DEVOLOPMENT OF MEAN AND MEDIAN BASED ADAPTIVE SEARCH ALGORITHM FOR MOTION ESTIMATION IN SNR SCALABLE VIDEO CODING

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF TECHNOLOGY

IN

ELECTRONIC SYSTEMS AND COMMUNICATIONS

BY

RAJKUMAR MAHARAJU

ROLL NO. - 212EE1289

Department of Electrical Engineering

National Institute of Technology, Rourkela-769008

2015

DEVOLOPMENT OF MEAN AND MEDIAN BASED ADAPTIVE SEARCH ALGORITHM FOR MOTION ESTIMATION IN SNR SCALABLE VIDEO CODING

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

MASTER OF TECHNOLOGY

IN

ELECTRONIC SYSTEMS AND COMMUNICATIONS

BY

RAJKUMAR MAHARAJU

ROLL NO. - 213EE1289

Under the Guidance of

PROF. DIPTI PATRA

Department of Electrical Engineering

National Institute of Technology, Rourkela-769008

2015

National Institute Of Technology, Rourkela

CERTIFICATE

This is to certify that the thesis entitled "**DEVOLOPMENT OF MEAN AND MEDIAN BASED ADAPTIVE SEARCH ALGORITHM FOR MOTION ESTIMATION IN SNR SCALABLE VIDEO CODING"** submitted by **MR. RAJKUMAR MAHARAJU** in partial fulfillment of the requirements for the award of Master of Technology Degree in **Electrical Engineering** with specialization in **"ELECTRONIC SYSTEMS AND COMMUNICATION"** at National Institute of Technology, Rourkela is an authentic work carried out by him under my supervision and guidance.

To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other University / Institute for the award of any Degree or Diploma.

Date: **Prof. Dipti Patra**

Department of Electrical Engineering

National Institute of Technology

Rourkela

ACKNOWLEDGEMENT

Firstly, I would like to thank Prof. Dipti Patra, Department of Electrical Engineering, National Institute of Technology, Rourkela as my guide, who encouraged and challenged me throughout my project work and guided me on writing conference papers, thesis and to put effort on our workings. She is a perfect motivator and I am glad to have guide like her.

After that, I would like to thank, Prof. Anup Kumar Panda, head of department, electrical engineering, National Institute of Technology, Rourkela and professor of my specialization Prof. Prasanna Kumar Sahu, Prof. Susmita Das, Dr. Supratim Gupta and Prof. K. Ratna Subhashini.

I would also like thanks my fellow lab mates in Image Processing & Computer Vision Lab: Pranav, Harsha, Bedadatta, Smita Pradhan, Rajashree Nayak, Umesh, Sushree and Yogananda Patnaik for their help and support.

Rajkumar Maharaju

Abstract

Now a day's quality of video in encoding is challenging in many video applications like video conferences, live streaming and video surveillance. The development of technology has resulted in invention of various devices, different network conditions and many more. This has made video coding challenging day by day. An answer to the need of all can be scalable video coding, where a single bit stream contains more than one layer known as base and enhancement layers respectively. There are various types of scalability as spatial, SNR, temporal scalability. Among these three types of scalability, SNR scalability deals with the quality of the frames i.e. base layers includes least quality frames and enhancement layer gets frames with better quality. Motion estimation is the most important aspect of video coding. Usually the adjacent frames of a video are very much similar to each other. Hence to increase the coding efficiency to remove redundancy as well as to reduce computational complexity, motion should be estimated and compensated. Hence, in the scalable video coding, videos have been encoded in SNR scalability mode and then the motion estimation has been carried out by two proposed methods. The approach depends on eliminating the unnecessary blocks, which have not undergone motion, by taking the specific threshold value for every search region. It is desirable to reduce the time of computation to increase the efficiency but keeping in view that not at the cost of much quality. In second method, the search method has been optimized using 'particle swarm optimization' (PSO) technique, which is a method of computation aims at optimizing a problem with the help of popular candidate solutions. In block matching based on PSO, a swarm of particles will fly in random directions in search window of reference frame, which can be indexed by the horizontal and vertical coordinates of the center pixel of the candidate block. These algorithm mainly used to reducing the computational time by checking some random position points in the search window for finding out the best match. PSO algorithm estimate the motion with very low complexity in the context of video estimation. Both the methods have been analyzed and performance have been compared with various video sequences. The proposed technique out performs to the existing techniques in terms of computational complexity and video quality.

Contents

List of figures

List of tables

CHAPTER 1

Introduction

1.1 Video coding

Video coding has been a very active research area for a long time. While transmission bandwidth and storage capacity have been growing dramatically, the demand for better vide coding technology has also been growing, the reason is the ever-increasing demand for higher quality of images and video, which requires ever-increasing quantities of data to be transmitted and/or stored. Again various type of devices with different resolution and displaying capability made the video coding more challengeable.

Now a days we availed number of video coding techniques. While considering the turn coding we refer to compression of the visual signals in this thesis along with the techniques for conventional coding also we have focused on scalability profile in video coding in order to make it more reliable for various applications [2][10]. Basically we aim at declining the data needed to represent a video without losing quality with the prospective of biological vision. But the issue is the video signals receive by the receiver cannot be replicated as the original signals given to the encoder due to inclusion of some distortion in between.

On the other hand the delay, which can be defined as the time difference between an image or video frame is at the encoder and the reconstructed video present at the decoder can stand as a critical for two way communications for such applications delay is the fourth dimension in video coding, This delay can be circumvented by utilizing motion estimation in this thesis along with scalability in video coding we also discuss how to perform motion estimation with minimum computation time along with less degradation in the video quality.

1.1.1 Video encoder
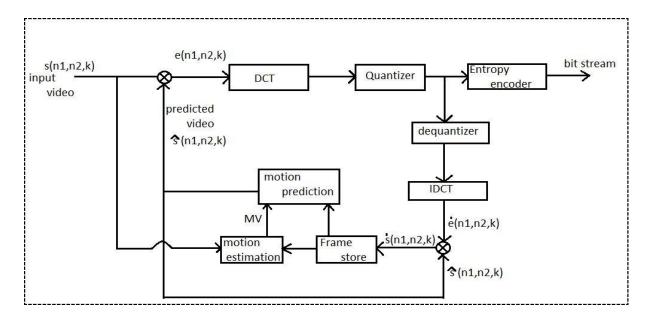
Figure 1.1 Block diagram of video encoder

Figure 1.1 describes the method of execution of a video encoder basically encoder takes a raw video as input the output generated is a compressed bit stream. The decoder works reversely i.e. it is given the compressed bit stream as input then it attempts to rebuild the original raw video given as input to the encoder the blocks shown in the finger have been illustrated as follows. The input is (n1, n2, k) where the symbols signifies ask third dimensions of time n1, n2 are horizontal and vertical dimensions [13]. If the process of video coding will be classified broadly it will be divided in to two sections one interframe mode of encoding second one intraframe mode of encoding extending further to the first mode of operation i.e. interframe mode can be described as the methodology of prediction of the error exist in between the blocks of the a single frame feedback loop has to be applied, the scheme of video encoding basically can be initiated with preprocessing. This preprocessing can be demising spatial resolution conversion etc.

 This aims at increasing the quality of service so that the end user terminals and the application targeted to be achieved can be served better [2] [12]. Also it enhances the coding efficiency remarkably but as this is a lossy one. The video signals after going through preprocessing should be go to the one-to-one stays the transformation so that visual data can be suggested to an efficient mode of compression. By utilizing the above said, the redundancy and the correlation exist

between the transformed coefficients can be cut down up to a remarkable extant the statistical distribution of the coefficient can be shaped to make subsequent entropy more efficient. There exists various kind of transformation such as differential the predictive mapping discrete cosine transform (DCT), wavelet transform etc. As collaboration of this technique gives a refined output, improved systems of compression use a hybridization of the above said and also supports a wide variant in the modes present at the stage of transformation in order to achieve adaptive decomposing of a visual signal. The module of quantization where the coefficients are mapped into a set of discrete values consequently need fewer vets the represent theses coefficients. Quantization in encoding is referred as a lossy process as it attains the loss of a remarkable amount of information. When the compression is lossy one it can be controlled by adjusting the size of the tape of quantization. The process of quantization can be classified broadly as scalar quantization (SQ) and vector quantization (VQ) the scalar quantization can be defined as the process of quantization where the quantization of the coefficients is being carried out independently.

One the other hand we can defined better quantization as process of quantization where various coefficients presents are made into clusters and then undergoes quantization cumulatively. Another important process in video encoding is entropy coding which referred as process which can be carried out either by using the statistical schemes of by utilizing the methods which are distortionary based. In encoding we used many coders, but some of the frequently used coders can be enlisted as Huffman coder, Arithmetic coder etc.

1.2 Scalability in video coding

Video coding has become challengeable day by day because of rise in number of internet users along with a huge group of people having various devices with different resolutions. This results in a difficult issue when number of users attains to access a single video though various links of communications. To elucidate the problem stated above more clearly an example can be taken. Suppose there is a mpeg1 video this can be downloaded at the speed of 1.5 mbps in real time for a playback applications in a user terminal which is contacted to a server via high speed link lets consider another user who uses a modem with the connection of 56kbps so he cannot get sufficient amount of the number of bits in time for the playback in real time. To answer the problem raised the concept of scalability in bit stream of vides is required this concepts uses the idea of achieving the ability to reconstruct the meaning full visual information after the decoding just the bit stream

compressed partially. With the concept for scalability lets loop back again to the example stated above. If the video stream utilized in the problems stated is scalable, end the user having high speed connection will be able to download the full video with a good quality but the user having low speed connection will able to avail only a subset of the entire bit stream and get a visual presentation of lower quality

The example stated here put light on the one direction of scalability the deals with bandwidth and scalability. A scalable bit stream can have various adaptive variations it can support Scalabling a bit stream with many aspects it also proved itself resilient to error and also holds good during transmissions for graceful degradation.
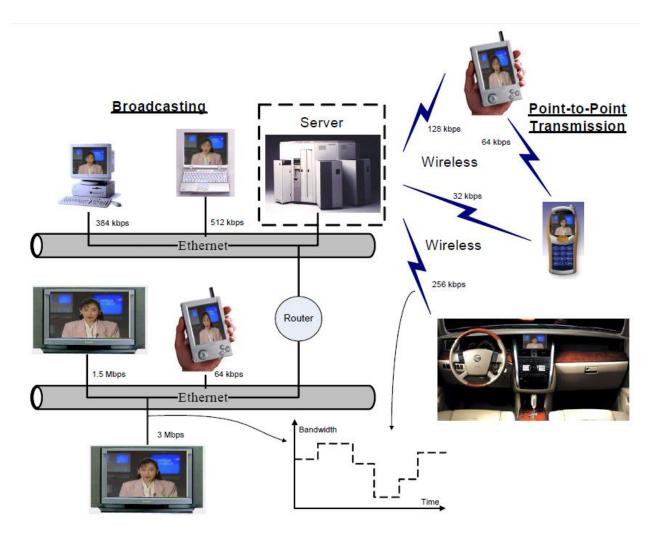
Figure 1.2 Application Frame work for Scalable video coding

The figure 1.2 gives a pictorial description of the basic frame work of scalable video coding applications which is subjected to number of fields such as internet transmissions where scalability supports variations in the transmission bitrate, discarding the selective bits as well as adjustment of the source rate in contrast to modem rates, variations in bandwidth of different channels and the diversity in the ability of different devices [5]. Extending further availability of internet, large use of multimedia applications, development in wireless technology have made the concept of scalability of major important which makes the users able to access the desired media application anywhere any time and any devices. The encoders used in scalable applications can be divided in to two groups based on granularity i.e. coarse granularity or fine granularity.

When the fine granularity is at its extreme, the truncation of the bit stream can be carried out at any point the number of bits retained is directly proportional to the quality of the image reconstructed. These kind of bit streams can be called embedded. When the coders are embedded we can expect various functionality such as control of bit streams in a precise manner which is advantageous for lot of applications. Then network filters can pick some of the selected number of bits for transmission from a bit stream which is embedded. In order to match with the bandwidth availed. Scalability in video coding can be included if for a single video various versions have been provided from perspectives like amplitude resolutions, spatial resolutions, temporal resolutions, frequency resolutions etc. The hybridization of all can also be a good alternative.

We can achieve scalability at frame level or object level the inclusion of scalability in video coding makes the encoder to lose some coding gain in contrast to the coders which are not scalable more over the scalable coding is designed with an intention to decline the coding efficacy degradation but still realizing the need of scalability.

1.3 Motion estimation and compensation

A video is a combination of temporal samples collected over a certain amount of time this sequence. Hence, a video sequence is a collaboration of number of frames. As we are dealing with a signal varies time it is obvious that the frames are subjected to a huge amount of similarity this similarity is mode when the video is in slow motion and less if the sequence moves fast.

Since each frame in video sequence is of very short period of time hence a great deal of similarity is expected referred as redundancy [2] [11]. To eliminate this motion estimation and compensation

can we applied. Consequently it gives a remarkable gain in coding. Motion compensation is a way followed to exploit the redundancy from frame to frame by compensating the moment in the objects whereas estimation can be defined as gathering the information about the amount of displacement of the objects which is accomplished by obtaining the motion vector. The pictorial representation of it has soon below.
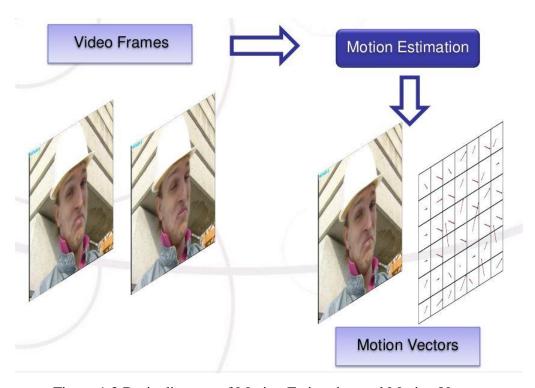
Figure 1.3 Basic diagram of Motion Estimation and Motion Vectors

There are many motion estimation algorithms available, for example, block matching where the motion vector is obtained by searching through all the possible positions according to a matching criterion.

1.3.1 Motion vector prediction

Motion vectors play a great role in estimating the motion it contains vary significant amount of information which is inevitable for decoding with a good amount of coding gain. Before transmitting the motion vectors to the decoder side a prediction process should be executed by utilizing the predecessors i.e. the adjacent motion vectors presents at upper and left positions. The encoders follow a number of ways for the prediction of the motion vector. This prediction can be done on the basis of availability of the blocks adjacent to it. If one wants to follow the prediction

of motion vector in a direct mode, it can be done by using either the motion vectors manifested out of the adjust blocks in spatial manner or using the motion vectors from the blocks of the frames used for reference.

1.4 Motivation

The advancement in technology has presented as many multimedia tools starting from mobile Tv to video telephonic and video conferencing etc. also the raise in the number of internet users has created a huge group of users demanding multimedia applications in better quality. Independent of the device capability, bandwidth availability at any place at any time.

When attempt has been made by number of users to access the same video through various link of communication established, it becomes a difficult problem to serve the clients a better quality of service.

Along with it sometimes the visual signals are transmitted to the networks future not reliable and subjected to a variance in condition of transmission. These things can be managed in a graceful manner with the addition of scalability in vide coding. Extending further the network infrastructure can be different and also the devices used for decoding can have variety in displaying capacity. The computational complexity is also an obstacle in video encoding while subjected to real time implementation. Moreover to overcome such problems we need a standard of video coding which is flexible enough to adapt the situations properly.

1.5 Literature review

Marta Mrak, Ebroul Izquierdo [3]. Scalable video coding targets seamless delivery of and access to digital content, enabling optimal, user-centered multi-channel, and cross-platform media services, and providing a straightforward solution for universal video delivery to a broad range of applications.

Shipeng Li,Weiping LI [2]. They had explained Basic fundamentals, color space, Pixel Quantization, video scanning and frame rate, Basic principle of Image and video coding, Entropy coding, motion estimation and compensation .etc.

Jillani, Rashad, and Hari Kalva [4]. SVC design includes Spatial, temporal, and fidelity scalability. Spatial scalability and temporal scalability describe cases in which subsets of the bit stream represent the source content with a reduced picture size or frame rate respectively. With quality scalability, the sub stream provides the same spatio-temporal resolution as the base layer bit stream, but improves fidelity where fidelity is often informally referred to as signal to–noise ratio(SNR).

Iain E. G. Richradson [11]. Block matching algorithm (BMA) can be used for finding out the motion vectors and implementing it in very efficient way. In this first, divide the blocks in to non-overlapping blocks which is sized 16×16, 8×8 and also even 4×4, then each block will search by block wise in the reference frame until the finding out the best match. Then calculate and generate the motion vector.

In **Nasir D. Memon, Khalid Sayood [13],** they had done the research to achieve the lossless compression between the successive frames. In this paper they explained and used both spatial and temporal correlation for designing an adaptive video compression technique.

Avijit Kundu [14], had developed a modified search algorithm for diamond search method in motion estimation technique, and explained briefly about different existing search methods such as Full Search (FS), Adaptive Dual Cross Diamond Hexagonal Search and Three Step Search (TSS).

Xiaolin Chen, Nishan Canagarajah [15], this paper describes lossless compression scheme by using adaptive backward pixel to pixel based on fast predictive motion estimation. In this method it doesn't require of block motion estimation technique, it search for best match by pixel by pixel to generate the motion. In this method gives good results than Full search in terms of computation time

1.6 Thesis object and contribution

The main theme of Thesis is as follows:

- To increasing the quality of video in a single bit stream by generating the more and more enhancement layers with improving the quantization value until it's getting the original quality of video at the output of encoding which is input of decoding.

- To design a block matching technique which is having a new search algorithm that is eliminate to search all unnecessary blocks, where motion could not occur and gives accurate search points to find a best block for generating motion vectors without degradation of much video quality and within less computational complexity.

- To design a block matching based on particle swarm optimization (PSO) in Scalable video coding, a swarm of particles will fly in random directions in search window of reference frame, which can be indexed by the horizontal and vertical coordinates of the center pixel of the candidate block. These algorithm mainly used to reducing the computational time by checking some random position points in the search window for finding out the best match.

- Study and comparison of different Scalability types such as SNR, Spatial, Temporal, and Frequency in various video sequence such as cif, qcif and avi.

- Comparison the Performance of Signal to noise ratio (SNR), Peak Signal to noise ratio (PSNR), Structural similarity index (SSIM) in various video sequence.

- Study and comparison of different Motion Estimation techniques in various video sequence in terms of PSNR (dB), SNR (dB), SSIM and computational time.

- The popular swarm based search algorithm, i.e. Particle Swarm Optimization and new search algorithm has been implemented for motion estimation in scalable video coding.

CHAPTER 2

Background Theory

2.1 Scalable video coding

The scalability concept comes up with an encoder which is able to provide multilayer bit stream where each layer differ from each other in many aspects such as resolution in spatial domain, various temporal resolution etc. The layers can be referred as base layer and enhancement layer where the base layer includes the minimum information needed to construct the sequence, the enhancement layers contains the information to raise the output. Moreover from the base layer to enhancement layer the resolution of the video sequence increases gradually. The scalability can be executed in various modes which are as follows. The first one, called spatial scalability, enables encoding one video sequence represented by several spatial resolutions into one bit stream. Depending on the channel capacity the decoder may receive and decode only part of the bit stream or the whole bit stream, achieving lower spatial resolution video sequence or full spatial resolution video sequence respectively.

Another one is temporal scalability which enables to produce a bit stream in such a way that the decoder may decode from part of the bit stream a video sequence with reduced frame-rate. It means that, for example, every second frame was dropped and remaining frames have full spatial resolution and image quality. The last one is quality (SNR scalability) scalability which enables to generate an encoded bit stream from the input video sequence which contains layers, each representing a different image quality. The output video sequence quality depends on the number of layers the decoder is capable to receive. The figure shown gives the description about the decoders working one the principle of scalability.
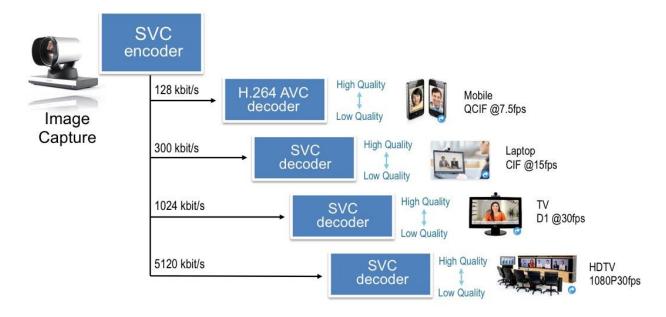
Figure 2.1 Different scalability Decoders

As shown in the figure 2.1 a sequence has been captured and given to the svc encoder which made it scalable and distribute in multiple layers [5]. When given to the decoder, the decoded the sequence according to the users need suppose a mobile having QCIF@7.5 fps is sent the lowest layer i.e. 128kb/se. according to the ability of the devices the bit streams are being sent. As shown in the figure 300kb/s, 1024kb/sec, 5.20kb/sec have been sent to the laptop, Tv, and HDTV respectively which are having frame rates as follows. CIF@15fps, D1 @30fps, 1080p 30fps.

2.1.1 SNR/Quality scalability

Out of the various mode of scalability SNR scalability is of prior importance this is also known as quality scalability it can be defined as the process to represent of video with various accuracies in the pattern of colors to achieve this one needs to quantized the values of color in the original or transformed domain along with refinement in the step size of the quantization [1][6]. Here the variations in the accuracy in the quantization can results in various PSNR values among the video gone through quantization and the raw video taken as input basically the decoding of the base

layers is a source for the outcome of a lower quality video. And other hand the decoding of the available enhancement layers bring up the quality of the reconstruct vide0 sequences the first layers is a resultant of course quantization subjected to the raw frame. The next layer is an outcome of the difference of the quantization among the frame given as in input and the same i.e. rebuilt from the very first layer, by implementing the quantization which is finer than the one being utilize to construct the first layer.

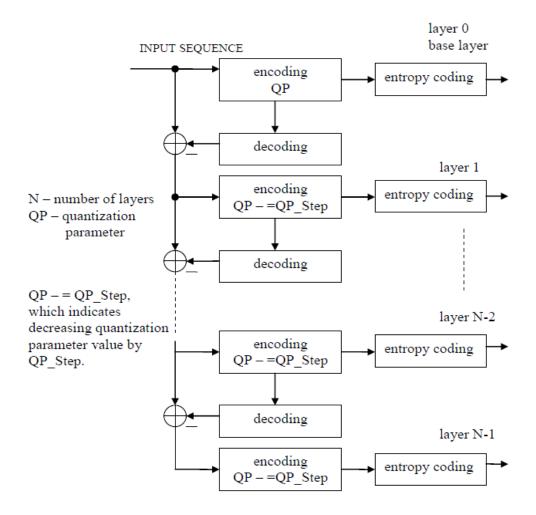
Figure 2.2 Encoding in SNR scalability

An encoder with two-level quality scalability is depicted as shown in Figure 2.2. For the base level, the encoder operates in the same manner as that of a typical block-based hybrid coder. For the enhanced level, the operations are performed as follows:

1. The raw video frame is given as an input.
2. Then it has gone through transformation by using DCT.

3. Afterword, it is quantized at the base level
4. The coefficients obtained at the base level can be constructed again by executing inverse quantization on it.
5. Then the DCT co efficient available at the base level gets subtracted from the available original coefficients of DCT.
6. The residual obtained in the previous step is identified and later goes through quantitation using a quantization parameter which is smaller than the parameter used in base level.
7. The bits which have gone through quantization are encoded using variable length coding.

Since the enhanced level uses smaller quantization parameter, it achieves better-quality than the base level. The decoder operation is depicted as shown in Figure 2.3. For the base level, the decoder operates exactly as the non-scalable video decoder. For the enhanced level, both levels must be received, decoded by variable length decoding (VLD), and inversely quantized. Then the base-level DCT coefficient values are added to the enhanced-level DCT coefficient refinements. After this stage, the summed DCT coefficients are inversely DCT-transformed, resulting in enhanced-level decoded video.
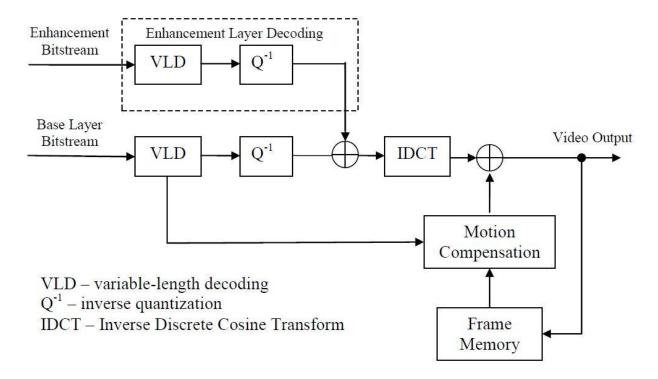

Figure 2.3 Decoder of SNR scalability

Bit plane coding can also be incorporated in scalable profile of video coding. Extending further the SNR scalability can also be achieved by using the same. The basic difference between the method followed conventionally and the biplane coding technique is. where the former one consider the coefficients obtained from the transformed as a matrix of two dimensional along with integer values and the later one considers the coefficients as several two dimensional matrices. For constructing one matrix there will be a need of one bit values [6]. Explaining future these one bit values can be defined as the bits which are adjust to each other for presenting every coefficients in a binary format. In order to make this perceptional clear lets take an example, suppose where we are having 8x8 DCT block and for the block the bit plane referred as and array consisting of 64 bits. The coding of the biplane and the matching pursuit coding of image residue are the methodology for achieving the scalability in the mode of Fine granularity known as Fine granular Scalability. This can be described more precisely by the Figure 2.4 shown.
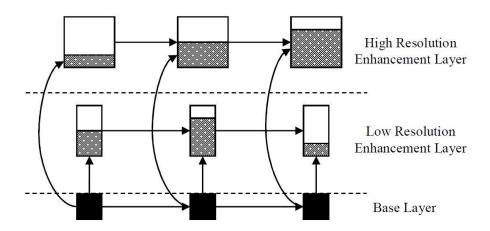
Figure 2.4 SNR FGS scalability

FGS scalability can be also achieved by macroblock reordering. The reordering is done in order to encode into a bit stream first the most important macroblocks and later less important ones. The most important macroblocks represent part of the image which is subjectively more important for human observer. So, this technique of FGS tries not to lose, after the bit stream cut, subjectively the most important parts of the coded picture.

2.1.2 Spatial scalability

Spatial scalability is defined as representing the same video in different spatial resolutions or sizes as shown in Figure 2.5. By decoding the first layer, the user can dis-play a preview version of the decoded image at a lower resolution. Decoding the second layer results in a larger reconstructed image. Furthermore, by progressively decoding the additional layers, the viewer can increase the spatial resolution of the image up to the full resolution of the original image.
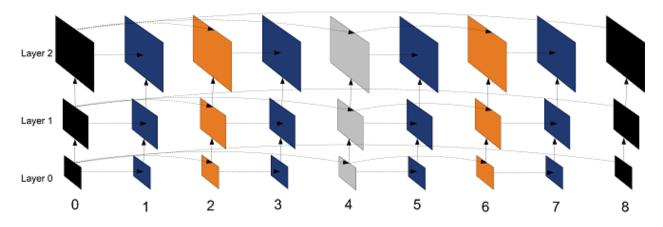
Figure 2.5 Different spatial resolution in spatial scalability

To produce such a layered bit stream, a multi-resolution decomposition of the original image is first obtained. The lowest resolution image is coded directly to produce the first layer (i.e., the base layer). To produce the second layer, the decoded image from the first layer is first interpolated to the second lowest resolution and the difference between the original image at that resolution and the interpolated one is coded. The bit stream for each of the following resolutions is produced in the same way first form an estimated image at that resolution, based on the previous layers, then code the difference between the estimated and the original image at hat resolution transformed, quantized and VLC-coded. For the enhanced layer, the operations are performed as follows:

1. The raw video is spatially down-sampled, DCT-transformed and quantized at the base layer
2. The base-layer image is reconstructed by inverse quantization and inverse DCT;
3. The base-layer image is spatially up-sampled;
4. Subtract the up-sampled base-layer image from the original image;

2.1.3 Temporal scalability

For such a structure the number of layers of reduced spatial resolution and number of layers of reduced temporal resolution is determined at the encoder side. In the case of hybrid codecs there may be distinguished two methods for obtaining temporal scalability. The first one which is used for most of hybrid codecs divides encoded frames into the following types

- frame which can be used as reference for other frames,
- frame which cannot be used as reference for other frames.
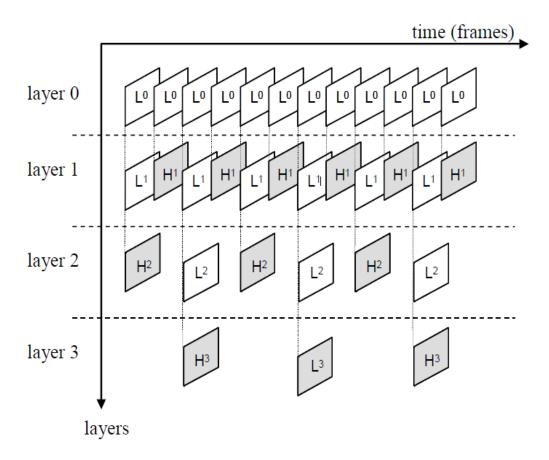-

Figure 2.6 Temporal scaling in Temporal scalability

Figure 2.6 shows temporal scaling in temporal scalability. If a frame cannot be a reference frame it can be dropped in the communication channel if there is not enough bandwidth to deliver the bit stream to the decoder or at the decoder side if such a decoder has no computational power to decode this frame [3] [4]. Both types of frames may be encoded: as access point frames, it is a frame which is encoded only byte use of spatial prediction modes (this frame does no use other frames for prediction); as a frame using one directional temporal prediction as well as two-directional temporal prediction (those frames use other frames for prediction).

Temporal scalability enables different frame rates for different layers of the contents. Typically, tempo-rally scalable video is encoded in such an efficient way: making use of temporally Up-sampled pictures from a lower layer as a prediction in a higher layer. The block diagram of temporally scalable codec is the same as that of spatially scalable codec. The only difference is that the spatially scalable code uses spatial down-sampling and spatial up-sampling while the temporally scalable encode uses temporal down-sampling and temporal up-sampling. The simplest way to perform temporal down-sampling is by frame skipping [7] [8].

For example, temporal down-sampling with ratio 2:1 can be achieved by discarding one frame from every two frames. Temporal up-sampling can be accomplished by frame copying. For example, temporal up-sampling with ratio 1:2 can be realized by making a copy for each frame and transmit the two frames to the next stage. In this case, the base layer simply includes all the even frames, and the enhancement layer all the odd frames. For motion compensation, a base layer frame will be predicted only from previous base layer frames, whereas an enhancement layer frame can be predicted from both base layer frames and enhancement frames.

2.2 Motion estimation

2.2.1 Backward

It is most normally utilized motion estimation method in video sequence. In this method current frame act as the candidate frame, and in the previous frame arrange the search region then find the best match gives motion vector as shown in Figure 2.7. This strategy is also called forward motion estimation.
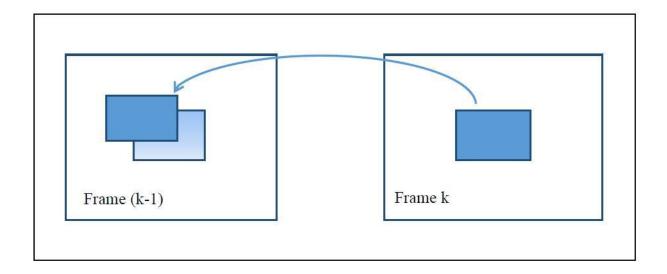
Figure 2.7 Backward Motion Estimation

2.2.2 Forward

It is precisely the converse of in Backward motion estimation (see Figure 2.8). Here the present frame is our candidate frame and next frame goes about as our reference frame, which implies current frame is predicted from a next frame. On the other hand we can say search is forward. Also, this procedure is called as Backward motion estimation.
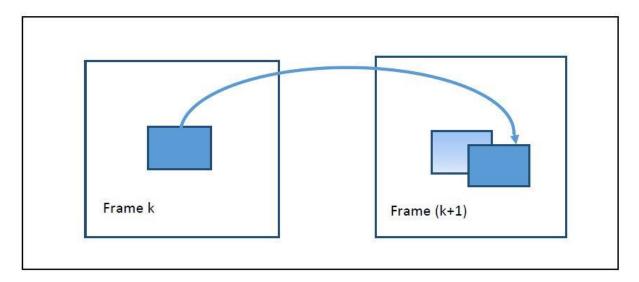
Figure 2.8 Forword Motion Estimation

2.3 Block based motion estimation

When doing video compression ideally we aim at achieving a refined output at the decoder which tends to be almost similar to the raw one which has been given as input to the encoder. To achieve this motion estimation plays the great role.

To accomplish motion estimation motion vectors are of major importance. If we analyses from ideal perspective, in a frame of a video sequence every pixels must be assigned to a motion vector which can indicates towards the pixel which is most alike to it. This requires a huge amount of evaluation for finding the motion vectors.

So we can state it precisely that the process explained above cannot be implemented practically due to involvement of huge amount of computational time [14][15]. Hence alternatively for real time applications we can group the pixels into a blocks either with fixed or variable sizes. The motion estimation can be carried out by the help of algorithms availed.

The following figure 2.9 shows the basic frame work for block matching.
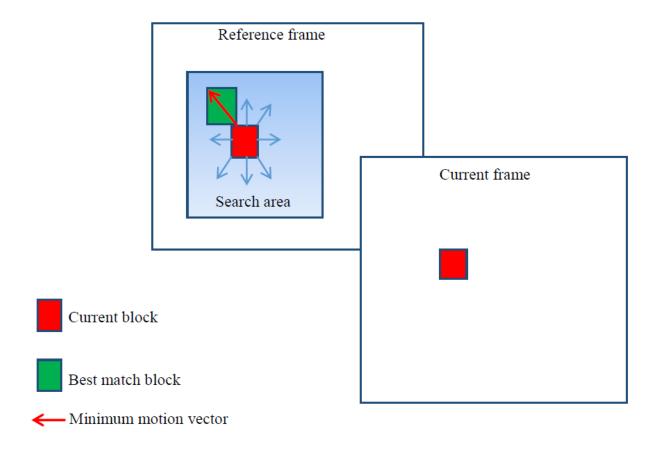
Figure 2.9 Basic diagram of Block matching estimation

Block matching algorithms are the widely used algorithms in this context. In block matching we can find the motion vector by carrying out a search operations across all the positions possible in accordance to the criteria being used for matching. Due to the simplicity and ease of application block matching are the best ones used for video coding systems. To bring down the computational complexity and making the motion estimation more efficient various methodology for search have been developed in order to achieve a proper match for a block. The methods can be enlisted as TSS, 4step search, Adaptive root search, Ds, zonal-based search etc.

2.3.1 Full search (FS)

The FS method gives a global optimal motion vector from all the candidate blocks with in the search window. If the search window size is 48×48 pixels and the block size is 16×16 pixels, total will be $16 \times 16 = 1024$ search points that needs to undergo mean absolute error (MAE) computation. It's an exhaustively search method. Therefore it is simple to implement but it takes high computational time.
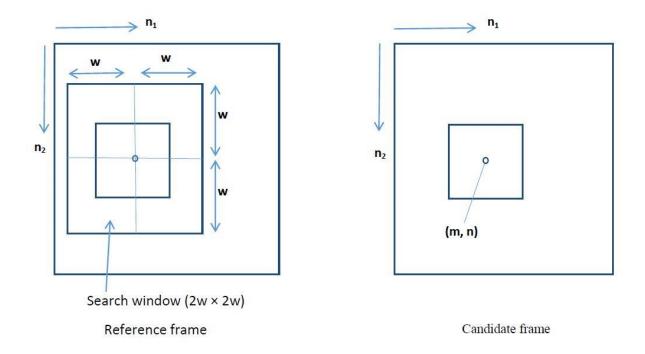
Figure 2.10 Full Search method
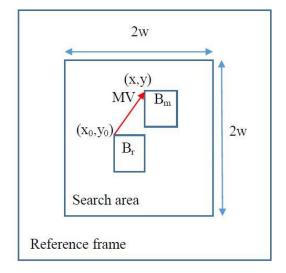
Figure 2.10 shows, a block of $N \times N$ pixels in the reference frame at the direction area (m,n), and consider a scope of $\pm w$ pixel of the searching window in both headings on the reference frame. The applicant block is contrasted with a block of size $N \times N$ pixels for each of the $(2w + 1)2$ search.

CHAPTER 3

A Mean and Median based Adaptive search algorithm for motion estimation

3.1 Introduction

In any sort of Block matching algorithm, the first step is partitioning the video frame into non-overlapping blocks. After that, for every block in the candidate frame, a search window is arranged in the reference frame. At that point the procedure if block matching is done, that is by matching each block in candidate frame with reference frame, to acquire a motion vector for every square in the applicant frame.
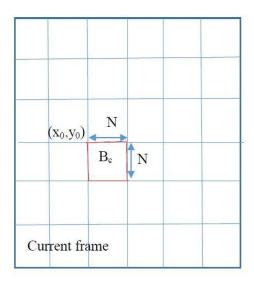
Figure 3.1 Block Matching Estimation

As shown in Figure 3.1, the main theme of block matching method. Here, in the current frame Bc is one of the block among all the blocks. Using with Bc start the searching for best match in reference frame, Br is the block which is having same co-ordinates in the reference frame corresponding to Bc. So the Br and Bc co-ordinates are same (x0, y0). Next we arrange a search region around Br and within this region the search has to be start until the best match found for Bc. Suppose after finding the best match as Bc is Bm. Then the motion vector generate as (x, y). As it is the process will be going on until the all blocks in the current frame found the best match.

$$E(\mathrm{MV}) = \left(\frac{1}{N^2}\right)^{\beta} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \left| I_c(\mathrm{x}_0 + \mathrm{m}, \mathrm{y}_0 + \mathrm{m}) - I_r(\mathrm{x} + \mathrm{m}, \mathrm{y} + \mathrm{m}) \right|^{\gamma} \qquad (1)$$

where $MV = (\mathrm{x} - \mathrm{x}_0, \mathrm{y} - \mathrm{y}_0)$

Here, *E (MV)* is the distortion or motion vectors of Bc to shows best block in reference frame. Motion vector is nothing but angle of displacement of the matched block (Bm) with respect to Br. In the equation (1). *Ir* and *Ic* are the intensity values of pixels in the both frames, and *N* is the measurement of the blocks (block size is *N×N*). Here, β and γ are decide the matching criteria. For β = 0, γ = 1, it is Sum of Absolute Error (SAE), for β = 1, γ = 1, its criteria is sum of Mean Absolute Error (MAE) and for β = 1, γ = 2 it is Mean Square Error (MSE).

There are many diverse motion estimation methods to get MV, we have officially examined numerous block matching procedures, yet these methods have a settled search design and are computationally extravagant. We can spare a ton of calculation exertion without losing the much quality of video by using new proposed search algorithm in block matching method which can eliminate the unnecessary blocks where motion could not occur. By utilizing this algorithm we could save a half of the computation time compared to congenital search in block matching algorithm.

3.2 Procedure of mean and median based adaptive search algorithm

Motion Estimation is a process where the objects are examined whether they are in motion or not and according to the degree of displacement, the motion vector gets generated. As it is a time consuming process, it is desirable to reduce the time of computation to increase the efficiency but keeping in view that not at the cost of much quality. There are many Search methods used for Motion estimation. The ideal one is the Full search method. In this method the first frame is taken as a Reference and second frame is taken as a current Frame. For the next incoming frame the current frame will act as a reference frame.

To perform the operation over the frames taken, Zero padding is required. The search region should be defined in such a way that it will cover all the blocks of the frame including the neighboring blocks. Search is being done for the current block by search window within the search region until some difference i.e. object in motion found. Then the absolute values are taken and Summing operation is performed.

The value found will be compared with the Threshold value. If it's less than the Threshold value then we will take the motion vector according to the coordinates of the corresponding frame. But the Full search is computationally intensive so we aim at developing a method which will do fewer evaluations than Full search method but give an output without losing much of the quality which will be acceptable from the prospective of biological vision. Figure 3.2 shows flow chart of new search method.

3.3 Flow chart

The flow chart of new search method shown here:

```
                          ┌───────────┐
                          │   START   │
                          └─────┬─────┘
                                │
                    ┌───────────▼───────────┐
                    │    Reading frames      │
                    └───────────┬───────────┘
                                │
                    ┌───────────▼───────────┐
                    │    Add Zero padding    │
                    └───────────┬───────────┘
                                │
                    ┌───────────▼───────────┐
                    │   Define Search area   │
                    └───────────┬───────────┘
                                │
        ┌───────────────────────▼───────────────────────┐
        │                 Search Window                   │
        └───────────────────────┬───────────────────────┘
                                │
            ┌───────────────────▼───────────────────┐
            │          Current block (C_B)           │
            └───────────────────┬───────────────────┘
                                │
              ┌─────────────────▼──────────────────┐
              │   K= max ((A_B) or (M_B) or (V_B))  │
              └─────────────────┬──────────────────┘
              ┌─────────────────▼──────────────────┐
              │   B= max ((A_F) or (M_F) or (V_F))  │
              └─────────────────┬──────────────────┘
                                │
                         ◇  K > B  ◇ ──No──→ ┌─────────────┐
                                │            │ Shift block │
                               Yes           └─────────────┘
                                │
                    ┌───────────▼───────────┐      ┌─────────────┐
   ┌─────────────┐  │  Searching for best   │      │ Shift block │
   │ Shift block │  │   match of C_B in R_B │      └─────────────┘
   └─────────────┘  └───────────┬───────────┘
                                │
                         ◇ Motion occur ◇ ──No──→
                                │
                               Yes
                                │
                    ┌───────────▼───────────┐
                    │  d = sum(abs(C_F – R_F))│
                    └───────────┬───────────┘
                                │
                    ┌───────────▼───────────┐
                    │     Store vectors      │
                    │     x1=V, y1= U.       │
                    └───────────┬───────────┘
                                │
                          ┌─────▼─────┐
                          │    End    │
                          └───────────┘
```
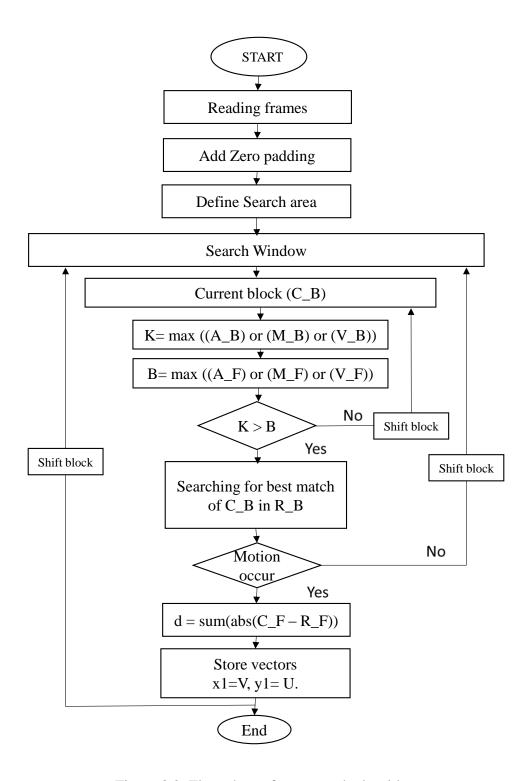
Figure 3.2. Flow chart of new search algorithm

3.4 Steps for proposed algorithm

The proposed method has been given as follows:

1. Take the Current Frame and Reference frame.
2. Divide the Frame into non-overlapping Rectangular blocks of equal size.
3. Do Zero padding for both the Frames.
4. Define Search region for Reference Block.
5. Find the Mean (A_B), Median (M_B), and Variance (V_B) of the Block, which is going Through Search.
6. Find the Maximum value between Mean (A_B), Median (M_B), and Variance (V_B) of the Block.
7. Find the Mean (A_F), Median (M_F), and Variance (V_F) of the whole frame.
8. Find the Maximum value between Mean (A_B), Median (M_B), and Variance (V_B) of the Frame.
9. Pick the Maximum number of value among three values from Both the Steps.
10. Compare, If (M_F) or (A_F) or (V_F) >= (M_F) or (A_F) or (V_F), then the block will be Skipped.
11. If (M_F) or (A_F) or (V_F) =< (M_F) or (A_F) or (V_F), then it will Search for the Current Block.
12. Then it will find the difference between the Current block and Reference block, then the Absolute value is taken.
13. The Value (d) obtained in the previous step will be compared with the Threshold value.
14. If, d > D (threshold), it will go to another block instead of evaluating the Current block.
15. If d < = D, the Motion Vector will be generated.

CHAPTER 4

Motion Estimation using Particle Swarm Optimization

4.1 Introduction

Motion Estimation is an important parameter while encoding a video. For motion estimation a block based matching methodology is being followed. Initially two frames are taken and they are segmented into several blocks. The comparison is being carried out between the frames taken for each block using some pre-decided matching search. Here the compression of the concerned frames are done by finding the best match by the utilization of matching searches. The procedure followed is known as motion estimation algorithm. Theses algorithms always strive to decline the time of computation by evaluating only some major points inside the search window used.

The goal is always to decrease the number of search points without losing the quality of the output much. As motion estimation takes a large amount of time to be completed, hence need of such algorithm to bring down the time needed is inevitable [9]. The perfection of the algorithm used is better than the conventional. It's a proven fact that the consecutive frames in a video sequence have always a great amount of similarity. So it is possible to remove the similarity in order to enhance the coding gain. During video encoding one should certainly follow the motion compensation and motion estimation. The block matching technique is the most popular method for estimating motion. The algorithm followed in this work is accomplished by using the search pattern which is cross shaped in the two steps followed initially. It makes the motion estimation between the blocks faster. The block matching can also be viewed as a problem of optimization as it aims at obtaining the most suitable matching block inside a search space specified. Particle swarm optimization (PSO) can be implemented in this context. It has multifarious benefits such as revamping the block matching efficiency of the algorithm by determining the initial individuals on the basis of fixed and random points.

4.2 Working principle of PSO

Particle swarm optimization is a well-known optimization algorithm which is a blending of two streams i.e. social science and computer science. Here the concept of 'swarm intelligence' is also being used. Swarm intelligence is an attribute of a system which is an outcome of the interaction of unsophisticated agents along with their environment and resulting in coherent global functional patterns. The description below explains major traits of PSO:

$$V_i(t+1) = wV_i(t) + c_1 r_1 \left[P_i(t) - X_i(t) \right] + c_2 r_2 \left\lfloor P_g(t) - X_i(t) \right\rfloor \qquad (2)$$

$$X_i(t+1) = X_i(t) + V_i(t+1) \qquad (3)$$

Where i is the index of the particle, i = 1, 2. . . M; w the inertia weight; c1, c2 the positive acceleration constants; r1, r2 the random numbers, uniformly distributed within the interval [0, 1]; t the number of iterations so far; g the index of the best positioned particle among the entire swarm. With the reference to social science it can be concluded that the interactions occurring socially builds up the intelligence level of a human being. Explaining further, comparing the situations, evaluating something, gaining ideas and information from day to day experiences made human being to be adaptable to the surrounding and construct a base for characteristics, behavior etc. culture and cognition are unavoidable results coming out from human sociality. Culture is an outcome of a group of individuals adopt a same process of learning. Next comes swarm intelligence, which can be explained by following traits:

1. The population need to accomplish simple space and computational time.

2. The quality factors must be responded by the population.

3. The activity by the population should not be carried by across the channels which are excessively narrow.

4. The environmental variations must not affect the population behavior every time.

5. The behavioral mode must be adaptable by nature so that it can be adjusted when the computational price is of prior importance.

The particles are manipulated obeying the above equations

4.3 Motion estimation using PSO:

As discussed earlier the block based motion estimation is a popular method for motion estimation. Initially the frames from the video sequence are divided as non-overlapping blocks. The prediction process is carried out from the block that is of same size of the contemporary reference frame. The aim is to find the best alternative for the block in the current frame inside the search window of the reference frame for minimizing the matching metric. This block matching, if analyzed vividly can be viewed as an optimization problem as it tries to achieve the best match within the search window. So to address this problem particle swarm optimization can be used as it is an effective algorithm for optimization problem. It has been implemented in this work using following algorithm shown Figure 4.1.

The population is first chosen which we need to deal with. Then the particles located at random positions are being initialized. In the swarm for each particle position and velocity is determined. From all the positions a certain location of a particle is chosen to be the best one and can be referred as p_{best}. Then the position of each individual particle gets compared with the best value found. If it has been observed that if in any case, the position of an particle is better than the best position decided previously then the current position is settled as the best one i.e. $p_{best.}$. At last when all the comparisons will be finished then all the p_{bests} found are compared and amongst them the best value for the position is being picked up which is referred as global best i.e. $g_{best.}$ After finding the global best value the velocity and position of the particles are updated.
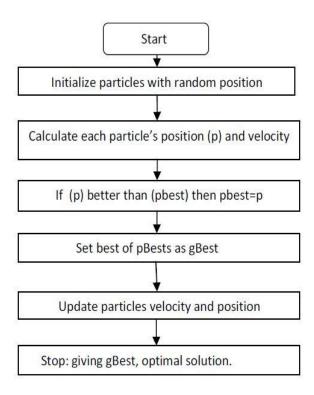
4.3.1 Flow chart

Figure 4.1 Flow chart of PSO

Finally g_{best} is being fixed as the optimal solution. Afterward, here comes the stopping criteria, which is based on the following: before starting the algorithm certain number of iteration have been fixed. After reaching that value there will be no more iteration and the process will be stopped. Using this we will be able to make some decrement in the number of search points, hence reducing computation time but not losing much image quality.

CHAPTER 5

Results & Analysis

5.1 Conventional search in SNR scalability

In this section one video 'News.avi' has been taken and encoded in SNR scalable domain. Initially to accomplish the task intra-coding has been taken place for the first frame and it has been stored inside buffer. The adjacent pictures have gone through predictive coding. During the process the step size for quantization will be 16 and scale for quantization will be 4 for the coefficients of DCT. This is the set up that constructs base layer. Then quantization has been carried out with a step size 4 for the disparity among the DCT and the DCT after quantization and de-quantization in the generated base layer for the all coefficients present [1]. The information obtained here additionally can be referred as enhancement layer. Variable length coding has been implemented on both base and enhancement layer.

Then the encoded outputs are either subjected to transmission or storing. If the client needs a video of lower quality then decoding is needed to be applied only on base layer and if the requirement is a video of higher quality then both of the layers will be decoded. Precisely, it can be said that viewed that the base layer has undergone through noticeable amount of quantization distortion while the amount is less in the enhanced layer. To prove the above said the plots of signal to noise ratio, peak signal to noise ratio as shown in Figure 5.1 (a),(b) respectively. And SSIM has been shown below Figure 5.2.

The values have been computed by taking certain number of frames such as 40 to 59. It should be marked that the scale chosen for quantization aims at generating the huge differentiation between the base layer along with the enhancement layer.
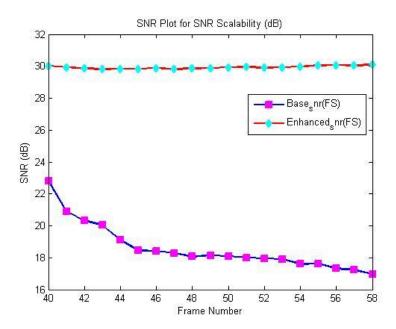
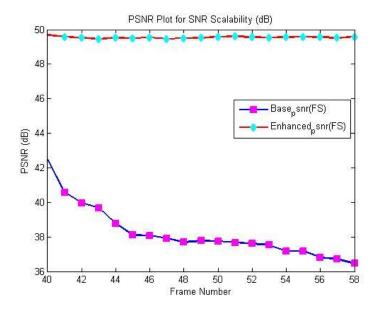Figure 5.1 (a) SNR plot for SNR scalability of "News.avi" sequence.

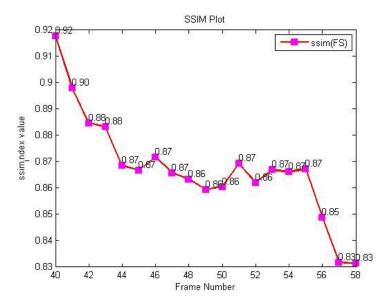Figure5.1 (b) PSNR plot for SNR scalability of "News.avi" sequence.

Figure 5.2 SSIM plot for of "News.avi" sequence.

5.2 Motion estimation using PSO.

Comparatively PSO is the best optimization technique for motion estimation than full search in terms of computation time. But we cannot achieve quality of reconstructed frames as much as full search. Figer 5.3 (a), (b), (c)SNR, PSNR and SSIM plots of "News" sequence respectively.
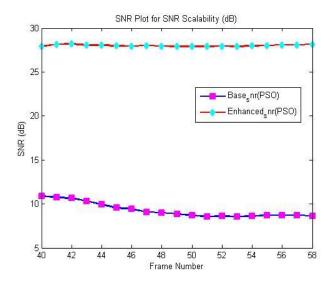

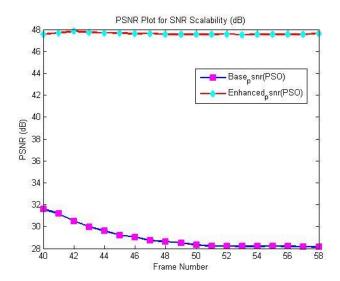Figure 5.3 (a) SNR plot for "News.avi" using PSO algorithm

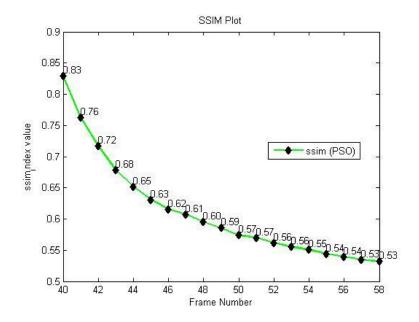Figure 5.3 (b) PSNR plot for "News.avi" using PSO algorithm

Figure5.3 (c) SSIM plot for of "News.avi" using PSO algorithm

5.3 Mean and median based adaptive search algorithm

As the Full search is computationally intensive and the PSO is too best in terms of computational complexity with degradation of quality of video so we aim at developing a method which will do fewer evaluations than Full search method but give an output without losing much of the quality which will be acceptable from the prospective of biological vision.

From the results shown it has been verified that the conventional search method takes huge amount of computation time. But with the proposed method we are able to reduce the computation time up to a considerable extent. The SNR & PSNR graphs shown are meant for measuring the quality. The time for computation has been reduced to 50-60% rather than the traditional technique.

Figure 5.4 (a), (b), (c) shows the SNR PSNR and SSIM plots respectively for the Rhinos.avi sequence.
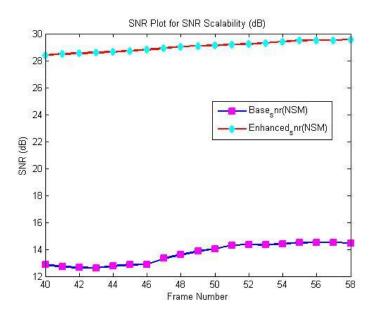

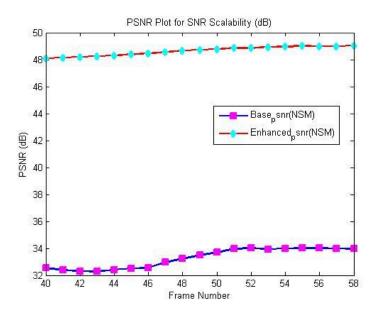Figure 5.4 (a) SNR plot for of "News.avi" using new search algorithm

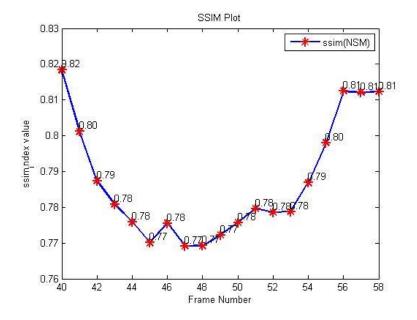Figure 5.4 (b) PSNR plot for of "News.avi" using new search algorithm

Figure 5.4 (c) SSIM plot for of "News.avi" using new search algorithm

5.4 Comparison with proposed and conventional results

The three sections depicted above are needed to be analyzed. Here, used different video sequences for the purpose of analysis a comparative study. Figure 5.5 (a), (b), (c) SNR, PSNR and SSIM plots for "News" sequence with different search techniques.
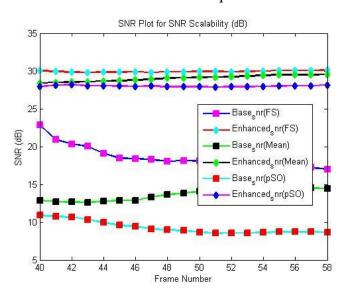

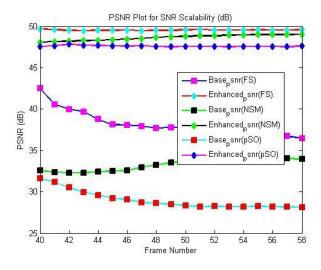Figure 5.5 (a), Comparison of SNR for "News.avi" with different search methods

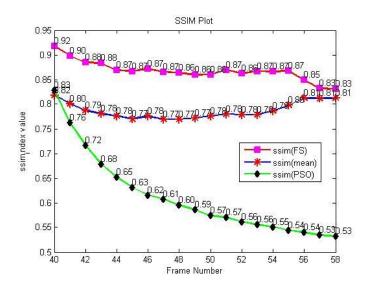Figure 5.5 (b), Comparison of PSNR for "News.avi" with different search methods

Figure 5.5 (c), Comparison of SSIM for "News.avi" with different search methods

The same procedure was carried out for the sequence of 'coastguard_cif.avi' and the graphs of SNR, PSNR shown respectively in Figure 5.6 (a), (b).
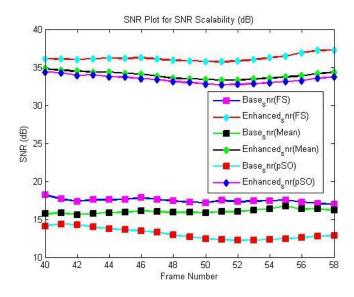

Figure 5.6 (a), Comparison of SNR for "coastguard_cif" sequence with different search methods
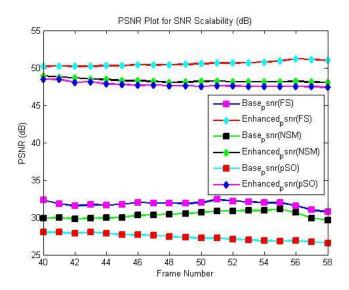
Figure 5.6 (b), Comparison of PSNR for "coastguard_cif" sequence with different search methods

The results obtained from the above sections are tabulated below with the intention of comparison. The average is being done in Base and Enhancement layer for SNR and PSNR values and comparison has been given between conventional, new search method and PSO. This method is carried out for all the four sequences and the experimental values are depicted in Table I and II which is shown below. Table III the computation time and SSIM has been shown for both conventional, new search method and PSO.

Table I. SNR values in dBs for both Base layer (BL) and Enhancement layer (EL)

Video Sequence	Conventional Search		New Search Algorithm		PSO	
	BL	*EL*	*BL*	*EL*	*BL*	*EL*
Foreman	19.47	33.61	18.12	33.23	14.35	31.25
coastguard	17.48	36.22	16.03	33.94	13.08	33.40
News	18.59	29.93	13.65	29.02	09.27	28.00
Football	12.26	32.05	10.84	31.62	09.05	30.86

Table II. PSNR values in dBs for both Base layer (BL) and Enhancement layer (EL)

Video Sequence	Conventional Search		New Search Algorithm		PSO	
	BL	*EL*	*BL*	*EL*	*BL*	*EL*
Foreman	33.08	47.23	28.98	47.28	26.37	46.95
coastguard	31.79	50.55	30.32	48.28	27.35	47.73
News	38.20	49.54	33.27	48.64	28.98	47.60
Football	28.18	48.10	26.95	47.23	24.85	46.83

Table III. Computation time in sec. for Full search, new search method and PSO

Video Sequence	Computational Time (sec)			SSIM		
	Conventional search	New Search algorithm	PSO	Conventional search	New Search algorithm	PSO
Foreman	184.50848	93.92231	25.45256	0.8530	0.6949	0.5450
Rhinos	172.12640	80.50817	22.56324	0.8924	0.7423	0.6256
News	112.25648	48.56324	10.25468	0.9193	0.8364	0.6325
Football	176.35648	95.32647	24.56324	0.7670	0.6853	0.5268

From the results shown it has been verified that the conventional search method takes huge amount of computation time. But with the PSO and proposed method (new search algorithm), we are able to reduce the computation time up to a considerable extent. The SNR & PSNR graphs shown are meant for measuring the quality.

CHAPTER 6

6. Conclusion and future work

In the modern era the developing technology has compelled the streaming makers to build a bit stream which can be an answer to the need of various group of client and gadget. The solution found is referred as scalability in video coding. The scalability again works in various mode. In this work the video sequence has made to be scalable in SNR mode. Again motion estimation is a important part of video encoding. To make the encoding more effective and to reduce the computational complexity the process of motion estimation should be molded accordingly. In the thesis a new search algorithm has been developed which will reduce the evaluation time 50-60% with negligible decline in resolution. Again attempt has been made to address this with PSO which reduces time up to 80% but with the cost of much reduction in resolution.

In future this work can be developed with the following aspect:

- ➢ Combined scalability can be collaborated with the proposed scheme.
- ➢ Motion estimation should be developed with more precise concept to decrease the computation time further.
- ➢ The suggested scheme can also be realized by hardware
- ➢ Developed prediction structure can be blended with the work done to enhance the coding gain.

References

[1] Thyagarajan, Kadayam S. *Still Image and video compression with MATLAB*. John Wiley & Sons, 2011.

[2] Li, Shipeng, and Weiping Li. "Image and Video Coding." *Encyclopedia of Telecommunications*. Vol 2, pages 1025-1058, 2003.

[3] Mrak, Marta, and Ebroul Izquierdo. "Scalable Video Coding Fundamentals. *"Encyclopedia of Multimedia*. Springer US, 2008. 771-775.

[4] Jillani, Rashad, and Hari Kalva. "Scalable Video Coding Standard" *Encyclopedia of Multimedia*. Springer US, 2008. 775-781.

[5] Schwarz, Heiko, Detlev Marpe, and Thomas Wiegand. "Overview of the scalable video coding extension of the H. 264/AVC standard." *Circuits and Systems for Video Technology, IEEE Transactions on* 17.9 (2007): 1103-1120.

[6] Błaszak, Łukasz. "Advanced Scalable Hybrid Video Coding." (2006).

[7] Lee, Ying. *"Scalable video"*. Diss. Massachusetts Institute of Technology, (2000).

[8] Yuan, Yufei. *"Wavelet video coding with application in network streaming."* Diss. University of Alberta, 2005.

[9] Jacob, Asha Elizabeth, and Immanuel Alex Pandian. *"An efficient motion estimation algorithm based on particle swarm optimization."(IJESS), ISSN: 2231-5969, vol-3, 2013"*

[10] *Information technology-JPEG2000 image coding system*, ISO/IEC15444, 2000.

[11] Richardson, Iain E. *H. 264 and MPEG-4 video compression: video coding for next-generation multimedia*. John Wiley & Sons, 2004.

[12] Yao Wang, Joern Ostermann, and Ya-Qin Zhang, "Video Processing And Communications,*" Printice hall signal processing series, 2002*.

[13] Nasir D. Memon, Khalid Sayood, *"Lossless Compression of Video Sequences," IEEE transaction on Communications*, vol.44, no. 10, 1996.

[14] Avijit Kundu, "Modified Block Matching Algorithm for fast Block Motion Estimation," *International conference on Signal and Image Processing*, 2010.

[15] Xiaolin Chen, Nishan Canagarajah, and Jose L. Nunez-Yanez "Backward Adaptive Pixel-based Fast Predictive Motion Estimation," *IEEE signal processing letter*, vol.16, no. 5, 2009.