ADAPTIVE RATE CONTROL ALGORITHM TO IMPROVE THE PERFORMANCE OF H.264 VIDEO CODEC IN VARYING NETWORK CONDITIONS

A Thesis by

Swapnil Shingvi

M.S., Wichita State University, December 2006.

Submitted to the Department of Electrical and Computer Engineering
And the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Master of Science

December 2006

ADAPTIVE RATE CONTROL ALGORITHM TO IMPROVE THE PERFORMANCE OF H.264 VIDEO CODEC IN VARYING NETWORK CONDITIONS

I have examined the final copy of this thesis for form and content, and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science with a major in Electrical and Computer Engineering.

Dr. Ravi Pendse, Committee Chair and Advisor

We have read this thesis and recommend its acceptance:

Dr. Kamesh Namuduri, Committee Member

Dr. Krishna Krishnan, Committee Member

DEDICATION

This thesis is dedicated to my beloved parents, all my family members and friends who have supported and encouraged me throughout my life. Without their support and encouragement this thesis would not be a success.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Ravi Pendse, for all his support and valuable guidance over the entire course of my academic career at WSU and for giving me an opportunity to do my Master's thesis under his guidance in the field of Video Coding. I am thankful to my committee for taking the time to work with me in this endeavor.

I would also like to thank Nagaraja Thanthry, Amarnath Jasti and Vijay Ragothaman and all my friends at the Advanced Networking Research Center (ANRC) at Wichita State University for their help they have given me during this time.

Finally, I am forever indebted to my parents and family for supporting and encouraging me throughout my life to achieve higher goals.

ABSTRACT

This thesis involves proposing an efficient algorithm, which takes into consideration the NALU (Network abstraction Layer Unit) and tries to improve the H.264 codec performance over a network. . Along with the bandwidth instabilities in the Internet and other issues like the packet drop cause the video frames to be dropped, which makes the video, appear distorted. The main idea behind the algorithm is to vary the bitrate by varying the Quantization parameter, both of which are stored in the reference table. The reference table is calculated based on the analysis of the sequences of different video clips .By changing the bitrate based on the bandwidth available the codec performs better and produces good results of (Signal to Noise Ratio)SNR. The research is based on the development of the reference table which when referred helps to change the quantization parameter of the codec for coding a Group of Frames in a video by periodically checking the network statistics. At the cost of reduction in the resolution due to the network conditions the video frame drop is avoided which helps improve the video quality at the decoder..

The entire simulations for this research were carried out using the JM Reference Software H.264 Encoder and decoder Version JM 10.2. Packet drop was simulated by dropping the individual NAL units from the video stream.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AVC .... Audio and Video Coding

CBR.. Constant Bit Rate

CODEC .... Coder/Decoder

DVD... Digital Versatile/Video Disc

GOF... Group Of Frames

GOP... Group Of Pictures

HDTV .... High Definition TV

JVT .... Joint Video Team

MPEG .... Moving Picture Experts Group

MTU.. Maximum Transmission Unit

NAL .... Network Abstraction Layer

NALU .... Network Abstraction Layer Unit

QP .... Quantization Parameter

RTP .... Real-Time Transport Protocol

SNR.. Signal to Noise Ratio

VBR .... Variable Bit Rate

CHAPTER 1

INTRODUCTION

The internet is not naturally suitable for real time traffic. This is due to the fact that compared to the other applications, which are based on pure text, multimedia requires very high bandwidth due to the high volume of data that needs to be sent to represent it. The multimedia data requires real time traffic in which the audio and video need to be played at a continuous rate at which it was sampled at the encoder side. If there is a delay in the receipt of the multimedia data at the receiver it may cause it to be out of sequence and hence the decoder may discard it or drop it. This may lead to Transmission Control Protocol (TCP) retransmission, which would further consume the available bandwidth and cause congestion, which is unacceptable in case of video and audio transmission. Moreover, multimedia traffic is bursty in nature and the increase in the bandwidth is not the solution for the efficient transmission of audio or video over the Internet. When resources are scarce, real-time traffic will suffer from the congestion. Improvement can be achieved by taking into consideration the available bandwidth, packet loss, and the transmission of multimedia over the Internet by introducing some sort of adaptability in the available codecs. One of the codecs, which has a significant reduction in the bitrate, compared to the other standards for the similar degree of encoder optimization, is the H.264 for the video.

This thesis involves improving the performance of H.264 codec by making it adaptable to the varying network conditions, like the bandwidth availability and the packet loss due to the burst in the traffic, by varying its bitrate. Analysis shows that maintaining a reference table for the Quantization Parameter (QP) and the target bit rate helps in the easy calculation of QP for the available bandwidth at any instance, compared to the methods, based on the previous history of frames and the bits per pixel calculation, which are complex in nature. The previous implementations have not taken into consideration about the packet loss, which when considered will reduce their performance by degrading the signal to Noise Ratio (SNR) value. With the cost of reduction of the resolution of the video we get better quality of video because there are no frame losses.

1.1 EVOLUTION OF MPEG STANDARDS FOR AUDIO/VIDEO CODING

MPEG stands for Motion Pictures Experts Groups. It is a standard, which refers to the compression and the storage of the video and the format of the files, containing this compressed video. The basic concept is to transform a stream of discrete samples into a stream of bits, which takes less space at the encoder. The decoder then knows how to inverse represent the compact stream of bits, which represents the original stream of samples of the actual video, which was encoded at the encoder side. There are sets of rules, which represent the semantics or the so-called algorithms with the mathematical calculations.

The various standards evolved are the MPEG 1, MPEG 2 and MPEG 4, consist of Video Home System (VHS) to the Digital Versatile/Video Disc (DVD) streaming over the web.

1.1.1 MPEG-1 [20]

This standard defines a low-resolution video sequence of about 352 by 240 pixels and the frame rate of 30 frames per second. The codec for the MPEG 1 needs a bandwidth of 1.5 Mbps. The application of the MPEG1 is the Video CD format. One of its disadvantages is that it only supports progressive video and has no support for the interlaced video.

1.1.2 MPEG-2 [20]

This standard defines a support for higher resolution video and increased video capabilities. The bit rate for this standard is 4 to 15 Mbps, providing a good quality of full screen video. The program stream for MPEG-2 is used in DVD standards. The video of MPEG-2 is very similar to the MPEG-1 and it provides the support for the interlaced video. It outperforms the MPEG-1 at 3 Mbps and above.

The decoders for the MPEG-2 are backward compatible and are fully capable of playing back the MPEG-1 video streams. MPEG-2 Video and Systems are also used in most High Definition TV (HDTV) transmission systems.

1.1.3 MPEG-4 [20]

Due to the need for Internet streaming video applications, which have a bandwidth range less than 64kbps – 4Mbps there was a need to have higher scalable codec's, which gave rise to the MPEG4 codec. MPEG4 adds the support for the Virtual Reality Modeling Language and 3D rendering. It was mainly designed to support the web streaming video, CD distribution, and `conversation for broadcast television, all of which benefit from compressing the Audio/Video stream. There are no complete implementations of the MPEG4 standard and most of its features are left for individual developers to decide and implement in their way to customize for their requirements.

1.1.4 H.264/MPEG-4 AVC [20]

H.264 is an emerging coding standard and provides a lot of advantages over the other coding standards. It provides reduction in the bit rate up to 50% compared to other coding standards for the similar degree of encoder optimization.

The emerging coding standard H.264 is mainly intended for video transmission in all areas where bandwidth or storage capacity is limited. The H.264 standard has a number of advantages that distinguish it from the other existing standards mentioned above, while at the same time-sharing common features with other existing standards. It offers good video quality at high and low bitrates. The heart of the H.264 codec is the Network Abstraction Layer (NAL), which helps the bit streams to be transported over different network segments.

1.1.4.1 Encoder [1]:

There are basically two dataflow paths in the encoder namely the "forward" path and the "reconstruction" path.

Encoder (Forward path):

In the forward path the input frame Fn is presented for encoding. A prediction macroblock is formed based on reconstructed frame in the modes called the Intra and the Inter for the prediction P. Each input frame Fn, which is sent to the encoder, is processed in the units, which are called macroblocks, which are the 16 x 16 pixels of the original image.

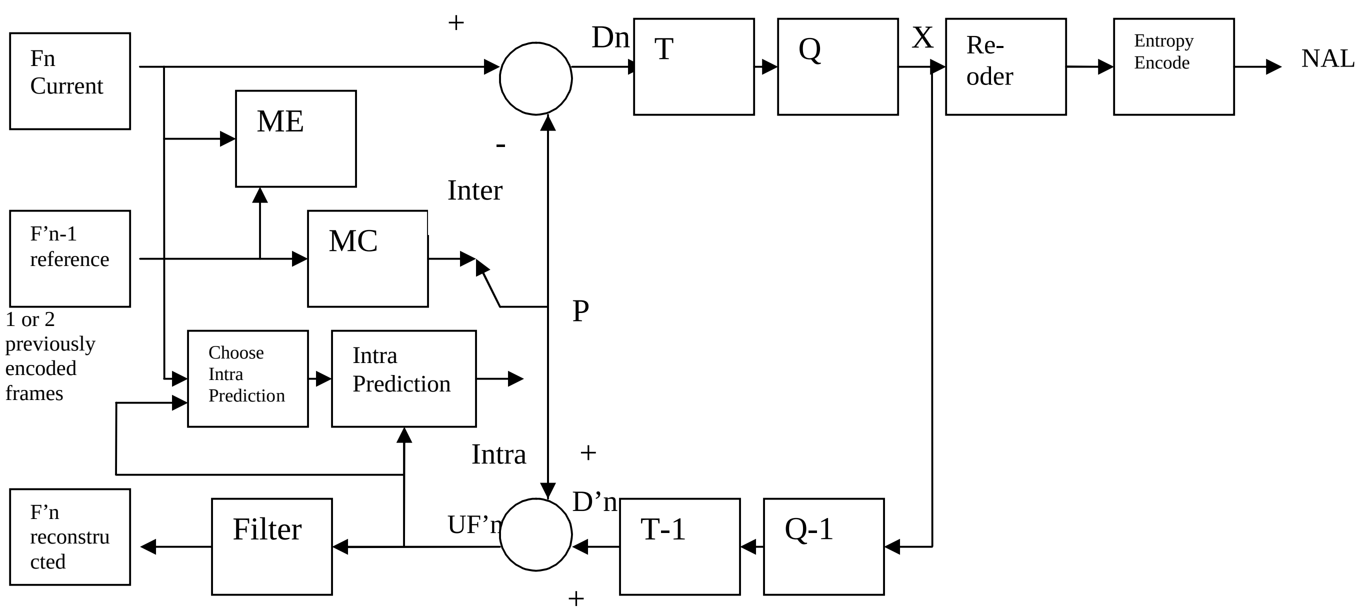


Figure 1.1: H.264/AVC Encoder

- Intra mode: P is formed from samples in the current frame and from the frames that have been encoded, decoded, or reconstructed.
- Inter mode: In this mode the prediction P is formed from the motion compensated prediction from one or more reference frames (usually the previous encoded frames). Each macroblock is predicted and formed from the previous couple of frames, which are encoded, decoded, and reconstructed.

The next step involves the creation of the difference macroblock Dn that is obtained by subtracting the current macroblock. Further, transformation and quantization of the Dn causes the X set of quantized transforms to be generated. Next takes the formation of the bitstream, which includes the macroblock along with the side information, which is the macroblock prediction mode, quantizer step size, etc. The transmission occurs at the Network Abstraction Layer (NAL), which is used for transmission or storage.

Encoder (Reconstruction path):

In the reconstruction path the quantized coefficients X are decoded, out of which the macroblocks are further derived from the reconstructed frame. These coefficients are further scaled and inverse transformed to produce a difference macro block Dn’. There may be some difference between the original macro block Dn and the difference macro block Dn’, which is the distorted version of Dn. After applying the prediction macroblock P, which is added to the Dn’, the reconstructed macroblock uF’n is created. A filter is applied to reduce the effects of blocking distortion, and the reconstructed reference frame is created from a series of macro blocks F’n.

1.1.4.2 Decoder [1]:

The compressed bit stream from the NAL is sent to the decoder. The quantized coefficients are produced by entropy decoding the data elements and reordering them. Using the inverse transformation and rescaling the Dn’ which is identical to the encoder Dn’ is produced.

Using the header information decoded from the bitstream, the decoder creates a prediction macroblock P, identical to the original prediction P formed in the encoder. P is added to Dn’ to produce uF’n which is filtered to create the decoded macroblock F’n. It should be clear from the Figures (Figure 1.1 and Figure 1.2) and from the discussion above that the purpose of the reconstruction path in the encoder is to ensure that both the encoder and decoder use identical reference frames to create the prediction P.

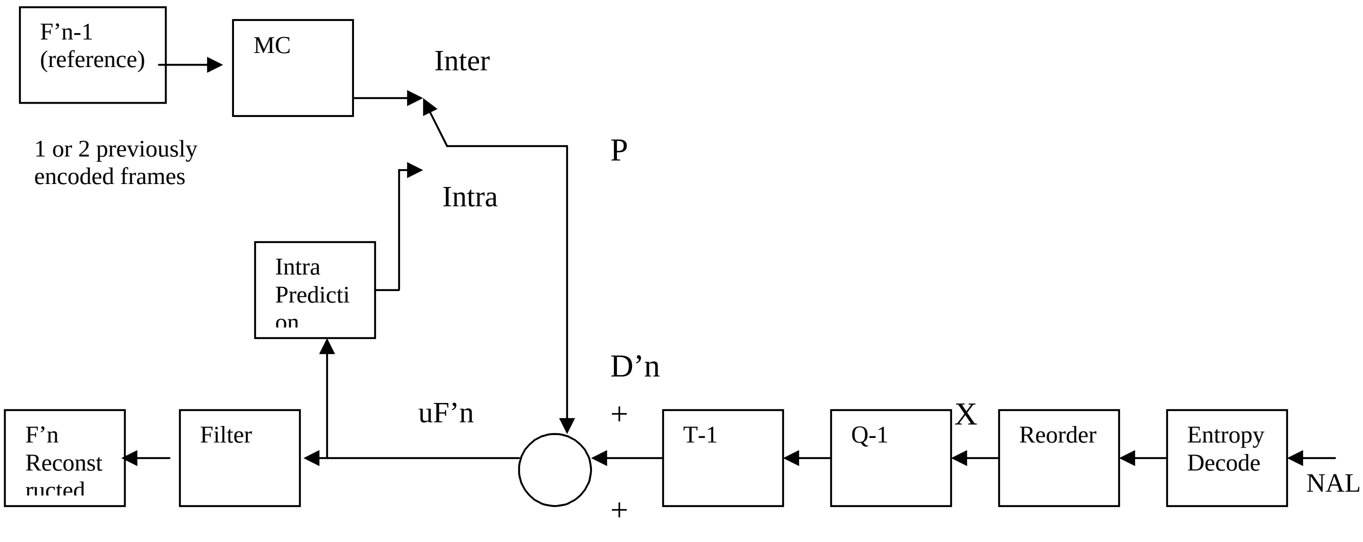


Figure 1.2: H.264/AVC Decoder

If this is not the case, then the predictions P in the encoder and decoder will not be identical, leading to an increasing error or "drift" between the encoder and decoder.

1.2 Objectives

The classical approaches for rate adaptation of the codec have been bit stream switching, which requires storing several pre-coded versions of the same video at different bitrates, or layered (scalable) video coding, which has coding efficiency and/or complexity penalties and is not suited to the varying network conditions over which the video is carried. The heart of the algorithm proposed here is a quantitative model, which includes the relationship between the Quantization Parameter (QP), the available bandwidth, and the packet loss count over the network. QP has no direct effect on the bitrates associated with overhead, prediction data, or motion vectors. The QP is taken into consideration because it influences the detail of information carried in the transformed Group of Frames (GOF). QP is initialized manually upon start of video sequence. Normally, a small initial Quantization Parameter (QPo) is chosen if the available channel bandwidth is wide and the packet loss is less, and a big QPo should be used if it is narrow and packet loss is more. As the bitrate is determined the encoder encodes the GOF using the target bit rate. The determination of the bit rate for every GOF makes the codec change the resolution of the video. At the cost of quality we get the complete video at the destination, which would not be possible if the rate had not changed. At a single bit rate the packet losses and the change in bandwidth causes the video to break in its flow, which is undesirable.

1.3 Organization of the Thesis

The organization of the remainder of this thesis is as follows: Chapter 2 provides the operation of the H.264 Video coding Layer, Chapter 3 gives a detailed understanding of the Network abstraction layer and the NAL units, Chapter 4 discusses the rate control parameters QP, bit rate, bandwidth and GOF into consideration, Chapter 5 presents the mathematical calculations, Chapter 6 shows the performance analysis of the H.264 codec based on the variable bandwidth and packet loss in a network using the proposed solution and Chapter 7 provides conclusions and future work

CHAPTER 2

OPERATION OF H.264 VIDEO CODING LAYER

2.1 Introduction to VCL [8]

The H.264 consists mainly of two conceptual layers, the Video Coding Layer (VCL) that defines the efficient representation of video, and the NAL that converts the VCL representation into a storage media or format suitable for transport layers such as bitstream format or packet stream transport, which are discussed later in detail in section 3.3 and 3.4 respectively. The division of this codec into these two layers helps derive the best possible output with the close integration of network abstraction layer and video coding.

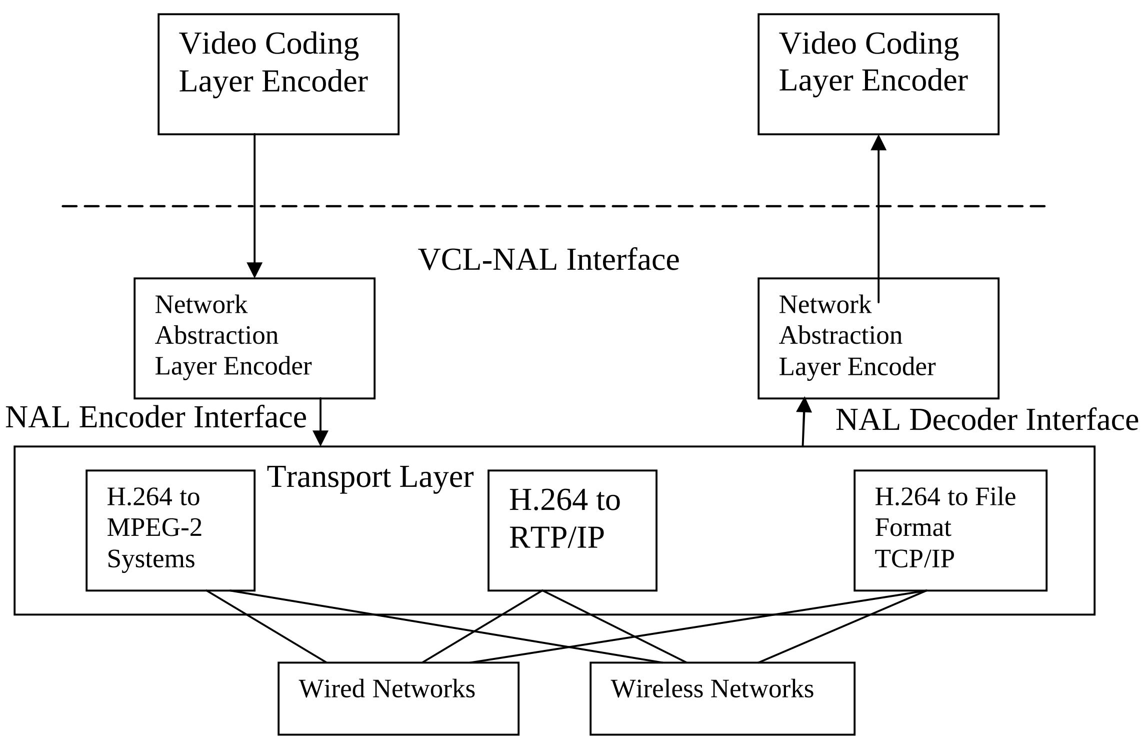


Figure 2.1: H.264 in a transport environment.

The block based hybrid video coding approach is followed by VCL design in which each coded picture is represented in block shaped units with the combination of luma and chroma samples called macroblocks. The block diagram of the VCL for a macroblock is shown in figure 2.2.

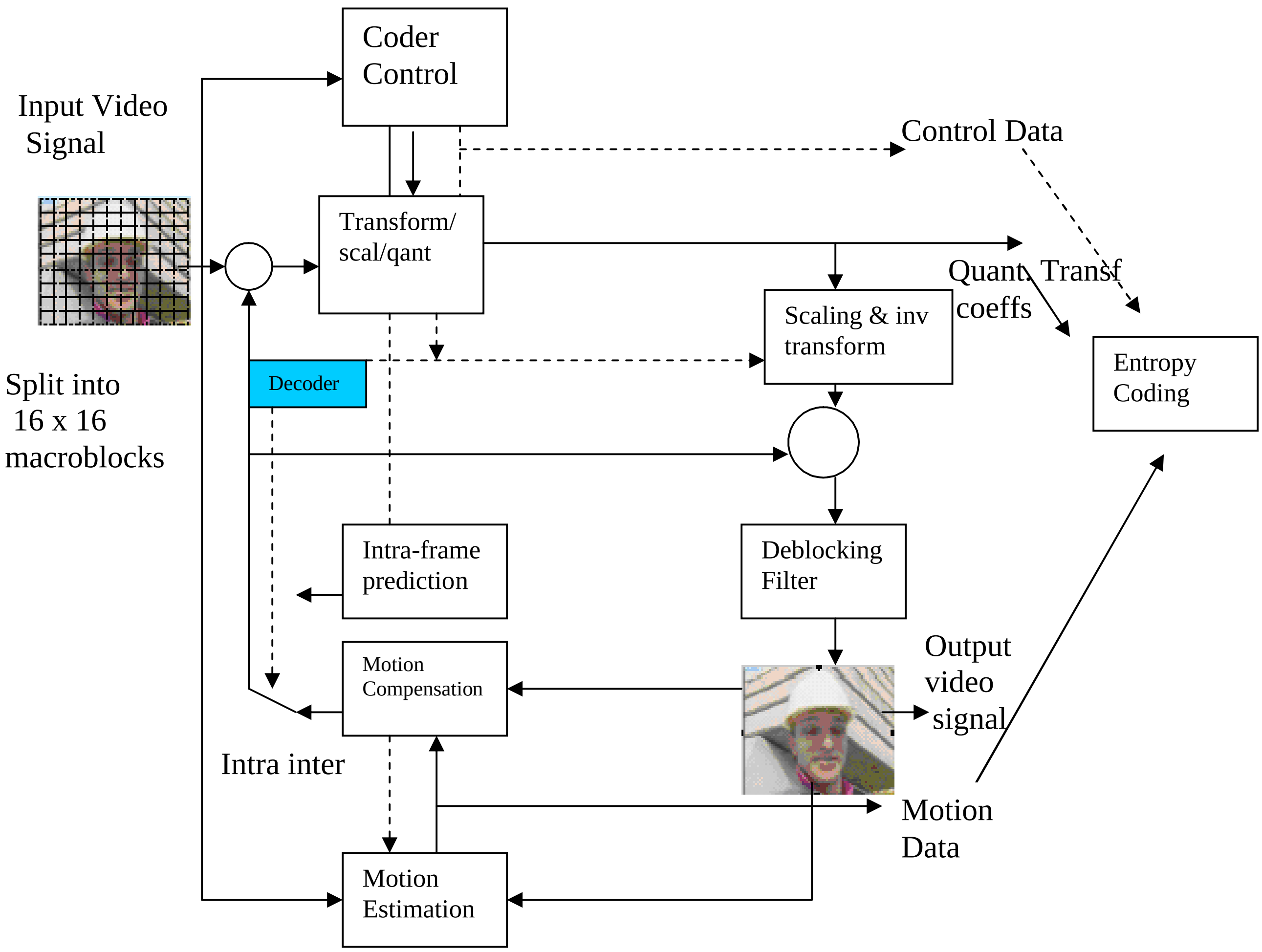


Figure 2.2: Basic coding structure of H.264/AVC for a macroblock

In the figure 2.2, on the left hand side, the picture is split into macroblocks. The first picture of a sequence is typically intra-coded, that means it uses information contained in the picture itself. Spatially neighboring samples of previously coded blocks are used to predict each sample of a bock in an intra frame. The encoding process takes place simultaneously in the encoder and decoder, which uses transmitted intra prediction side information to decide which and how neighboring samples will be used for intra prediction.

The encoder and decoder use motion data transmitted as side information by an encoding process, which includes reference pictures and spatial displacement to conduct inter prediction due to which remaining blocks of sequence are decoded. Inter coding uses information from previously decoded sample blocks.

Transform coefficients of the difference between the original and the predicted block which is called residual, are scaled and quantized for entropy coding and then transmitted and used as side information for either intra or inter prediction.

For the prediction of the next block or picture, the encoder contains the decoding (reconstruction path), which enforces the quantized transform coefficient to inverse scale and inverse transform in the same manner as at the decoder side, which outputs decoded prediction residual. This decoded prediction residual is added to the prediction, which is then provided as input to the deblocking filter, which in return results in the decoded video as its output.

2.2 Slices and slice groups

In H.264/AVC, the image is split into one more slices, which are sequences of microblocks. Slices are independent because their syntax can be transmitted from the bitstream if active sequence and image parameter sets are provided. If the reference pictures are the same at encoder reconstruction path and at the decoder, then regions represented by slices in the image can be perfectly decoded without using data from other slices.

Each slice group can be split into multiple slices, which contain sequences of macroblocks. Sets of macroblocks are mentioned by macroblock to slice group map that indicates identification number for each macroblock in an image indicating the slice group to which that particular macroblock belongs.

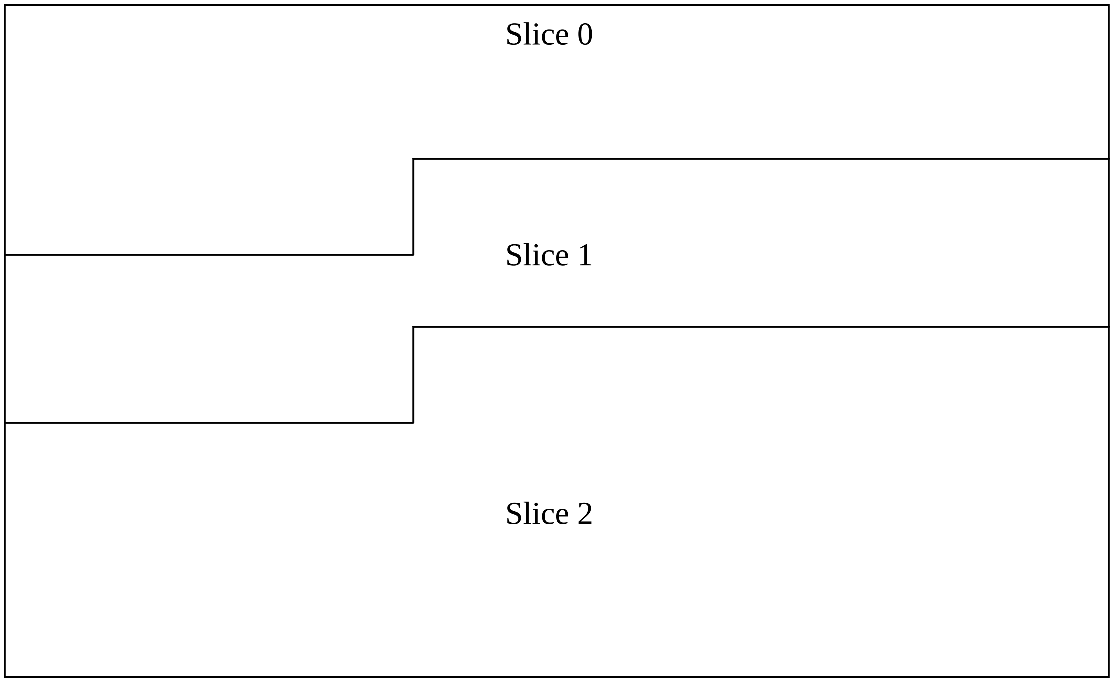


Figure 2.3: Partitioning of picture into slices

2.2.1 Macroblocks [2]

The picture is split into macroblocks that are the fundamental building blocks of a picture. Each macroblock consists of 16x16 luma pixels and 8x8 chroma pixels. Macroblocks are grouped into slices, which are again grouped as slice groups

2.2.2 Intra-frame Prediction I-Slice [20]

In H.264 intra prediction is performed using neighboring samples of blocks that are coded previously coded. This may cause an error to propagate in the macroblocks that are coded with inter prediction. So, a restriction is applied to intra prediction so that it can be coded using neighboring intra coded macroblocks only. In I-slices all macroblocks in a slice are predicted using intra prediction alone.

2.2.3 Inter-frame Prediction in P-Slices [20]

Some macroblocks of P-slice can also be coded in addition to the coding types like the I-slice. With the help of memory management control operations mentioned in the bitstream, the decoder duplicates the multi-picture buffer of the encoder. The index of the reference picture inside the multi-picture buffer should be signaled till the size of multi-picture buffer is set to one. That means it can be coded using inter prediction with at most one motion-compensated prediction signal per prediction block only.

2.2.4 Inter-frame Prediction B-Slices [20]

Some macroblocks of B-slice can also be coded in addition to the coding types available in a P-slice. These can be coded using inter prediction with weighted average of two distinct motion-compensated prediction signals per prediction block.

CHAPTER 3

NETWORK ABSTRACTION LAYER DETAILS

3.1 Introduction to NAL [16]

As seen, the video coding layer represents the content of video in an efficient manner whereas the NAL forms the interface between the codec and the outside world. The NAL unit consists of a single byte header and a bit stream representing the macroblocks of a slice in an encoded form. In order to provide support for both the stored media and streaming media the NAL unit specifies two different formats: one is the bitstream format and the other is Real Time Protocol (RTP) payload format. The only difference between the two is that the bitstream format uses a start code prefix to distinguish between two NAL units and the RTP format uses the RTP header.

3.2 NAL unit structure

The structure of Network Abstraction Layer Units (NALU) is as follows:

```
+---------------+
|0|1|2|3|4|5|6|7|
+-+-+-+-+-+-+-+-+
|F|NRI|  Type   |
+---------------+
```

F – 1 bit: It represents the “forbidden_zero_bit”. According to the H.264 specification, if the value of this bit is 1 it then indicates that there is a syntax violation in the NAL unit. At the NAL decoder interface it is assumed that the NALUs are delivered in the order in which they were transmitted. If the packets are not received correctly, or if they consist of bit errors, then this bit is set by the media aware network element. The decoder or any other gateway can decide whether this errored NALU is decoded or disposed. This also helps prevent error propagation.

NRI – 2 bits: If these bits are “00” they indicate that the content of the NAL unit is not used to reconstruct reference pictures for inter prediction and hence such NAL units can be discarded if the value of these bits is greater than 00 then it indicates that the NAL units are required to maintain the integrity of the reference pictures. The highest priority for transport is 11, followed by 10, then 01, and finally 00 is the lowest.

Type – 5 bits: It represents the nal_unit_type. This component specifies the NAL unit payload type. This either indicates the VCL "slice type" like the intra slice, p-slice, or b-slice, or high-level information such as the random access points and the sequence and picture parameter set information. The NAL units, which are taken into consideration in this thesis, are as below:

NAL Unit Type	Content of NAL unit	NRI (binary)
1	non-IDR coded slice	10
2	Coded slice data partition A	10
3	Coded slice data partition B	01
4	Coded slice data partition C	01
5	IDR coded slice	11
7	Sequence parameter set	11
8	Picture parameter set	11

The NAL unit types 6,9,10,11 and 12 have the lowest priority with the NRI = 00. Only these NAL units are allowed to transmit as 1 NAL unit in a packet. The types from 13 to 23 are reserved for ITU-T and ISO/IEC and the semantics of the types 0, 30 and 31 are not defined. The remaining bytes consist of the payload data. The payload data in the NAL unit is interleaved as necessary with emulation prevention bytes, which are bytes inserted with a specific value to prevent a particular pattern of data called a start code prefix from being accidentally generated inside the payload.

3.3 NAL units in bit stream format use

The H.264 defines a byte stream format for certain systems, which require the delivery of the entire or partial NAL unit stream as an ordered streams of bytes or bits within which the location of the NAL unit stream needs to be identifiable from patterns within the coded data itself. In this format each of the NAL unit is prefixed with a bit string, which represents the start of the unit. The bit string is usually 16-24 bits long and depends on the implementation of the encoder and decoder. The boundaries of the NAL unit can then be identified based on the start code prefix. For systems that operate and provide streams of bits without alignment to byte boundaries there is an addition of a small amount of additional data added for the decoders to decode the stream. In order to prevent the start code prefix the emulation prevention bytes are added in the stream.

3.4 NAL units in packet-transport system use

In the systems, which use the packet networks like the Internet Protocol (IP), and the RTP systems, a packet structure is formed which is framed by the transport protocol used by the system. In such systems there is no need to use the start code prevention bytes, which would be a waste of data carrying capacity, so instead the NAL units can be carried in data packets without start code prefixes.

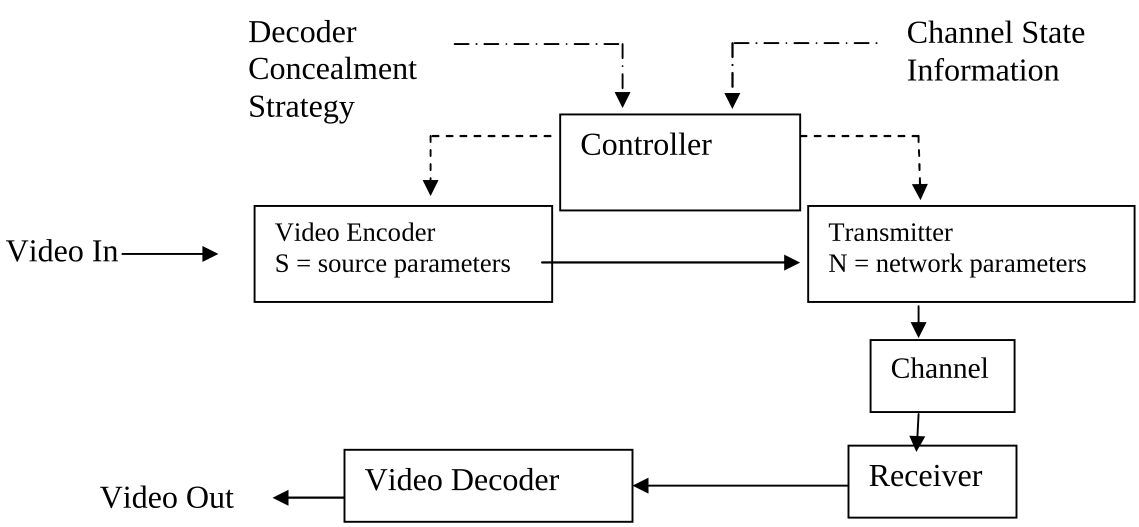


Figure 3.1: Packet-based video transmission system diagram

3.5 Parameter sets [2]

To solve the problem of packet-lossy environments, a new approach was taken in H.264/AVC video. The synchronous, real-time media transmission consists only of packets that are completely independent. That is, without depending on other packets each packet can be reconstructed. The slice layer was considered the smallest self-contained unit, because the size of the slices can be adjusted to fit into the Maximum Transmission Unit (MTU) size of almost all-demanding systems. In higher layers whole information is carried asynchronously. During the existence of a session a few pictures and GOF parameters may change. Parameters that change very frequently are added to the slice layer. All other parameters are collected in a Parameter Set. Encoder and Decoder(s) maintain a synchronized set of such Parameter Sets. Figure 3.2 depicts this concept. The Parameter Set information packet is a method specified by Joint Video Team (JVT) standard to carry parameter sets in special NALU types. The Parameter Set concept allows the usage of different logical channels or even different out-of-band control protocols to convey parameter sets from the encoder to the decoder.

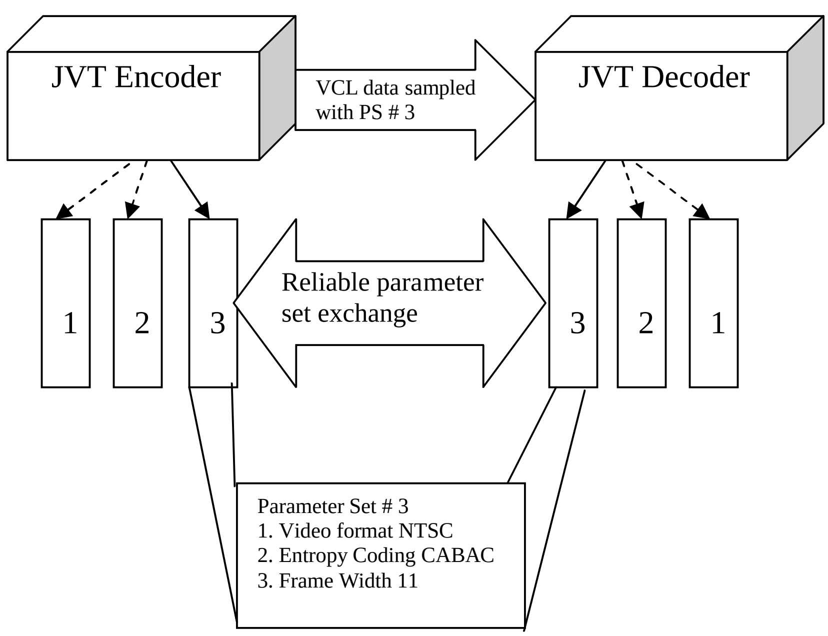


Figure 3.2: Parameter set concept

A parameter set is supposed to contain information that is expected to rarely change and offers the decoding of a large number of VCL NAL units. There are two types of parameter sets:

- sequence parameter sets, which apply to a series of consecutive coded video pictures called a coded video sequence, and
- picture parameter sets, which apply to the decoding of one or more individual pictures within a coded video sequence.

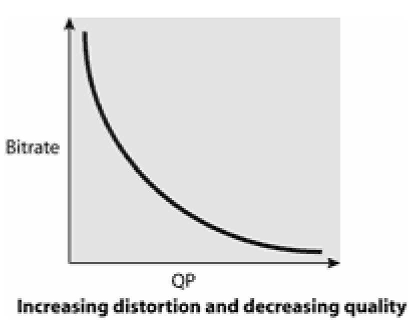
CHAPTER 4

RATE CONTROL PARAMETERS

4.1 Introduction to Rate Control Mechanism [13]

The purpose of the rate control algorithm in this thesis is to dynamically adjust the encoder parameters to achieve the target bit rate.

MPEG and the H.26* families of encoding schemes are inherently lossy processes, which achieve compression not only by reducing the amount of data that needs to be sent, but also by making small quality compromises, which are minimally perceptible. The main factor affecting the spatial detail is the QP. When the QP is small it helps retain all the detail of the video and when the QP is increased some of the detail is reduced, which results in the reduction of the bitrate. Hence, varying the QP varies the bitrate. Therefore, developing a direct relationship between the QP and the bitrate would help make the codec adaptable. Reduction of bitrate leads to an increase in distortion and some loss of quality. On the other hand, it would increase in bitrate and provide more details.



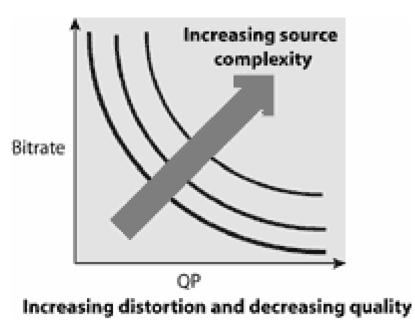


Figure 4.1: For a particular source frame **Figure 4.2: When source complexity varies**

Figure 4.1 and 4.2 depict the relation between bitrate and QP. As shown in the figure 4.1 if we decrease bitrate, we have to increase QP; this increases distortion and consequently, decreases quality. As shown in the figure 4.2 we move from one curve to another of this kind along with the source complexity variations during a sequence.

The network bandwidth is unpredictable and varies during the session. Some problems may arise due to varying bandwidth like delay and packet loss. To resolve problem we can apply a rate control algorithm, which dynamically adjusts encoder parameters - mainly QP to achieve a target bit rate suitable for the available bandwidth.

There are few types of rate control methods-

- GOF (Group of frames) layer rate control.
- Frame layer rate control.
- Macro block level rate control, etc.

4.2 GOF Layer

The GOF layer consists of one or more consecutive frames in which the first frame in the GOF is always an intra frame.

The functionalities of GOF are as follows,

- Carries information attributed to a group of frames
- Random Access Functionality
- Not only for storage applications many other applications need random access functionality. Some examples of applications are, random access for storage application, like DVD, broadcasting channels, splicing, editing, etc.
- Re-synchronization Tool for Error Recovery
- In real applications GOF structure is considered to be a useful tool, and hence in JVT codec it is supported at an appropriate layer.

This thesis implements GOF layer rate control, which uses the available network bandwidth. It can be either a Constant Bit Rate (CBR)(Figure 4.4) or a Variable Bit Rate (VBR)(Figure 4.3). In this thesis CBR case is implemented.

In VBR encoding two values are provided by the user, uncompressed video source and a value for QP. We get output video of moderately constant quality, but at the cost of a dramatic change in bitrate because in the real world the complexity of pictures changes continuously.

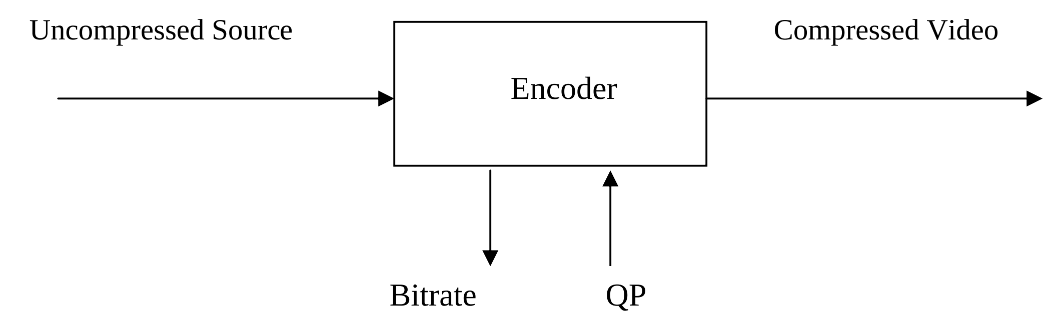


Figure 4.3: Open Loop Encoding (VBR)

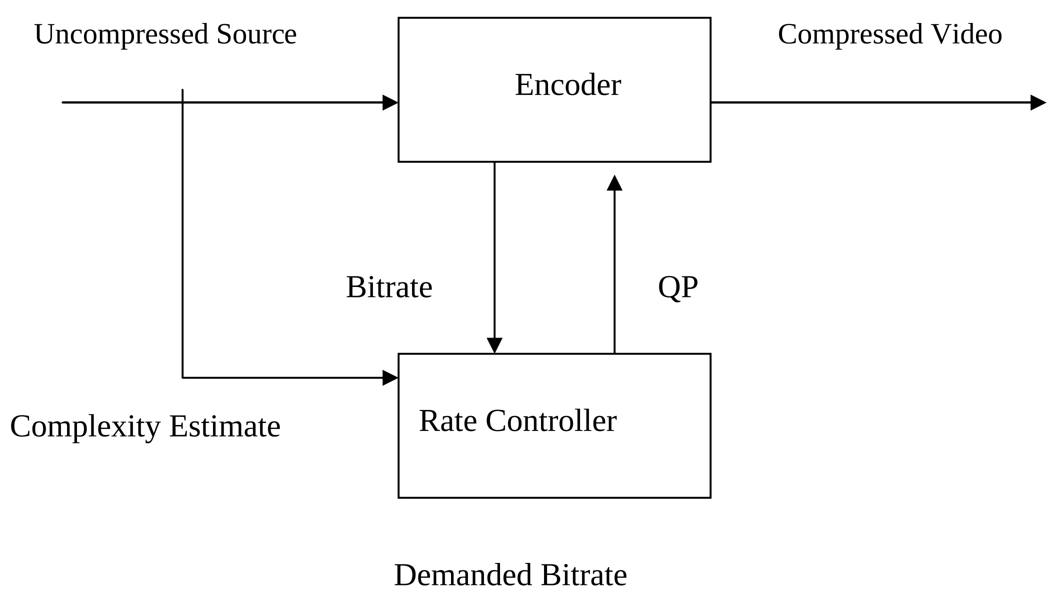


Figure 4.4: Closed Loop Rate Control (CBR)

In CBR the user instead of QP, provides the encoding demanded bitrate. On the basis of estimates of the source complexity one must dynamically vary QP for every frame (or group of frames) to get the proper allocation of bits to work with.

4.3 Terminologies

The terminologies used in the thesis to perform the GOF level rate control are as follows,

4.3.1 Group of Frames (GOF)

In the GOF concept all the P and B frames follow an I-frame until occurrence of the next I picture. For example a typical MPEG GOF structures may look like, IBBPBBPBBI. Although H.264 follows one I frame per video sequence, for more efficient output video the adaptive rate control concept requires a repeating GOF structure. Therefore, IntraPeriod parameter is never set to 0 for H.264 rate control to work properly.

4.3.2 Bitrate

Bitrate refers to the bits per second consumed by a sequence of pictures, i.e., bitrate = (average bits per picture) / (frames per second).

4.3.3 Quantization

A quantization parameter is used for determining the quantization of transform coefficients in H.264/AVC. The parameter can take 52 values. These values are arranged so that an increase of 1 in quantization the parameter means an increase of the quantization step size by approximately 12% (an increase of 6 means an increase of the quantization step size by exactly a factor of 2). It can be noticed that a change of step size by approximately 12% also roughly means a reduction of bit rate by approximately 12%.

4.3.4 Quantization Parameter (QP)`

The QP determines the step size for associating the transformed coefficients with a finite set of steps. Large values of QP represent big steps that crudely approximate the spatial transform, so that most of the signal can be captured by only a few coefficients. Small values of QP more accurately approximate the block's spatial frequency spectrum, but at the cost of more bits. In H.264, each unit increase of QP lengthens the step size by 12% and reduces the bitrate by roughly 12%.

4.3.5 Basic unit

Basic unit expresses the granularity on the basis of which QP is adjusted in the feedback control loop. If the basic unit is a frame, then the adjustments to QP are uniform across the frame.

In this thesis, the basic unit is a frame. Initially, most H.264 applications will probably use the picture as the basic unit, but ultimately all or few numbers of frames are expected to yield the best possible results with uniform bitrate and uniform quality.

CHAPTER 5

MATHEMATICAL ANALYSIS

5.1 QP Calculation using reference paper method [11]

The reference paper method has implemented a static initial table, showing an approximate mapping between the quantization parameters and the bits per pixel obtained. Every time a new I-frame is being coded, the desired bitrate is read and the target bits-per-pixel (bpp) indicator is computed according to the frame size and frame rate. The initial quantization parameter is then chosen from the table as the one ensuring the closer bpp indicator. Every time a GOF terminates, and right before starting the following I frame, the bpp obtained for the last GOF is stored in the table together with its average quantization parameter, so updating the starting static values at each step to better fit over the sequence characteristics.

As the reference paper calculates the total number of bits allocated for a GOF based on the available bandwidth, it means that the bandwidth is assumed to be constant for the entire GOF. This is not the situation every time, because if there are more users or if there is a burst of traffic in the network while the GOF is encoded and transmitted then it may cause the packet to be dropped. This may cause the video quality to degrade due to loss of frames in the form of NAL units as seen in the analysis section. This thesis uses the direct relationship between the QP, which is in a reference table calculated on extensive simulations for various types of video frames based on the bandwidth as well as packet drop consideration and the bitrate. The results show that this reference table method for a direct relationship between the bitrate and QP performs better than the previous method.

5.2 Proposed Method

An efficient technique based on available bandwidth and packet loss for GOF (Group of Frames) level adaptive rate control for the H.264 video codec over a network is as follows:

5.2.1 Input Video variables:

F_r = Frame rate of video. This value along with the preset QP determines the bitrate of the codec

N_{gof} = The total number of frames in a GOF.

5.2.2 H.264 Codec Variables:

B_r = Default bit rate of the codec which is equavalent to the full available bandwidth (1.544 Mbps). This can be set manually based on the available full bandwidth and the initial QP.

C_{Br} = Current bitrate of the codec. This is selected based on the packet loss and the current available bandwidth of the network

QP = Quantization Parameter, which is stored in the table corresponding to the bitrate B_r calculated over a period of time for various sequences of video. Initial QP is chosen based on the available bandwidth for the frames and the packet loss estimation of the previous frames.

5.2.3 Network Parameters:

B = Available Effective bandwidth on the network calculated based on the lowest link. (Serial – 1.544 Mbps)

NALU = NAL units are variable in size and depend on the video complexity in the I, P, B frames.

D_l = Data Lost in the network (Packet loss due to corruption or congestion) in bits.

R_l = Random packet loss due to burst in traffic over the network, which further reduces the savailable effective bandwidth calculated over a GOF of video.

QP	Bit Rate (Kbps)
9	1540.66
10	1480.63
11	1211
12	982.46
13	871.64
14	752.29
15	665.92
16	584.19
17	509.01
18	440.7
19	397.78

Table 5.1: QP and corresponding bitrate with N_{gof} = 10

The relationship for the QP and the bitrate from the observations can be calculated as per the algorithm below

If (B = $C_{Br(i-1)}$ - random (Y)) then

QP = QP + 1

Else If (B = $C_{Br(i-1)}$ + random (Y)) then

QP = QP - 1

Where 100 <= Y <= 200 in these simulations calculated per the observation.

5.3 **Analysis**:

Probability of data lost in terms of NAL units over the network when GOF of video is sent can be calculated as shown below

D_{lp} = mod [(C_{Br} - B) / B] + R_l if C_{Br} > B

= 0 + R_l if C_{Br} < B (would be in most of the cases) (5.1)

D_l = 8 * Dlip bits (5.2)

The Dl, is dropped from the bitstream of the encoder output due to the lack of bandwidth or other environmental parameters, which cause the NALU to corrupt. This creates a loss of frames in the video, which is unacceptable.

We determine the current bitrate needed using the following formula,

$$C_{Br} = (\mathrm{B} - D_l) \tag{5.3}$$

This value of C_{Br} is used as an index to the reference table, which consists of the C_{Br} and the QP. The required QP is determined to further encode the next GOF or partial GOF, which is based on this QP. This way the bitrate is reduced and the data sent is adapted as per the available bandwidth,

This shows we can consider the available bandwidth as well as the packet loss over a network when H.264 codec is used for video transmission. This is the efficient way to tune the bit rate of the codec so that even though at the cost of little less resolution in the video we will get no frame/packet loss causing the distortion or breaking in the video sequence. Hence, we can improve the performance of the codec.

CHAPTER 6

PERFORMANCE ANALYSIS OF THE PROPOSED SOLUTION

6.1 Encoder/Decoder Software

The JM Reference software encoder and decoder version 10 was used in the simulations. This encoder/decoder provides the output of the encoding and decoding in a detailed manner. The advantage of the encoder is that it can be run using a parameter file (encoder.cfg and decoder.cfg), which describes a list of values such as the frame rate, QP, start frame to encode and number of frames to encode which determine the GOF of the encoder. This output of the encoder consists of the console print of the various values like the bitrate used, the number of the bits used to encode, and the average SNR values for the Y, U and V components over the group of frames of the video encoded. The encoder also produces a reference output of the video after its encoding is done. The decoder takes the encoded bitstream along with the reference output produced by the encoder as input and then uses it to decode the stream and produce the corresponding output video file.

6.2 Testing Objectives

The research work consists of analyzing the behavior of the coder/decoder for ".yuv" type video format video clips. Initial QP is chosen based on the network bandwidth available as per the reference table. The reference table is created based on the extensive simulations for manually provided QP values ranging from 0 – 51 for different video files. This table determines the QP corresponding to its bitrate, which is used by the encoder. By varying the QP the bitrate for the GOF gets varied per the available bandwidth, which is taken as a random number. This helps the encoder encode the video file using the available bandwidth conditions. The packet drop probability is calculated based on the available bandwidth and the current bitrate. This helps to drop the NAL units from the H.264 encoded stream manually. A comparison between the decoded stream, which consists of the dropped NAL units, and a decoded stream, which consists of the stream that is encoded using an appropriate QP, is done. The comparison consistently shows that varying the QP per the network bandwidth produces a better result than the one produced after the random drop occurs in the H.264 encoded stream.

6.3 Testing Procedure

The research conducted had the following steps.

- Create the reference table consisting of the bitrate and the QP needed for it.
- Set the basic parameters in the parameter file of the encoder like the framerate of the video and the GOF, which is (Start Frame + Number of frames encoded).
- Run the encoder with the other parameters, which would change for every Group of Frames of video like the QPISlice, QPBSlice and QOPSlice, which correspond, to the QP for the I, B and P frames of video. These QP values are varied from its range, 0-51, which makes the encoder encode the frames with the varying bitrate. These values of the bitrate and the corresponding QP are stored in a reference table.

A random drop of NAL units was performed in the encoded video stream, which simulates the drop that would have occurred if there was less bandwidth available than that currently available, as well as if there was corruption in the NAL unit during its transit in the network

- The packet drop probability is calculated based on the available bandwidth over the
- network and the current bitrate of the codec used.
- The NAL units which have the size equal to the packet drop probability are calculated and the number of bytes, which may be dropped, are chosen. These NAL units are then randomly dropped off the encoded video stream.

- Now the encoded H.264 video stream consists of the missing NAL units which would represent the packet loss in the network
- This encoded video stream is then decoded and the SNR values are calculated.

The SNR values are also calculated based on the bitrate, which would be needed as per the reference table by choosing the appropriate QP.

- The value of the current bit rate C_{Br} is calculated as per the equation 3.5 above.
- This value is used as a pointer in the reference table to choose the QP so that it will make the encoder encode the video at the desired C_{Br} .
- Once the video is encoded, the SNR values are calculated

Observations show that the reference table helps select the most appropriate bitrate values which helps improve the SNR value in the case the packet dropped stream is decoded.

Table 6.1. shows the values calculated for QP=11 for a video "test.yuv" which consists of 90 frames at the rate of 30 frames per second. The N_{gof} = 3.

The table shows the SNR calculation for QP = 11, which is chosen as the initial value based on the reference table.

The full bandwidth over the network = 1544 Kb. Hence as per the table 1486Kb is chosen as the current bandwidth and QP =11 is selected.

Further, as the NAL units are individual entities and are transmitted over a single RTP packet, the drop of the bytes in the table is in terms of the NAL units and the bytes corresponding to them.

Available Bandwidth	1486 Kb	1350 Kb	1210 Kb	1100 Kb
Probability of packet dtop	0	0.1007	0.228	0.3509
Bytes dropped per probability	0	1711.9	7866	16930
Bytes to be dropped	0	17000	34500	48250
Bytes dropped during simulation	0	4000 - 1 NAL Units	9000 - 2 NAL Units	20000 - 3 NAL units
SNR Y	49.53	40	39.05	37.17
SNR U	49.83	33.85	31.91	28.05
SNR V	50.09	32.56	34.39	28.9

Table 6.1: SNR calculation for bandwidth variation with N_{gof} = 3

When the packet drop occurs the SNR values are shown in the table to be reducing, meaning that the frames are lost and the video quality is poor.

Table 6.2 shows the values of the SNR when the QP is adjusted as per the reference table based on the calculated required bitrate.

Available Bandwidth	1486 Kb	1350 Kb	1210 Kb	1100 Kb
Variable QP	11	12	13	14
SNR Y	49.53	48.7	48.09	47.32
SNR U	49.83	49.23	48.64	48.09
SNR V	50.09	49.47	48.95	48.41

Table 6.2: SNR calculation for bandwidth variation by changing QP with N_{gof} = 3

The table 6.2 shows that the SNR calculation over the same bandwidth when compared to the one when the packet is dropped in table 1 has significant improvement as with the acceptable low resolution video there are no packet/frame drops.

Table 6.3 shows the values calculated for QP=9 for a video "test.yuv" consisting of 90 frames at the rate of 30 frames per second. The GOF = 10. Here, the QP is 9 because the number of frames encoded is 10, which is determined from the reference table as per the bandwidth. This shows that when the frames increase, the QP required for the bandwidth decreases and hence we can have better quality for more frames in N_{gof} .

Available Bandwidth	1486 Kb	1350 Kb	1210 Kb	1100 Kb
Probability of packet dtop		0.1007	0.228	0.3509
Bytes dropped per probability		1711.9	7866	16930
Bytes to be dropped		17000	34500	48250
Bytes dropped during simulation		5000 - 1 NAL Units	10000 - 2 NAL Units	16000 - 3 NAL units
SNR Y	51.04	35.24	34.99	34.71
SNR U	51.23	27.66	26.83	25.9
SNR V	51.88	28.8	27.33	26.49

Table 6.3: SNR calculation for bandwidth variation with N_{gof} = 10

Table 6.4 shows the improvement in the SNR when the QP is chosen and changed, which changes the bitrate for same value of the available bandwidth as table 6.3.

Available Bandwidth	1486 Kb	1350 Kb	1210 Kb	1100 Kb
Variable QP	9	11	11	11
SNR Y	51.04	48.68	48.68	48.68
SNR U	51.23	49.01	49.01	49.01
SNR V	51.88	49.99	49.99	49.99

Table 6.4: SNR calculation for bandwidth variation by changing QP with N_{gof} = 10

This way we conclude that if the reference table is created for each network and maintained based on which QP is chosen for the bandwidth the performance of the coder improves.

Table 6.5 and Table 6.6 – Comparison between variable and constant QP for bandwidth variation

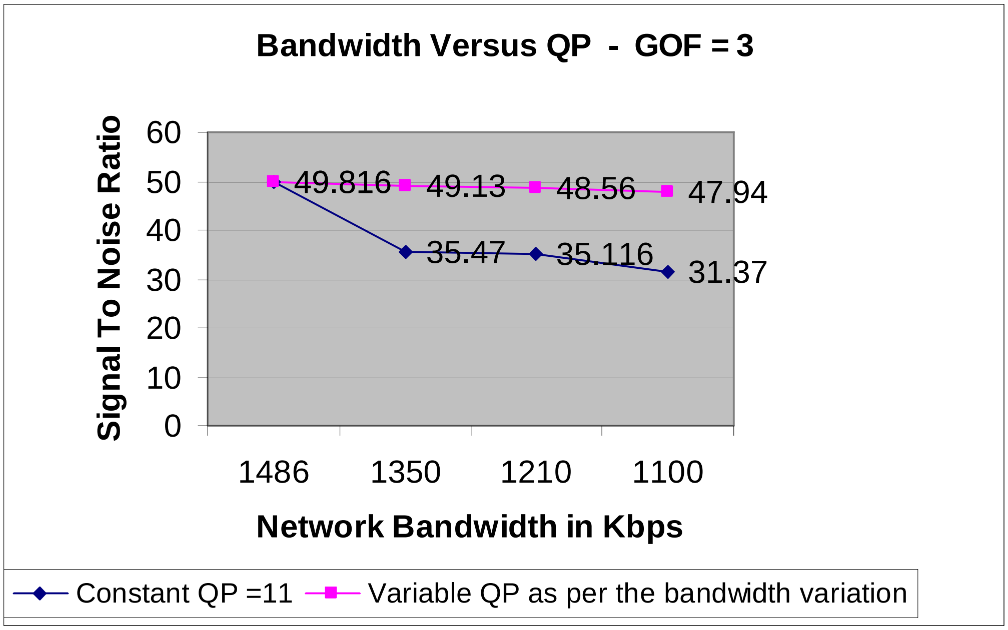


Table 6.5: Bandwidth versus SNR with N_{gof} =03

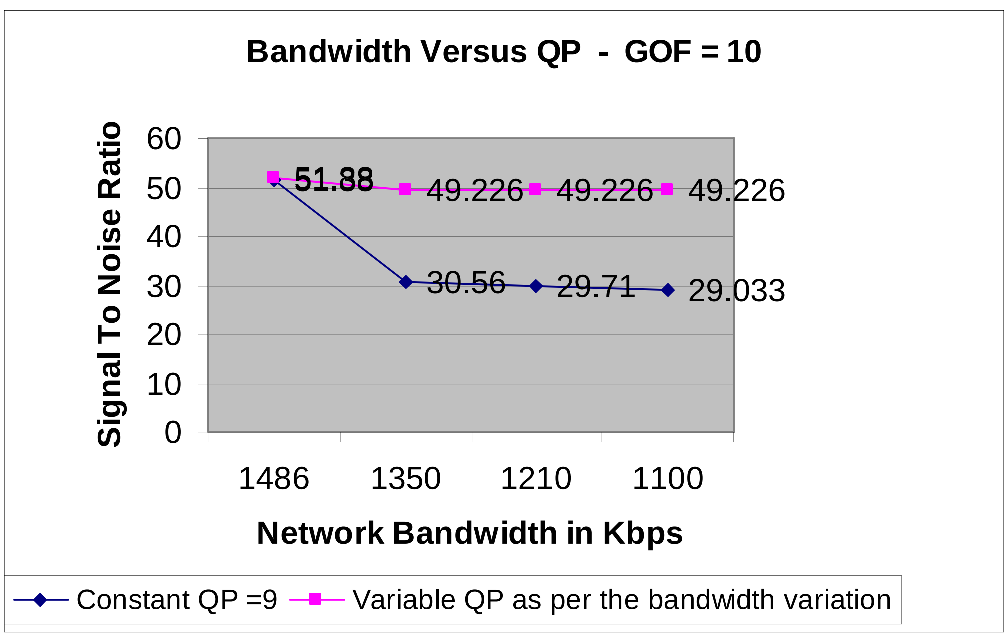


Table 6.6: Bandwidth versus SNR with N_{gof} =10

Table 6.7 shows the relationship between the QP and the available bandwidth.

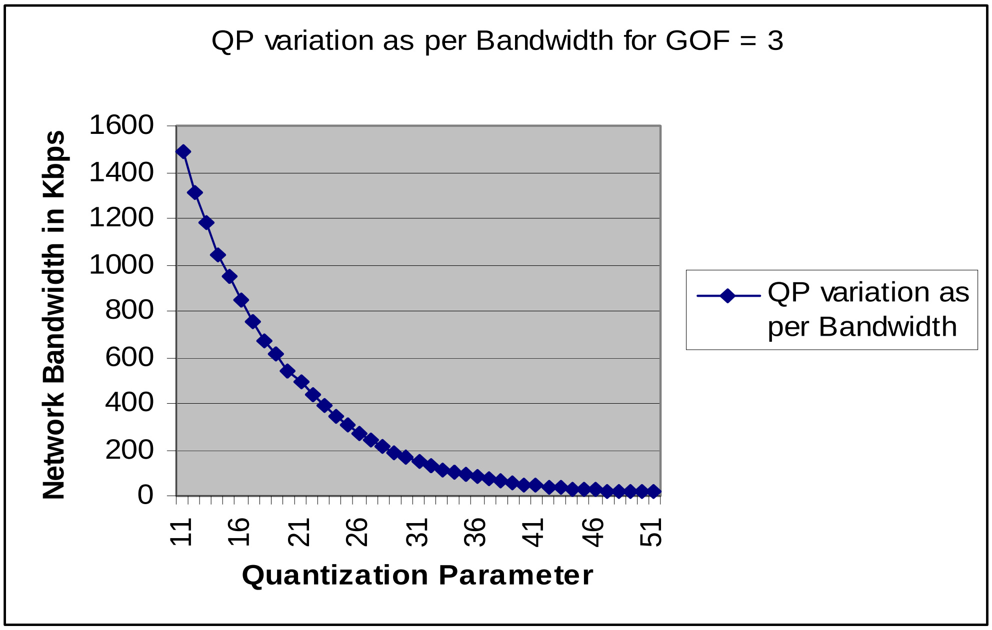


Table 6.7: QP versus Bandwidth

CHAPTER 7

CONCLUSION AND FUTURE WORK

The solution provided improves the performance of the video quality over a network using H.264 codec. With the varying available bandwidth conditions there may be packet losses, which may lead to dropping of frames and bad video quality determined by the SNR output.

The research implemented was based on:

Changing the Quantization parameter of the H.264 Reference JM coder/decoder to make it adaptable to the network. Due to network conditions there may be less bandwidth available for the video to reach the client in the exact encoded form. There may be issues like the packet drop and the packet corruption, which again leads to the packet being dropped by the decoder. The H.264 video stream is encoded in NAL units, which are sent in the RTP payload over the network. The entire NAL unit consists of one slice of picture (I, P, B) If the packet is lost then the whole slice of video is lost.

By changing the quantization parameter of the coder based on the current available effective bandwidth a significant improvement in the SNR ratio was observed as seen in the analysis section.

The reference table is created by running the encoder for various video sequences, and stored which consists of the QP and the bitrate relationship. Upon detection of a bandwidth change, the reference table is consulted and the appropriate QP of the encoder is set. The comparison of the SNR of the encoded stream with the packet loss, and the encoded stream with appropriate QP shows this solution improves the video quality. At the cost of reduction of video resolution we get better video performance as the frames are not dropped in the transit of network.

In the future, this reference table can be created at the macroblock, frame or pixel level, and the performance can be compared to this GOF level to see if it improves the video performance.

REFERENCES

[1] Iain E G Richardson, " H.264 / MPEG-4 Part 10: Overview", www.vcodex.com.

[2] Jorn Ostermann, Jan Bormans, Peter List, Detlev Marpe, Matthias Narroschke, Fernando Pereira, Thomas Stockhammer, and Thomas Wedi, "Video coding with H.264/AVC: Tools, Performance, and Complexity", IEEE CIRCUITS AND SYSTEMS MAGAZINE, FIRST QUARTER 2004.

[3] Thomas Wiegand, Gary J. Sullivan, Gisle Bjontegaard, Ajay Luthra, "Overview of the H.264 / AVC Video Coding Standard", IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, JULY 2003.

[4] Rui Zhang, Shankar L. Regunathan, Kenneth Rose, " Video Coding with Optimal Inter/Intra-Mode Switching for Packet Loss Resilience ", IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 18, NO. 6, JUNE 2000.

[5] 1 ITU-T Rec. H.264 / ISO/IEC 11496-10, “Proposed Draft of Adaptive Rate Control”, Final Committee Draft, Document JVT-H014, May 2003.

[6] 1 ITU-T Rec. H.264 / ISO/IEC 11496-10, “Advanced Video Coding”, Final Committee Draft, Document JVT-E022, September 2002.

[7] WENGER Stephan, WANG Ye-kui, HANNUKSELA Miska M., " RTP payload format for H.264/SVC scalable video coding", Journal of Zhejiang University SCIENCE A.

[8] Ralf Schafer, Thomas Wiegand and Heiko Schwarz, "The Emerging H.264/AVC Standard", Heinrich Hertz Institute, Berlin, Germany, EBU TECHNICAL REVIEW - January 2003.

[9] Simone Milani, Luca Celetto, Gian Antonio Mian, "A Rate Control Algorithm for the H.264 Encoder", Dept. of Information Engineering, University of Padova.

[10] Stephan Wenger, " SGN-3156 Video Compression H.264/AVC Network Abstraction Layer", Tampere University Of Technology.

[11] Fabio De Vito, Tanir Ozcelebi, Reha Civanlar, A. Murat Tekalp, Juan Carlos De Martin, "RATE CONTROL FOR GOP-LEVEL RATE ADAPTATION IN H.264 VIDEO CODING", Department of Electrical Engineering, Koc University.

[12] Xiaoqing Zhu, Eric Setton and Bernd Girod, "CONTENT-ADAPTIVE CODING AND DELAY-AWARE RATE CONTROL FOR A MULTI-CAMERA WIRELESS SURVEILLANCE NETWORK", Information Systems Laboratory, Stanford University, Stanford.

[13] www.pixeltools.com/rate_control_paper.html

[14] www.pixeltools.com/h264_paper.html

[15] Fabio De Vito and Juan Carlos De Martin," SNR Control for GOP-level Constant Quality in H.264 Video Coding", Dipartimento di Automatica e Informatica Politecnico di Torino.

[16] S. Wenger, M.M. Hannuksela, T. Stockhammer, M. Westerlund, D. Singer, " RTP Payload Format for H.264 Video", http://www.rfc-archive.org.

[17] Xiaoqing Zhu, Eric Setton and Bernd Girod," CONTENT-ADAPTIVE CODING AND DELAY-AWARE RATE CONTROL FOR A MULTI-CAMERA WIRELESS SURVEILLANCE NETWORK", Information Systems Laboratory, Stanford University.

[18] Thomas Stockhammer, Institute for Communications Engineering, Munich University of technology, Miska M. Hannuksela, Nokia Mobile Software, Nokia Corporation, Stephan Wenger, Communication and Operating Systems, Technical University Berlin, "H.26L/JVT CODING NETWORK ABSTRACTION LAYER AND IP-BASED TRANSPORT".

[19] Gary J. Sullivan, Microsoft Corporation, Pankaj Topiwala, FastVDO LLC, and Ajay Luthra, Motorola Inc., " The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions", SPIE Conference on Applications of Digital Image Processing XXVII Special Session on Advances in the New Emerging Standard: H.264/AVC, August, 2004.

[20] Multimedia API for Linux Programming Guide – January 2006, Intel Corporation.

[21] Aggelos K. Katsaggelos, Yiftach Eisenberg, Fan Zhai, Rnadall Berry, Thrasyvoulos N. Pappas," Advances in Efficient Resource Allocation for Packet-Based Real-Time Video Transmission", PROCEEDINGS OF THE IEEE, VOL. 93, NO. 1, JANUARY 2005.

[22] Israel Cidon, Asad Khamisy, Moshe Sidi, "Analysis of Packet Loss Processes in High-speed Networks", IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 39, NO. 1, JANUARY 1993.

[23] Li-Wei Kang and Jin-Jang Leou, "AN ERROR RESILIENT CODING SCHEME FOR H.264 VIDEO TRANSMISSION BASED ON DATA EMBEDDING", Department of Computer Science and Information Engineering, National Chung Cheng University.

[24] K.W. Roger Cheuk, S.H. Gary Chan, K.W. Alice Mong, C.-M. Martin Lee, S.-S. Shirley Sy, "Developing PDA for Low-Bitrate Low-Delay Video Delivery", Department of Computer Science, Hong Kong University of Science and Technology

[25] Krzysztof Hebel, "RTP: Real-time Transport Protocol", Multimedia Communications Laboratory Electrical & Computer Engineering Department, University of Waterloo