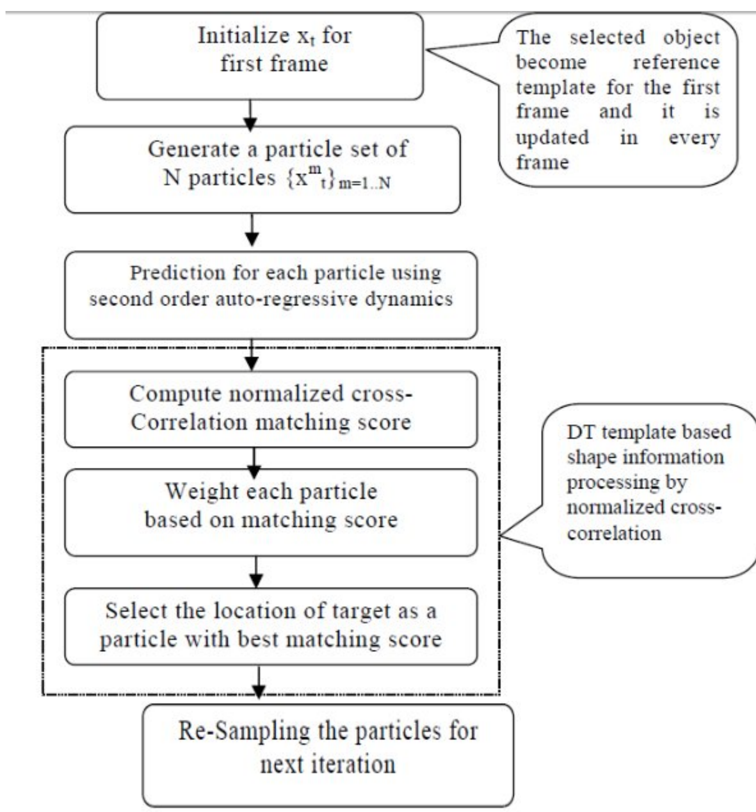


Fig 3.7 Algorithm flowchart of shape based system

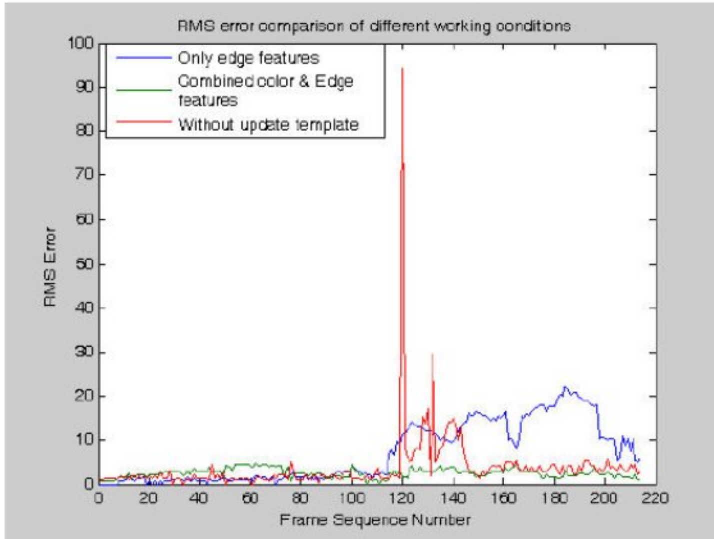


Fig 3.8 RMS error at tracking feature points for each frame

Experimental results of the proposed figure are shown in the above figure which displays the comparison in the obtained root mean square error in tracking the desired feature points using only edge features, combined color and edge features, and without updating the template. It can be observed that combined color and edge features give least root mean square value. Without updating the template while tracking shows very high peak of rms for few frames, otherwise rms value is small, as desired, which is an improved result for tracking the object in a non-stationary scene. Hence in the integrated framework of particle filtering, the proposed algorithm based on instant/adaptive DT template matching by normalized correlation is a robust approach.

For future work the proposed approach can be extended into multiple object tracing system or learning based system.

ECSE-626 Project: An Adaptive Color-Based Particle Filter

Fabian Kaelin McGill University Montreal, Canada

In the field of computer vision, the presented algorithm makes use of an Adaptive Color Based Particle Filter for target tracking in the video scene. The proposed approach is reliable and robust against changes in illumination or scale. The method takes the advantage of simple linear dynamical model of the particle filter and combines it with the likelihood model based on color histograms. The algorithm supports adaptive nature because the target model is updated over time.

Target model can be defined as :

$$s = \{x, y, \widehat{x}, \widehat{y}, H_x, H_y, \widehat{a}\} \tag{21}$$

Where x,y : coordinates of the centroid

$\widehat{x}, \widehat{y}$: velocity of the object

H_x, H_y: Dimension of the object

a: area of the target in the bounding box

Linear Dynamic model:

$$S_{t+1}=A*S_t + W_t \tag{22}$$

The algorithm can be outlined in the five following steps as:-

1. Resampling the particles to avoid degeneracy
2. Propagate each particles according to our dynamical model
3. Update the weight $\pi^{(n)}$ of each particle according to our likelihood model
4. Estimate the posterior state $p(x_t \mid Z_t)$ of the target, given the new frame z_t
5. Adapt the target's color distribution q to increase reliability and robustness

Degeneracy is the very common problem which appears in the Particle Filter linear model. After every iteration of sampling process, some of the particle carry the 0 weight value and with number of iteration later, we are left with very less ratio of the particles we started with. Small amount of particles to represent the target is not sufficient and they loose the track fairly, called as Degeneracy.

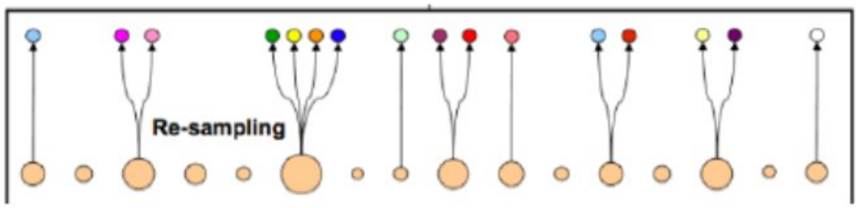


Fig 3.9 Resampling illustration from [1]. The particles with higher weights (bigger circles) are chosen with a higher probabilty, but the total number of samples stays the same

To avoid the problem of degeneracy, after every iteration of sampling the particles with weight value 0 are replaced by the particles of highest weight. It is similar to keeping multiple copies of the particles with highest weight value. It can be shown diagrammatically by the schema shown in fig .

Propagation of the particles: Particles in every frame will be representing the target location, and assigned with some weight values. Now, the weight values should be assigned in such a way that, the particles always point to the true target in the frame. With the motion the target position will be changing, so as the locations of the particles, to provide synchronization between two movements, the particles are assigned weights in such a way that the velocity of the target is captured. Velocity and scale of the target is controlled by the covariance parameter, used to define the weight to the particle. For slow motion of the target, covariance value need to be small and for high amount of motion, covariance value needs to be taken higher. It will be adjusted as per the sequence of the frames.

Update: Each particle has a hypothetical state of a target and a new observation, where it has to move. To assign new position to the particle, the likelihood of the particle to be at the next position needs to be calculated. For this first a color distribution is calculated and Bhattacharyya distance is computed, which provides the measure of similarity to the target q for every particle in the set. This Bhattacharyya distance is used to update the weights.

Weighting function is given by the equation as shown below:

$$k(r_i) = \begin{cases} 1 - r_i^2 & r_i \leq 1 \\ 0 & otherwise \end{cases} \tag{23}$$

where $r_i = ||v_i - c_i||/b$; b is a normalizer,

$b=sqrt(H_x^{\wedge 2}+H_y^{\wedge 2})$

it normalized the distance from the target center c = (x; y) to the actual pixel v_i that we are.

Because we want a probability distribution (independence of the size of the rectangle), we normalize the color distribution by a factor f:

$$f = \sum_{i=1}^{I} \frac{1}{k(r_i)} \tag{24}$$

where I is the number of pixels in the rectangle. The color distribution can be written as $p_s = \{ p_s^{(u)} \}$ u=1:8 where

$$p_s^{(u)} = f \sum_{i=1}^{I} k(r_i)\delta[h(v_i) - u] \tag{25}$$

Bhattacharyya coefficient, which is defined as

$$\rho[p, q] = \sum_{u=1}^{8} \sqrt{p(u)q(u)} \tag{26}$$

p,q are two discrete distributions, ρ is 1 if p and q are identical and decreases if p and q become different . Bhattacharyya distance can be calculated as :

$$d = \sqrt{1 - \rho[p, q]} \tag{27}$$

This distance is used to update the weights as:

$$\pi^{(n)} = \frac{1}{(\sqrt{2\pi}\sigma)} e^{-\frac{d^2}{2\sigma^2}} = \frac{1}{(\sqrt{2\pi}\sigma)} e^{-\frac{(1-\rho[p_{s}(n), q])}{2\sigma^2}} \tag{28}$$

Where σ is the variance of the Gaussian distribution, controlling the motion of the object/target. 0.1 variance value has been used in this work. Weights are normalized to limit the value between 0 and 1, it can be shown as:

$$\pi^{(n)} = \frac{\pi^{(n)}}{\sum_{i=1}^{N} \pi^{(i)}}$$

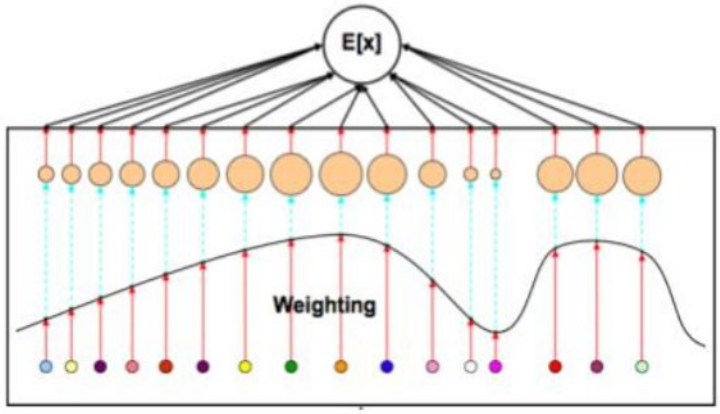


Fig 3.10 after updating the weights mean estimate is computed

One of the main advantages of the particle filter is that it supports multimodal distribution. Due to which a particle filter can consider multiple hypothesis, it is required when it not sure where the target is moving. It can be explained with the help of an example:

Fig 3.11 Example of a multimodal particle set tracking a face, but also caught onto an arm

While tracking a face, it is crossed by another face, the particles will get distributed to another face also, until one of the faces is chosen. Particles detected an arm of the similar color like face but Particle Filter discarded the arm soon and calculated the mean estimate as red box visible.

Experimental results

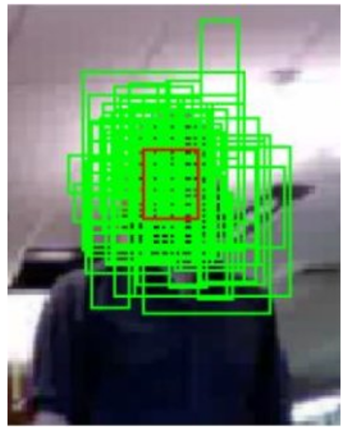


Figure 3.12
. Example of a particle set tracking a face, the green rectanglesare the particles and the red rectangle is their posterior(mean state)

Fig 3.13 Face tracking with cluttered background and directions and velocity changes - frames 13, 40, 69, 89, 140, 156 from top left to bottom right)

For the future work, it has been proposed that, the particle filter response can be improved if likelihood function is calculated not on the bases of the color information but also on the bases of shape of the target. Though adding the shape information for likelihood estimation for assigning weights to the particles, limits the algorithm to not be independent of rotation and it will loose its some of the robustness. But it would be interesting to see how the filter would perform if the likelihood would be based on mutual information as defined by Viola and Wells [2] instead.

Adaptive swarm optimization for locating and tracking multiple targets

Applied soft computing, @ 2012 Elsevier

To locate and track multiple targets is a crucial and challenging problem when environment is dynamic and uncertain in practical application. The main objective of this paper is to find solution of three fundamental problems i.e. identification of irregular target, locating multiple targets, tracking multiple targets. Firstly, to collect the target's shape from the available frame, an objective function is defined. Secondly, to locate multiple targets in the dynamic environment, adaptive PSO algorithm is used, which divides the swarm (set of particles) into multiple subgroups and adaptively adjust the number of particles in each subgroup, using competition and cooperation concept. To show the efficiency and high performance of the proposed algorithms,
A practical scenario of locating and tracking three ants is discussed in the paper.

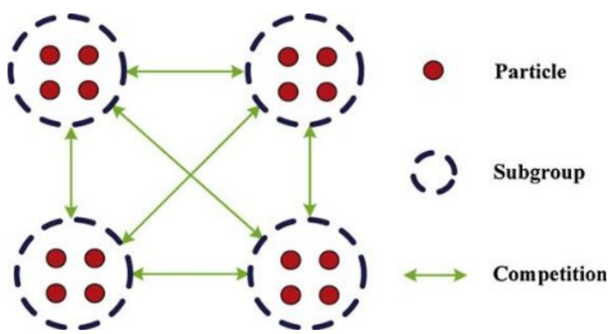


Fig 3.14 Main framework of the adaptive PSO algorithm. Red point and blue circle denote one particle in the subgroup and one subgroup, respectively. In addition, the green arrow denotes the competition among different subgroups.

Because of the easy implementation and less computational complexity of time, PSO algorithm has been successfully and widely utilized for locating and tracking purpose.

Let Position and velocity of the i^{th} particle be $X_i=\{x_{i1},x_{i2},\ldots x_{in}\}$ and $V_i=\{v_{i1},v_{i2},\ldots v_{in}\}$ respectively. Let $P_i=\{p_{i1},p_{i2},\ldots p_{in}\}$ and $G=\{g_1,g_2,\ldots g_n\}$ be the personal best previous position of the i^{th} particle and all particles, respectively.

The PSO algorithm can be summarized as below:

Step 1: Initialize the parameters including the particles position, the particles velocity, inertia weight and acceleration coefficients, etc.

Step 2: Calculate the fitness function $F(X_i(t))$ of the i^{th} particle and modify the best previous position $P_i(t+1)$ of the i^{th} particle by

$$P_i(t+1) = \begin{cases} X_i(t) & \text{if } F(X_i(t)) < F(P_i(t)) \\ P_i(t) & \text{if } F(X_i(t)) \geq F(P_i(t)) \end{cases} \tag{29}$$

Step 3: Current best position can be defined as:

$$G(t+1) = arg\,\min\{F(P_1(t)), F(P_2(t)), \ldots, F(P_n(t))\}. \tag{30}$$

Step 4: Velocity and position of the particles are updated by the equations model [31] :

$$\begin{aligned} V_{ij}(t+1) &= \omega(t)V_{ij}(t) + c_1 r_1 (P_i(t) - X_{ij}(t)) + c_2 r_2 (G(t) - X_{ij}(t)) \\ X_{ij}(t+1) &= X_{ij}(t) + V_{ij}(t+1) \end{aligned} \tag{31}$$

w = momentum coefficient of the previous velocity

large w value leads to exploration while small w value leads to exploitation.

C1 = cognitive factor responsible for previous best position.

C2 = social factor responsible for best position of all particles.

R1, r2 = uniformly distributed variables ranging from 0 to 1.

First term of velocity update equation, represents the momentum of previous velocity, second term of the same equation represents personal thinking of each particle, and third term denotes cooperation among all the particles.

Step 5: From step 2 to step 5 is repeated until stopping criteria is met.

Sometimes particles get trapped to the false locations. It happens due to imbalance in exploration and exploitation process of the particles. The value of w plays an important role in creating the balance between the global search ability and local search ability of the PSO algorithm, required to find the new best location for every particle. Local search ability is more powerful in the evolutionary process.

How to find fitness value for each particle, depends on objective function. It can be mathematically defined as:

$$F_{ij}(S, S_d) = \sum_{i=1}^{L} \sum_{j=1}^{W} \sum_{k=1}^{3} |S(i, j, k) - S_d(k)| \tag{32}$$

where S(I, j, 1), S(I, j, 2) and S(I, j, 3) denote the RGB values of the point (I, j), respectively. Sd(1), Sd(2) and Sd(3) denote the RGB values of the target, respectively. Additionally, L denotes the length of the target's window and W denotes the width of the target's window.

To locate multiple targets using adaptive PSO algorithm, an interaction matrix needs to be defined as:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ . & . & . & \cdots & . \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}$$

Where m : no. of targets, n: no. of particles in the PSO

Each element of matrix A, a_{ij} represents the distance between the i^{th} particle and central position of the j^{th} subgroup(target)

$$a_{ij} = \begin{cases} 0 & X_i \in G_j \\ 1 & \|(X_i, G_{cj})\|2 \leq T_d \quad X_i \notin G_j \end{cases} \tag{33}$$

Where Gj and Gcj denote the set and central position of the j^{th} subgroup (target) respectively. T_d is the exclusion threshold distance of each subgroup. It means that if distance of any point or particle from the center of the subgroup and the point does not belong to the subgroup, is in the range of threshold, the interaction element for i^{th} particle with the j^{th} subgroup will be assigned value 1. The value of the threshold plays an important role in the efficiency of an adaptive PSO method. Very low value of T_d may converge all the particles to one subgroup only.

$$\eta(t) = 0.5 \times \min_{j \neq i} \mathrm{Dist}(C_i(t), C_j(t))$$

The parameter η ($\eta_{min} < \eta < \eta_{max}$) denotes the smallest distance between the particle and the center position of the subgroup. Mathematically it can be expressed as η(t), as shown in the equation above.

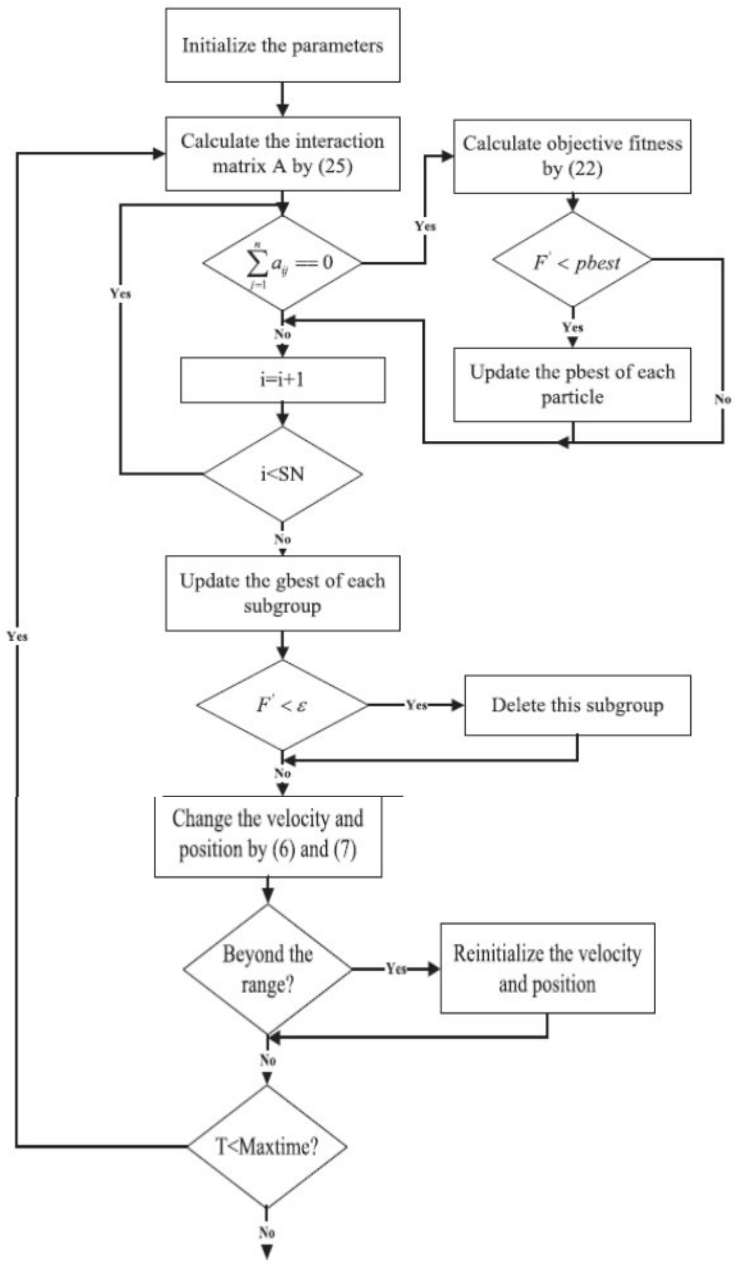


Fig 3.15 Adaptive PSO algorithm Flowchart

Small value η_{min} gives small covered region, large value η_{max} gives large number of particles. Therefore the value of η is selected in between minimum and maximum range.

If two objects to be tracked are adjacent to each other, adaptive PSO above algorithm does not accomplish the task, because particles in two subgroups tend to track the identical target. In order to track multiple adjacent targets a new algorithm needs to be framed. One of the two groups is responsible for tracking the central target, another subgroup needs to track external target. The

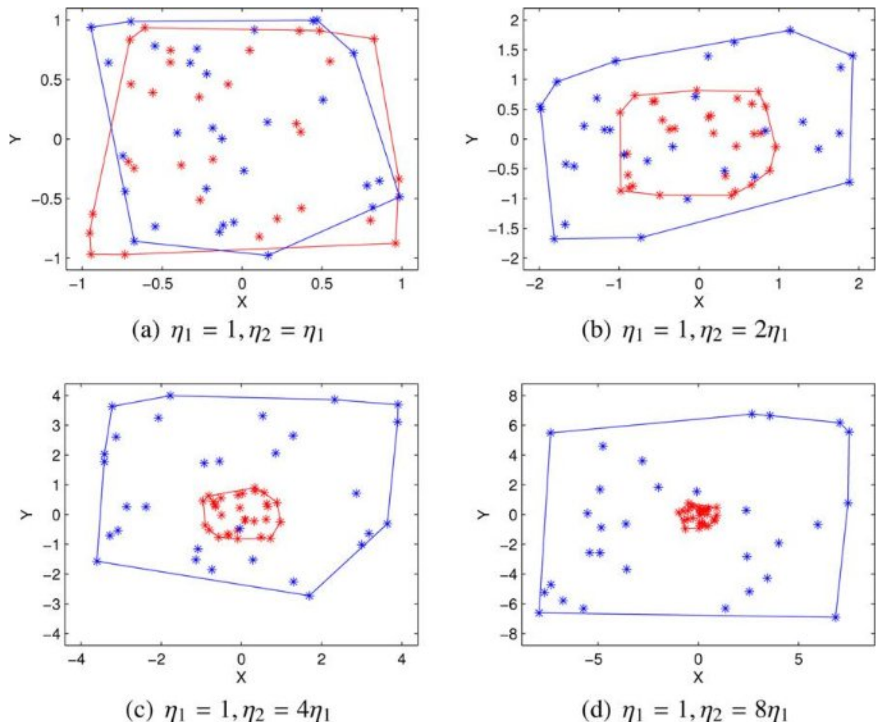


Fig 3.16 Different overlapping regions under different η1 and η2.

area of two regions can be viewed in the above figure, distinguishing the regions of two adjacent targets for different values of η.

Mathematically, two subgroups can be represented as model [34] and [35]:

$$X_i(t+1) = C_j(t) + \eta_1(2r_6 - 1)$$

$$X_i(t+1) = C_j(t) + \eta_2(2r_7 - 1) \quad (34\ \&\ 35)$$

where η1 and η2 : radius of internal and external subgroup, respectively.

R6, r7: two random variables in the range of [0,1].

Algorithm:

Step 1 :The parameters i.e. number of subgroups, number of particles in each subgroup, radii η1, η2 and max and min value of η are initialized.

Step 2: The minimum distance is calculated between two centers of the two subgroups as a parameter η.

Step 3: If the value of η is in the range of η_{min} and η_{max} , with help of model [31] defined by equations of position and velocity, is followed and adaptive PSO algorithm can be followed.

Step 4: But if the value of η is lesser than η_{min}, then model [34 and [35] defined by two separate position equations of the two adjacent targets to be tracked is followed.

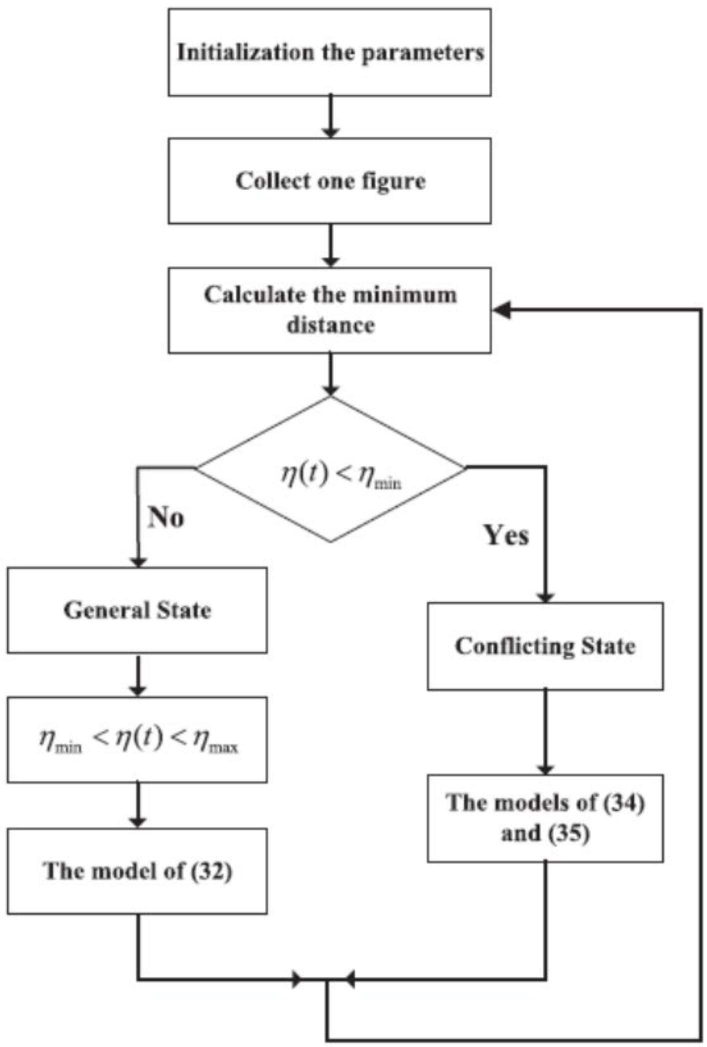
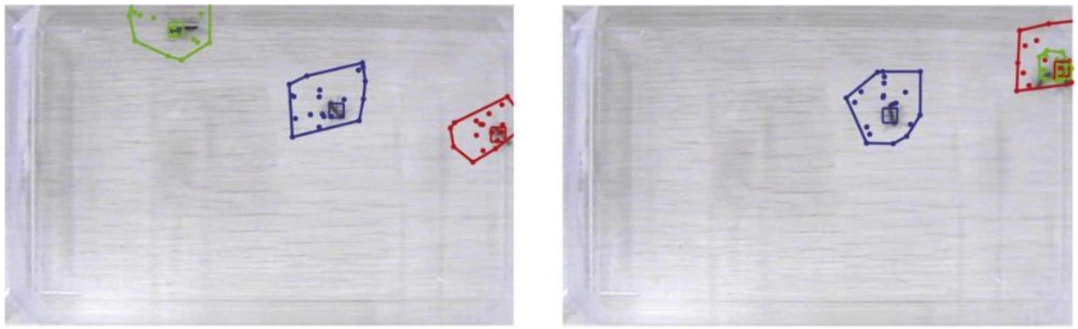


Fig 3.17 The proposed algorithm of tracking multiple targets.

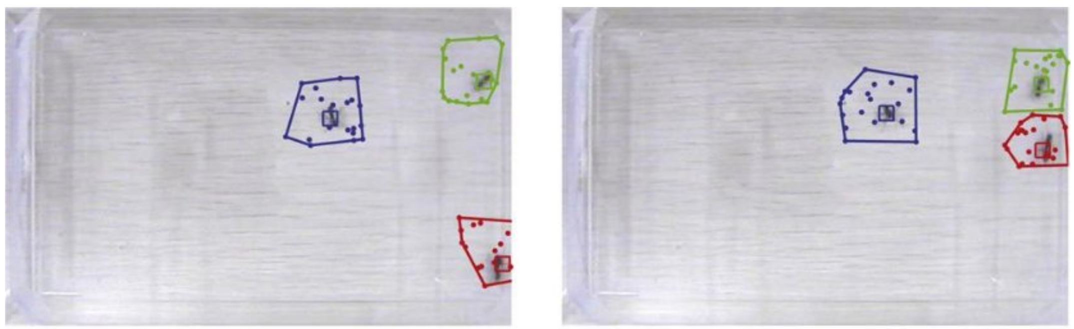
Step 5: A new frame is acquisitioned and again the minimum distance is calculated by as defined by step 2, followed by step 3 and step 4, iteratively.

Experimental results:



(a) t=10 (b) t=110

Fig Two subgraphs of tracking three ants in the early process.



(a) t=160 (b) t=213

Fig 3.18 Two subgraphs of tracking three ants in the end process.

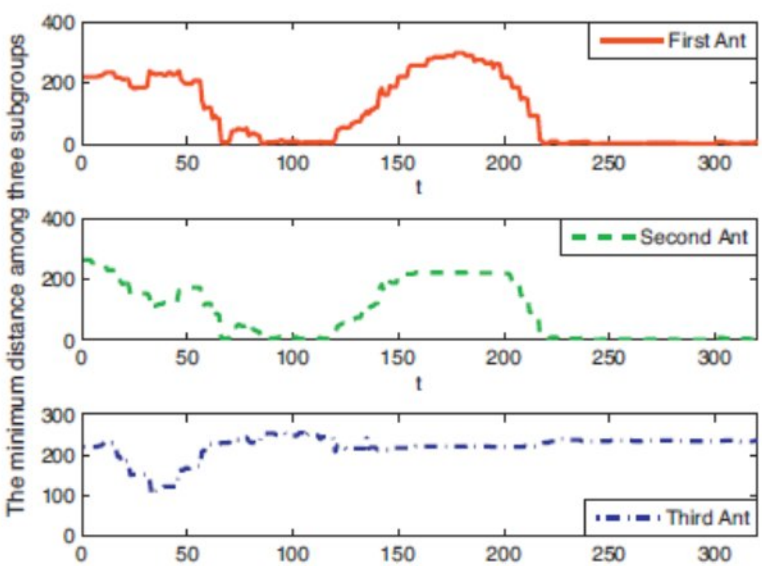


Fig 3.19 The minimum radius among the ants in the whole process.

In the future work, the object identification and locating it can be realized on the bases of shape and color both. The another practical problem that an object may suddenly jump out of the defined subgroup, needs to be worked upon.

<u>BLIND SEPARATION OF *L* SOURCES FROM *M* MIXTURES OF SPEECH SIGNALS</u>

New Mexico State University,
Klipsch School of Electrical and Computer Engineering

In the practical scenario we come across, a blind source separation problem. Where, we are not aware of the number of sources available to us for the number of mixtures to be separated. This paper discusses about the extension of successful and efficient kurtosis maximization algorithm used in speech separation, of two sources from two linear mixtures to handle the problem of arbitrary number of sources and mixtures.

GENERALIZED KMA (Kurtosis Maximization algorithm)

Let us assume there are L unknown speech signals $S=\{s_1,s_2\ldots..s_L\}$, which are mixed in linear fashion to produce M mixture signals as $X= x_1,x_2\ldots..x_M\}$. Our objective is to produce the unknown speech signals from the mixed available signals as output $Y=\{y_1,y_2, \ldots.y_L\}$, as approximated values of S.

Mathematically the generalized mixing model can be expressed as :

$$X(n)=A(n)^* S(n) \tag{36}$$

Where $X(n)=\{x_1,x_2,\ldots..x_M\}^T$ and

$S(n)=\{s_1,s_2,\ldots\ldots s_L\}^T$

A(n) is MxL mixing matrix

Our objective is to compute LxM W(n), a separation matrix, mathematically written as:

$$Y(n)=W(n)^*X(n) \tag{37}$$

Where Y(n) is the reconstructed estimated sources from the given mixture.

$$Y(n)=\{ y_1(n) \quad y_2(n)]^T \tag{38}$$

W(n) is a separation matrix.

A multidimensional objective function framed by kurtosis of the output signal (Y(n)) can be represented as:

$$J(n)=\{k_{y1},k_{y2}\ldots..k_{yL}\} \tag{39}$$

Where kurtosis of a signal (say) x can be calculated as :

$$\kappa_x \equiv \frac{E\left[x^4\right]}{\{E\left[x^2\right]\}^2} \tag{40}$$

To maximize the kurtosis separation matrix can be rewritten as :

$$\begin{aligned} \mathbf{W}(n+1) &= \mathbf{W}(n) + \mu \bigtriangledown \kappa_{\mathbf{y}} \\ &= \mathbf{W}(n) + \mu \begin{bmatrix} \frac{\partial \kappa_{y_1}}{\partial W_{11}} & \frac{\partial \kappa_{y_1}}{\partial W_{12}} \\ \\ \frac{\partial \kappa_{y_2}}{\partial W_{21}} & \frac{\partial \kappa_{y_2}}{\partial W_{22}} \end{bmatrix} \\ &= \mathbf{W}(n) + \mu \mathbf{C}(n) \end{aligned} \tag{41}$$

μ = step size

∇k_y = gradient of kurtosis of the output signals w.r.t. the elements of the separation matrix (W(n)).

C(n) = correction matrix

Experimental results

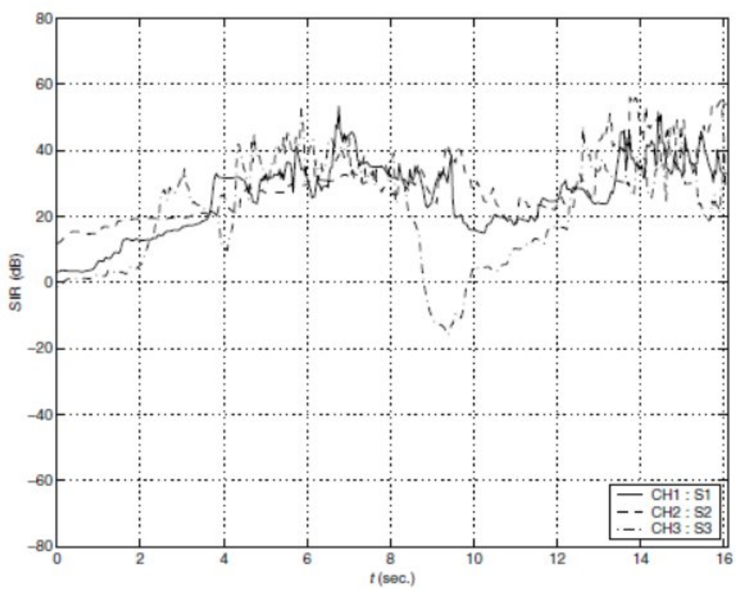


Fig. 3.20.
Signal-to-Interference Ratios of three source, three mixture simulation.(L=M)

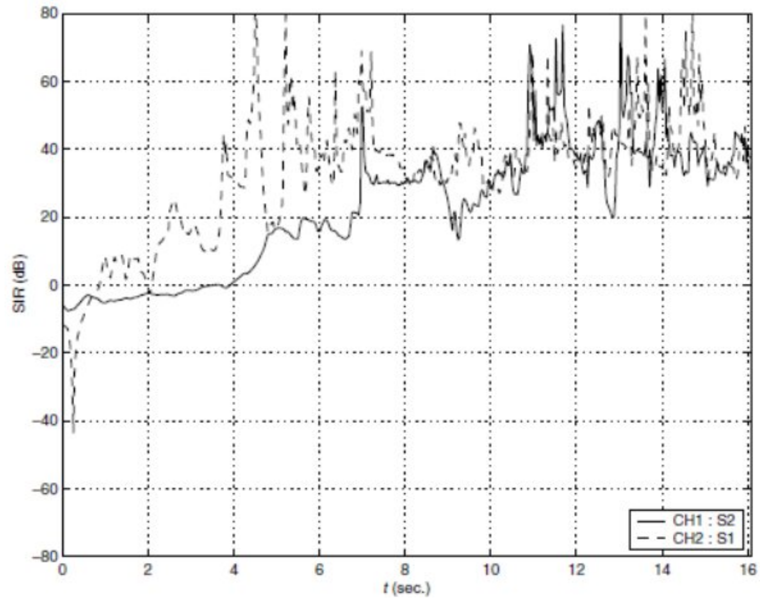


Fig 3.21

Signal-to-Interference Ratios of two source, three mixture simulation, (L<M)

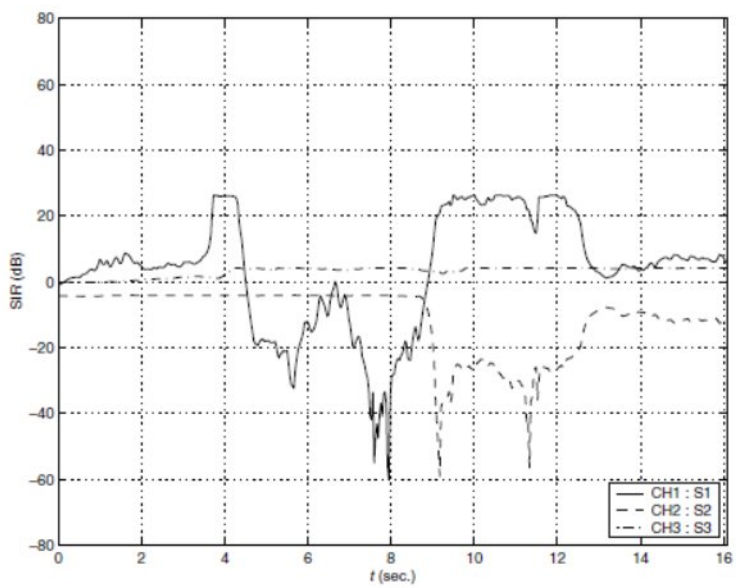


Fig. 3.22

Signal-to-Interference Ratios of three source, two mixture simulation (L>M)

This algorithm works for L<M and L=M but does not work well for L>M. In the future work the generalized algorithm needs to modified to handle L>M over-determined condition.

CHAPTER 4

PROPOSED METHOD

4.1 Introduction

To track any object first requirement is its trajectory record primarily as visual information. But at boarders where the field is at few kilometers of distances from the base stations or control rooms, only visual information is not sufficient to locate and track the object. We need to add one more information parameter about the object motion. In our work I have taken seismic signals generated by the object under motion on the ground which are recorded by seismic sensor placed in the field, with a sensing range of 30 m. In the market sensors with larger sensing range are also available but are costly for study purpose. Therefore we have limited ourselves to 30 m seismic sensors for acquisition of seismic signals. In the video the scene has been created in such a way that the object moves in the range of 30 m only. Instead of tank in the field, we have used a car and a scooter.

In this work I have divided entire approach in two building blocks

1. Seismic signal acquisition and processing
2. Visual signal acquisition and processing

And the results of these two modules are mixed together to identify that the object in the field is an intruder tank or not.

We can keep the strengths of seismic signals as templates for different objects. Seismic signal generated by a tank will be stronger than that of a scooter.

4.2 Flowchart

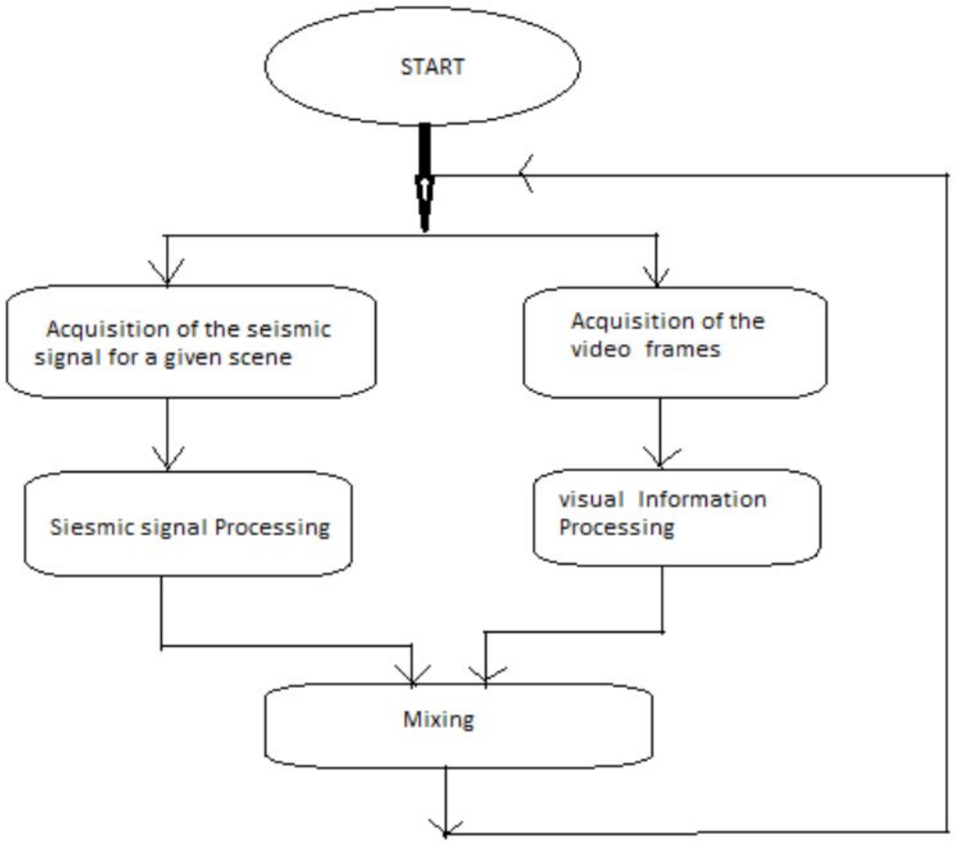


Fig 4.1 Flowchart Diagram of the proposed Algorithm

4.3 Visual Information Processing

Visual information is in the form of video frames with us, in which we have captured the object to be tracked. The tracking is achieved using the simple and efficient optimizing technique PSO (Particle Swarm Optimization). Before starting tracking basic preprocessing steps are followed i.e. morphological operations, background subtraction.

4.3.1 Flowchart of PSO algorithm

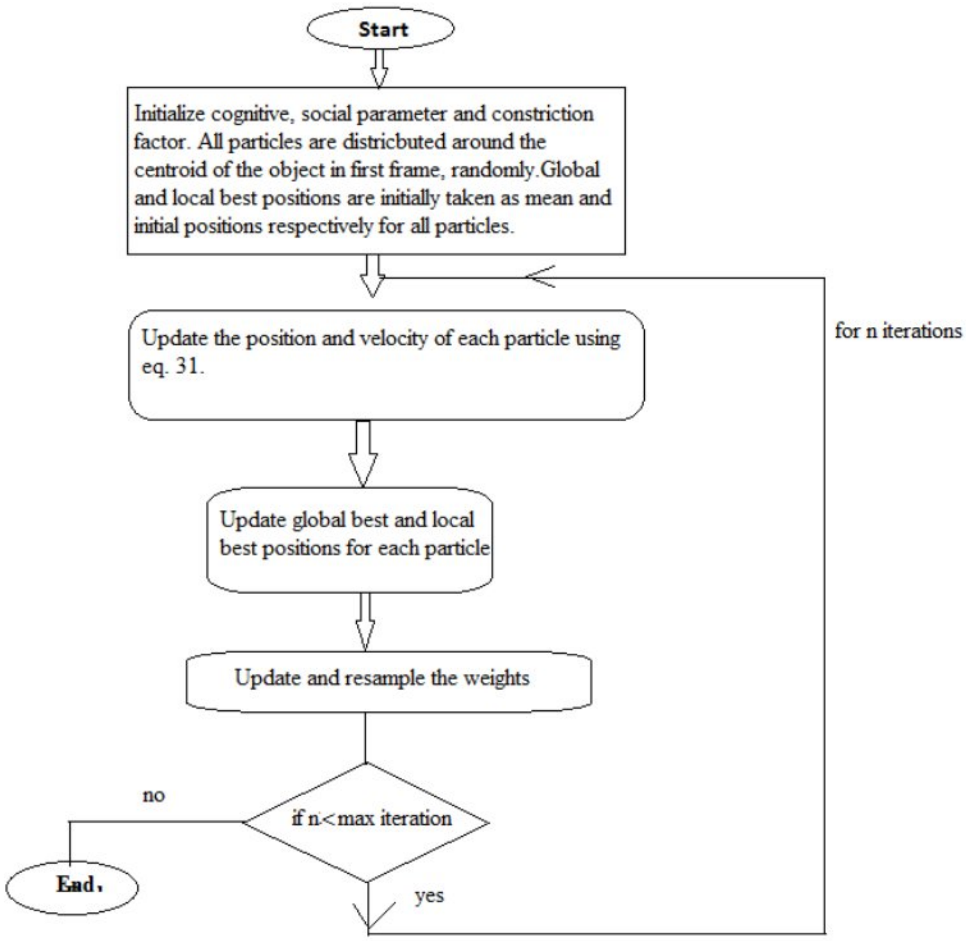


Fig. 4.2 Flowchart of PSO algorithm

4.3.1 Algorithm

Step 1 : Initialize the cognitive and social parameter, and constriction factor.

Step 2 : Perform preprocessing steps and compute area, centroid and length and breadth of the largest connected component available in the frame.

Step 3 : Distribute the particles around the centroid of the largest area of the body and assign them equal weights.

Step 4 : Initially the global and local positions are taken as the initial position of each particle. Velocity given to each particle is randomly assigned initially.

Step 5 : In the first iteration the velocity and position in both x direction and y direction are updated using standard equation of velocity eq. 31.

Step 6 : Distance of every particle's updated new position from its mean position is calculated. The position of the particle which is nearest to its mean position is taken as global best position.

Step 7 : Distance of every particle's previous position from its mean position is calculated. And the position of the particle which is nearest to its mean position is taken as the local best position.

Step 8 : Resampling of the weights is performed to avoid degeneracy.

Step 9 : Repeat Step 5 to 8 for every iteration.

End

4.4 Seismic signal Processing

In the latest paper by Raffi and Pardo ref [4], a unique method is presented for music/voice separation from a given song as a signal. Our proposed method for seismic signal processing is based on the same algorithm with some modifications in it. We are using mixed seismic signals, instead of music signal. In our algorithm we will be picking only a short duration of clip. In visual information processing the frames will be picked up after same duration, to provide synchronization between two modules. Now the short seismic signal clip is divided into four parts using 4-band Filter Bank according to the set of frequencies present in the clip. This helps to separate the part of the signal of the same type at the initial step, because two sounds will be around two different frequencies. For this individual signal out of four, third order cumulants are computed which are also called bi-spectrum, which gives information about the signal's skewness. It is symmetric or non-symmetric about the mean value. The bi-spectrum curve beat spectrum b is computed in the stage one of the algorithm. In the next stage, stage two, the period of the beat spectrum is computed after which, beat is repeating itself. Power spectrum V of the bi-spectrum is divided into sub-segments each of period equal to beat period p, after which beat information is repeating itself. Average of all sub-segments is computed and a repeating segment S is obtained. In the third stage each sub-segment of V power spectrum is compared with the repeating segment S and the minimum of two is assigned to the repeating spectrogram W. Finally, the soft mask is obtained by taking the ratio of W and V.

The same procedure is repeated for rest three parts of the short seismic signal clip. And for each of four parts of the short clip, we receive four masks. When we apply these masks on the original short clip, we can extract the unique pattern of the clip. The difference between each pattern is computed, and it is identified that each of the four patterns are same or not. If they are same, it means there is only one sound in the seismic signal clip. If there are more than one unique patterns, it means there is more than one object moving on the ground and generating the vibrations.

With every iteration a new clip of the signal is picked up and entire procedure is followed till the entire signal length is accessed. We can overview the entire algorithm in figure 4.1

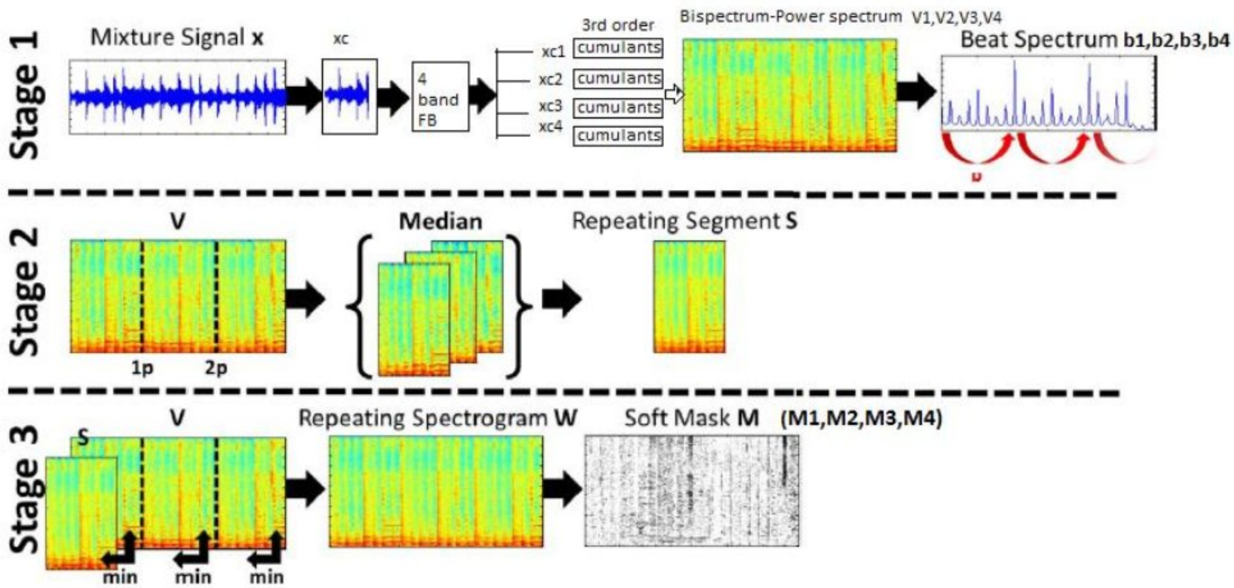


Fig 4.3

Overview of the modified REPET algorithm. Stage 1: calculation of the beat spectrum and estimation of the repeating period p. Stage 2: segmentation of the mixture spectrogram V and computation of the repeating segment model S. Stage 3: derivation of the repeating spectrogram model W and building of the soft time-frequency mask M.

where

1. xc is one short clip of the mixture seismic signal.
2. xc1, xc2, xc3, xc4 are four parts of xc divided using 4-Band FB on the bases of frequencies.
3. V1,V2,V3,V4 are the power spectrums of the bi-spectrum (Fourier Transform of the 3rd order cumulant of xci, i=1:4)
4. b1,b2,b3,b4 are the beat spectrums of the power spectrums V

CHAPTER-5
METHODOLOGY USED IN SEISMIC SIGNAL PROCESSING

The seismic signal processing algorithm is described in article 4.4. In which few parameters are a calculated i.e. 4-bands FB, Third order- Cumulants, Bispectrum, Power spectrum, Beat spectrum, period of the beat spectrum, and Mask. We will be discussing about the methods of calculating these parameters in detail, in this chapter.

5.1 4-Band Filter Bank :

The basic knowledge required for understanding the M-band Filter Bank has been discussed in chapter 2 in detail. Here we will be highlight the design of the filter bank, such that no aliasing takes place and perfect reconstruction is possible by combining the filtered segments into the original signal. In the M-band filter bank, we are aware of that, it has one analysis filter (combination of M filters) and synthesis filter (combination of M filters). Synthesis filter is designed in such a way that all aliasing effects are compensated.

For the coefficient values of analysis and synthesis filters (each M), the paper written by Trac. D Tran as mentioned in ref [5] is followed. And the coefficients for 4-Band filter bank are taken as specified in the table 2:

5.2 Third order Cumulants

Third order cumulants od a signal provides information about the skewness of the signal. For a symmetrically distributed signal x(n) skewness of x(n) is always equal to zero.

$$C_{3x}(k, l) = E\{x^*(n)x(n+k)x(n+l)\} \tag{42}$$

h0	-5/16	1/8	7/8	21/16	21/16	7/8	1/8	-5/16
h1	-61/ 512	9/256	95/256	141/512	-141/ 512	-95/256	-9/256	61/512
h2	-1/8	0	½	-3/8	-3/8	1/2	0	-1/8
h3	23/128	5/64	-61/64	249/128	-249/ 128	61/64	-5/64	-23/ 128
f0	0	1/16	3/16	¼	1/4	3/16	1/16	0
f1	5/128	-23/ 256	-249/ 256	-61/128	61/128	249/256	23/256	-5/128
f2	-1/16	-5/32	21/32	-7/16	-7/16	21/32	-5/32	-1/16
f3	-9/512	-61/ 1024	141/1024	-95/512	95/512	-141/ 1024	61/ 1024	9/512

Table 2: Ref [5] Coefficients of the 4 channel 8 tap LPPRFB, Top: Analysis Filters h_i, Bottom: Synthesis Filters f_i

Eq. 42 gives the mathematical definition of third order cumulant of signal x(n). Skewness is used to refer un-normalized quantities.

5.3 Bispectrum

$$\left|S_{3x}(f_1, f_2) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} C_{3x}(k, l) e^{-j2\pi f_1 k} e^{-j2\pi f_2 l} \right. \tag{43}$$

Bispectrum is defined as the FT of the third order cumulants, given by Eq. 43 as per Ref [6]. It is called Bispectrum because it is a function of two frequencies f1 and f2, from the expression.

5.4 Power Spectrum

Power spectrum of a signal x(n) is defined as the product of its fourier transform with itself.

Power spectrum V= S_{3x}*S_{3x} (44)

Where S_{3x} is the FT of 3rd order cumulants.

And V is power spectrum of the Bispectrum S_{3x}.

5.5 Beat Spectrum and its period

Beat has been defined in chapter 2. Beat spectrum can be computed from the given power-spectrum V of the signal using eq. 45 as per the definition in ref [4].

$$B(i,j) = \frac{1}{m-j+1} \sum_{k=1}^{m-j+1} V(i,k)^2 \, V(i,k+j-1)^2$$
$$b(j) = \frac{1}{n} \sum_{i=1}^{n} B(i,j) \qquad \text{then } b(j) = \frac{b(j)}{b(1)}$$
$$\text{for } i = 1 \ldots n \text{ (frequency) where } n = \frac{N}{2} + 1$$
$$\text{for } j = 1 \ldots m \text{(lag) where } m = \#\text{time frames.} \tag{45}$$

5.5.1 Algorithm to find the period of a beat spectrum

The algorithm is followed from ref [4].

$l \leftarrow$ length of b after discarding the longest 1/4 of lags
$\delta \leftarrow$ fixed deviation for possible shifted peaks
$J \leftarrow$ empty array of length $\lfloor l/3 \rfloor$
for each possible period j in the first 1/3 of b **do**
$\Delta \leftarrow \lfloor 3j/4 \rfloor$, $I \leftarrow 0$
for each possible integer multiple i of j in b **do**
$h \leftarrow \underset{k \in [i-\delta, i+\delta]}{\operatorname{argmax}} b(k)$
if $h = \underset{k \in [i-\Delta, i+\Delta]}{\operatorname{argmax}} b(k)$ **then**
$I \leftarrow I + b(h) - \underset{k \in [i-\Delta, i+\Delta]}{\operatorname{mean}} b(k)$
end if
end for
$J(j) \leftarrow I / \lfloor l/j \rfloor$
end for
$p \leftarrow \arg\max_j J(j)$

5.6 Mask

The Mask is computed by the ratio of the Repeating Spectrogram, and the power-

$$M(i,j) = \frac{W(i,j)}{V(i,j)} \qquad \text{with } M(i,j) \in [0,1]$$
$$\text{for } i = 1 \ldots n \text{ (frequency) and } j = 1 \ldots m \text{ (time)} \tag{46}$$

spectrogram of the Bispectrum as per the definition specified in ref [4]. When mask is multiplied with a sequence of the signal, the non-repeating pattern will be extracted from the sequence.

CHAPTER 6

SIMULATION AND RESULTS

It has been mentioned that our both visual signal processing and Audio Signal processing are executed in parallel. And later from the results of two modules the final conclusion is made that there is a threat on the field by the tracked object or not.

When Visual Signal Processing Module is simulated certain parameters are required to be initialized.

Number of particles in the swarm :10

Cognitive parameter c1: 0.1

Social Parameter c2 : 2-c1

Constriction factor : 1

Maximum iterations swarm has to make before reaching to the next position : 10

Positions of particles distributed around the mean position are defined as

```
x0=mean(1,1)+0.0001*randn(10,1);
y0=mean(1,2)+0.0001*randn(10,1);
```

variance =0.0001;

The Results of the PSO algorithm are shown in fig. 6.1. where when two objects are detected, using preprocessing operations, we have selected only the larger area object. Because our objective is to find the tank with a threat, which is of large size. So even if we have two objects in the frame we will be tracking only the bigger one. In the fig 6.1 the blue particles coinciding at one point on the car shows the accuracy of the tracking algorithm. Particles, visible in the frames 210,370,450,570 , are placed at the estimated values of position of the swarm through an algorithm.

Fig 6.1 PSO algorithm based visual tracking result when two objects are moving

In fig 6.2 the percentage of estimated mean square error is obtained, while estimating the new position of the Particle Swarm, which is a very small quantity, showing that the swarm did not mislead while tracking the car.

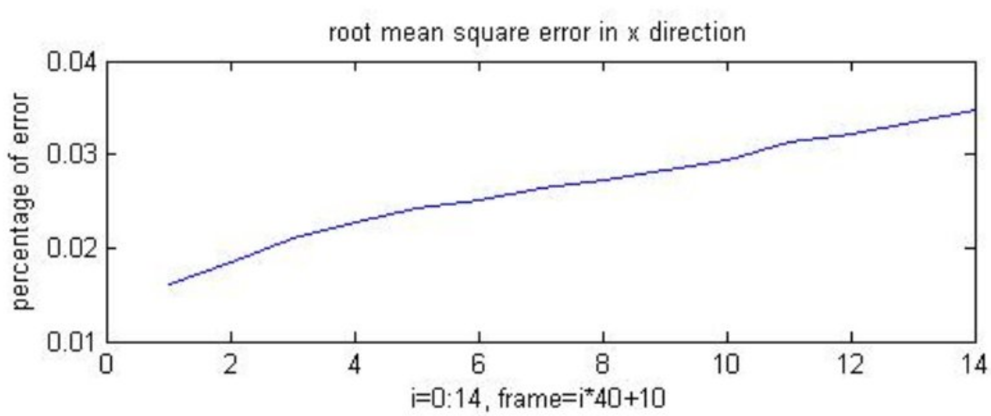


Fig 6.2 root mean square observed at every iteration from frame to frame

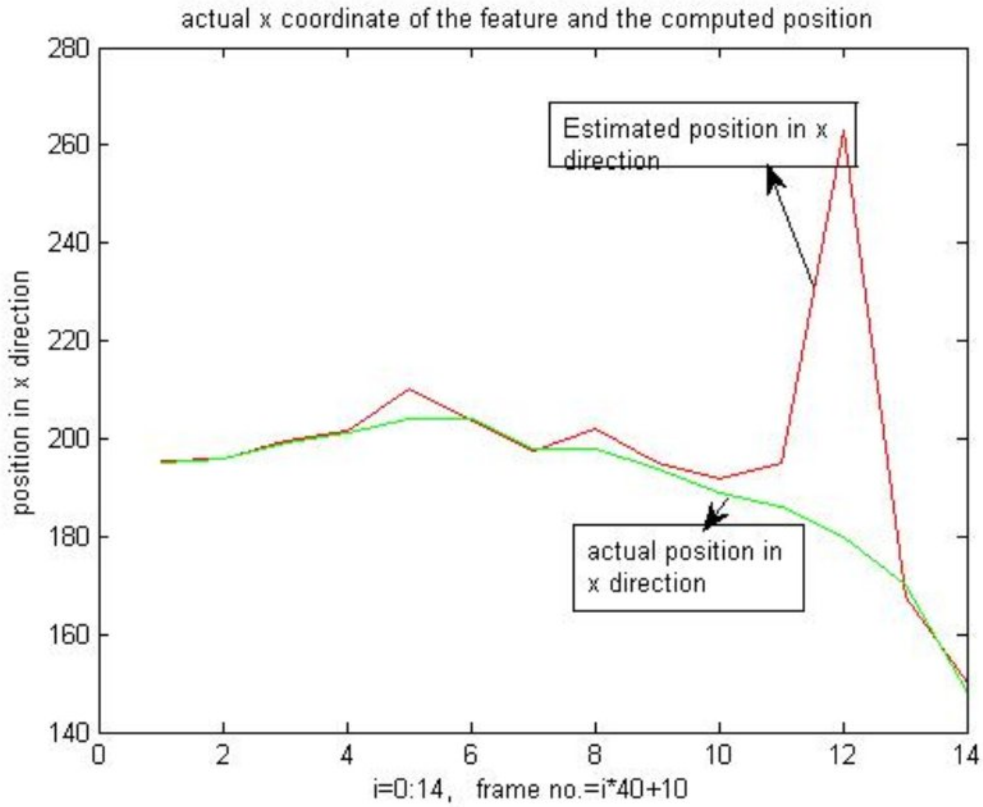


Fig 6.3 Actual path (green curve) and predicted path (red curve)
followed while tracking in x direction

Fig 6.3 is the curve between the actual position of the object in every frame excessed sequentially, and the estimated position of the swarm in every frame.

Results of Audio Signal Processing Module:

The audio sample rate is 5626.7 Hz.

A clip is made for 2 secs.

Therefore in a clip sample rate is 11253.4Hz.

Each clip is passed through a 4-channel FB and each filtered signal beat period is computed to separate the unique part of the each signal clip.

At the end all the clips unique patterns are joined together. And identified how many unique patterns are available.

If the are two objects on the ground only two unique patterns will be produced by the algorithm.

```
p_arr = [  65   129   129   129 ;    66   137   137   138 ;    90   137   137   133;
            66    78    78    78  ;    31   138   138   138;     66   125   125   118;
           70   133   133   134 ;    66    92    92    92  ;    66   136   136   119
           66   138   138   138 ;    64   135   135   137;     63   138   138   138;
           66    66    66    66  ;     66    66    66    66];
```

p_arr is as array of collection of four bands beat period value, for 13 voice clips of the entire voice/ seismic signal.

Same value of the period in all four bands of a clip represents that, that particular clip is symmetric in nature.

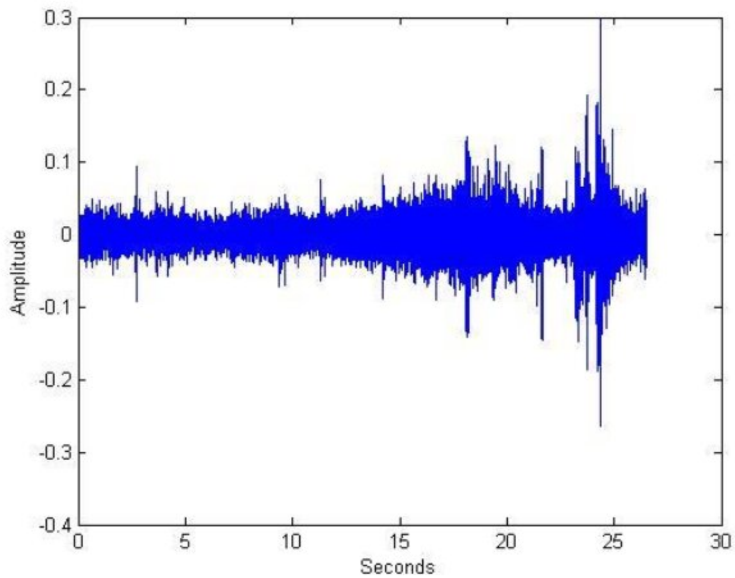


Fig 6.4 Original seismic signal for 26 secs long and 5Khz frequency, generated due to both car and scooter.

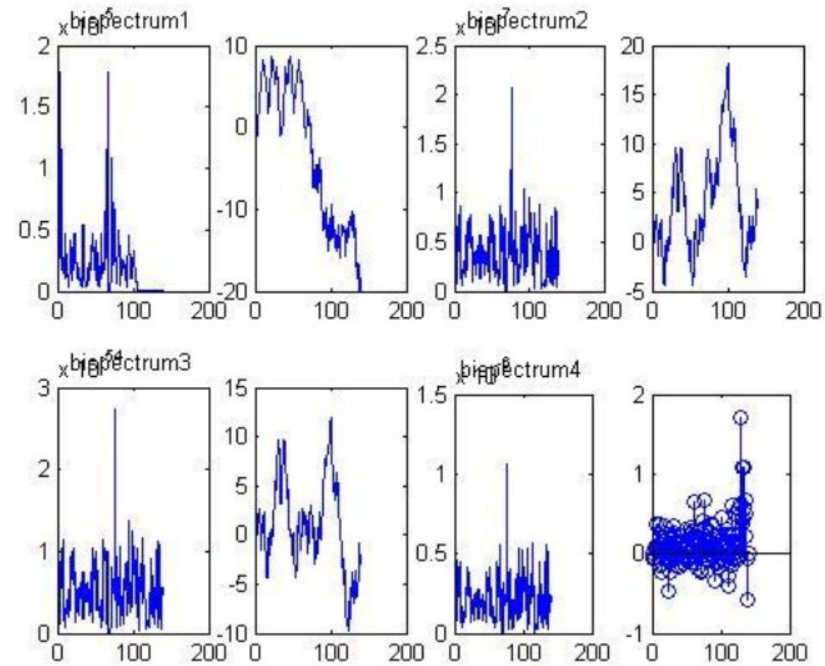


Fig 6.5 Bispectrum magnitude and phase curve of the
4-bands filtered signal of clip no. 2

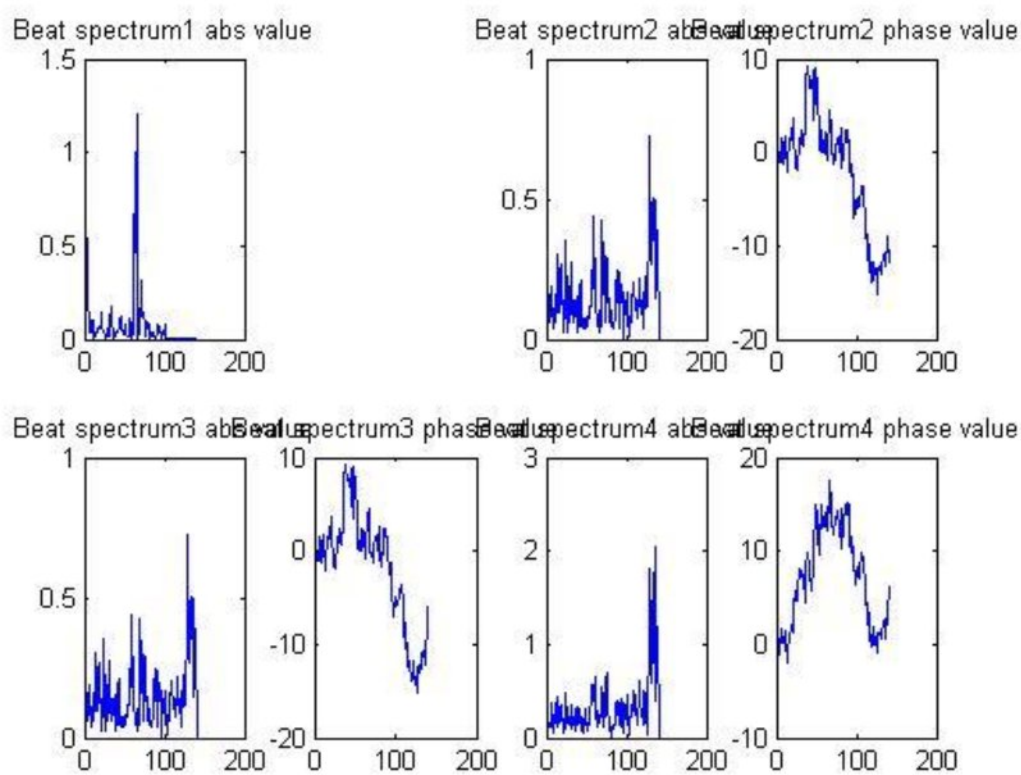


Fig 6.6 Beat spectrum magnitude and phase curve for 4-band filtered signal of clip no 2

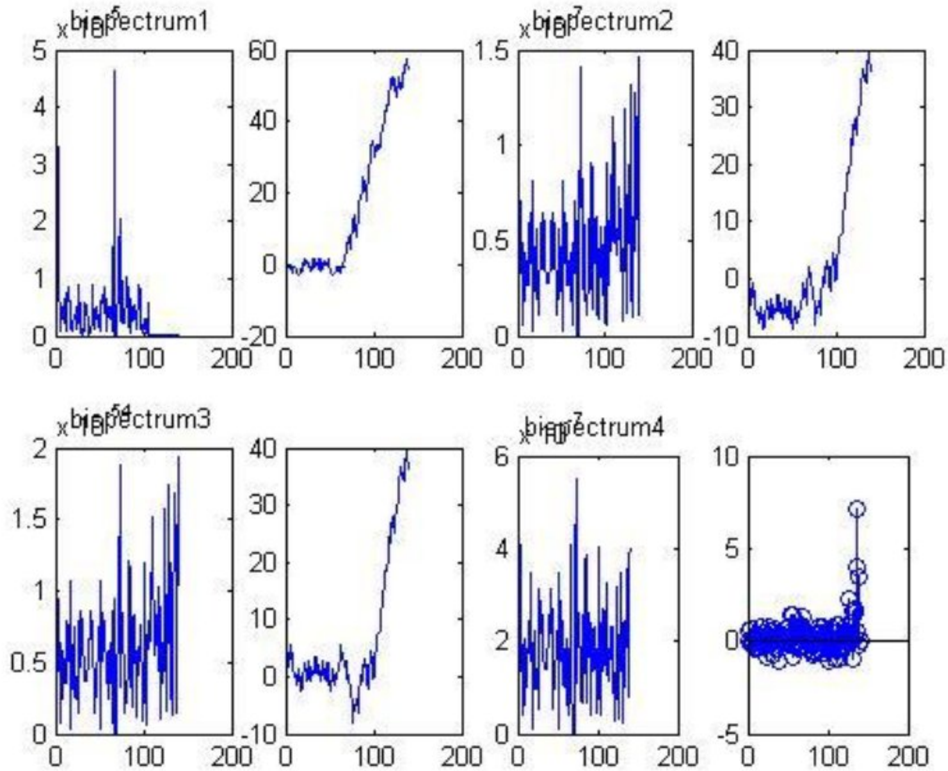


Fig 6.7 Bispectrum magnitude and phase curve of the
4-bands filtered signal of clip no. 8

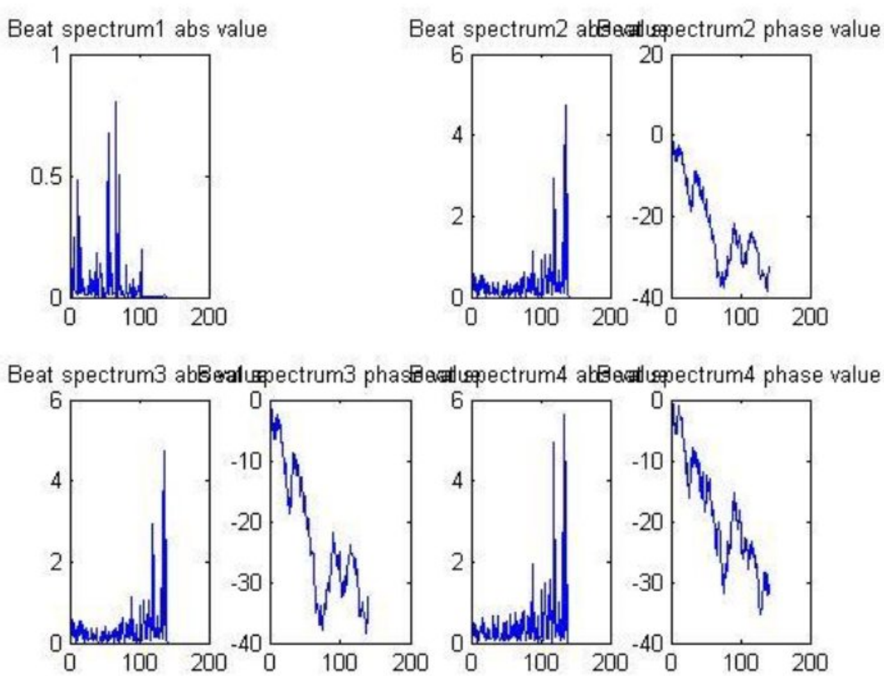


Fig 6.8 Beat spectrum magnitude and phase curve for 4-band
filtered signal of clip no 8

From p_arr matrix if extract the mask and the corresponding signal and join all the four band signals pattern for every clip. For all 13 clips we can observe that only two patterns are repeating. Hence for the given video seismic signal two objects are identified.

Mixing Module

Case1: In the mixing module the results obtained from audio and visual signal processing, that there are two objects, next task is to identify out of two there is (practical case :tank) car present. If yes, then the from visual signal tracking the large area object tracking is justified.

Case 2: If there is only one object in the screen, from visual signal it can be easily tracked. But from seismic signal it needs to be identified that the number of objects identified in the screen are same as actually moving on the ground, which is detected by finding the unique pattern to be only one. And then comparing it with the tank/car seismic signal template, to verify that the single object which is being tracked is tank only. Then an alarm can be made of threat, that there is only one tank in the field, which needs to be checked.

CHAPTER 7
CONCLUSION

It can be concluded that the proposed algorithm using two modalities to ensure security at boarders proves to be better approach rather than relying only on visual information. If the visual information is not clear, the fast and efficient Repeating pattern algorithm generates accurate results about the number of objects in the field and also the type of object.

7.1 Future work

In the future work the mixing module can be made much more efficient, by applying array processing concept with the two algorithms. The same algorithm can be worked out for moving cameras or foggy environment where the visual information captured is not that clear. Image enhancement and restoration module can be added into it.

CHAPTER-8

References:

[1] E. Maggio and A. Cavallaro. Hybrid particle filter and mean shift tracker with adaptive transition model. Acoustics, Speech, and Signal Processing, 2:221–224, 2005. 3, 4

[2] P. V. William and M. W. Iii. Alignment by maximization of mutual information. International Journal of Computer Vision, 24(2):137154, 1997. 5

[3] http://arduino.cc/en/Main/arduinoBoardDuemilanove

[4] Zafar Rafii, student Member, IEEE, and Bryan Pardo, Member, IEEE , "Repeating Pattern extraction Technique (REPET): A Simple method for Music/Voice Separation. IEEE Transaction on Audio, Speech and language processing, vol 21 no. 1 Jan 2013

[5] Trac D. Tran, Member IEEE, "M-channel Linear Phase Perfect Reconstruction Filter Bank with Rational Filter". IEEE Transactions on circuit and systems-I, Vol-49,N.7, July 2002

[6] *Ananthram Swami, Jerry M. Mendel ,Chrysostomos L. (Max) Nikias* Higher Order Spectral analysis (HOSA) Toolbox, for use with MATLAB.

[7] *J Control Theory Appl* 2013 11 (1) 42–53, DOI 10.1007/s11768-013-1088-0, "Object tracking based on particle filter with discriminative features", by Yunji ZHAO South China University of Technology, Guangzhou Guangdong 510640, China

[8] International Journal of Advanced Robotic Systems "Object Tracking Using Maximum Colour Distance Under Illumination Change" Department of Computer Science and Engineering, Chung-Ang University, Seoul, Korea, author E-mail: knchoi@cau.ac.kr DOI: 10.5772/52860, 2012