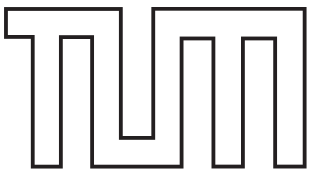


Technische Universität München
Chair of Media Technology
Prof. Dr.-Ing. Eckehard Steinbach

# Master Thesis

## Face Tracking for Optimized Bitrate Control in Low Delay Video Encoding


| | |
|---|---|
| Author: | Chethan Ningaraju |
| Matriculation Number: | 0365491 |
| Address: | SchrofelhofStraße 10-05-07 |
| | 81375 Munich |
| Advisor: | Mr. Muhammad Zafar Iqbal and Dr. Eugen Wige |
| Begin: | 15/07/2016 |
| End: | 15/01/2017 |

With my signature below, I assert that the work in this thesis has been composed by myself independently and no source materials or aids other than those mentioned in the thesis have been used.

München, December 29, 2016 _____

Place, Date Signature



München, December 29, 2016 _____

Place, Date Signature

# Abstract

Titel auf Englisch wiederholen.

Es folgt die englische Version der Kurzfassung.

# Contents

# Chapter 1

# Introduction

In recent years, there is increasing demand for high-quality video conferencing solutions. Due to availability of high-speed internet, video conferencing has proved to be an efficient alternative to face-to-face meetings. Video telephony has grown into a multi-billion dollar industry and has huge commercial significance. To address this growing need there has been constant improvement in low-delay video coding techniques in addition to better techniques to ensure low-delay transmission reliability at the network level. The tremendous increase in smartphone usage has led to an increase in video telephony over cellular networks whose bandwidth is highly constrained. Therefore, it is very important to develop methods of delivering high quality video with less bandwidth requirement.

The most commonly used video coding standards like H.264/AVC have been designed to exploit the spatial and temporal redundancies in the input video stream to achieve high data compression. The techniques of spatial and temporal prediction form the core principle of these video coding standards [TWL03]. However, after encoding the video the perceptual redundancies still remain since human attention does not focus on the whole scene but only a small region of fixation called region-of-interest (ROI) [MX14]. Therefore, reducing the perceptual redundancy gives a new dimension towards achieving lower bit-rate at acceptable perceptual quality. This work proposes a region-of-interest based bitrate control scheme for low-delay video encoding to exploit the perceptual redundancies.

In this work, the salient region of the frame which is the face of the participant in a video conference is identified. Since the attention of the viewer is mostly focused on the face of the other participants during a video conference call, improving the quality of the face region (ROI) can improve the overall perceptual quality. In this work, ideal capture conditions are assumed and results of the face tracking is used directly as supplementary information for the H264/AVC encoder's bitrate control. This work explores the methods of region of interest(ROI) based encoding to exploit the available bandwidth to encode regions that are of high importance to perception with higher quality. Face region in the input stream is allocated an above-average bit-count to yield a better visual quality than

the background regions. It is the aim of this work to develop and extensively evaluate the strategy of uneven bit-allocation and also to identify its limitations.

The remainder of this thesis is organized as follows. Chapter 2 gives an overview of the hybrid video coding used in H.264, functionality of the bitrate control module and the concept of Region-of-Interest(ROI) based encoding. In Chapter 3, a literature review is presented which discusses related earlier research works in the field of ROI-based encoding. An insight into limitations of the earlier works is also presented in this chapter. A detailed overview of the low-delay bitrate control module [SML02] used in this work is presented in chapter 4. Chapter 5 deals with explanation of the setup used in this work along with assessment techniques to evaluate ROI-based encoding approaches presented in this thesis work. The proposed bitrate control for ROI-based encoding is presented in Chapter 7. The conclusion for this thesis work along with a note on potential future work related to this topic is presented in chapter 8.

# Chapter 2

# Background

A brief overview of principles of hybrid video coding used in H.264 is presented in this chapter. The ROI-based encoding approaches discussed in this work are implemented in the form of intelligent bitrate control schemes. Therefore, an overview of the bitrate control module is presented to provide the reader with an understanding of the functionality of bitrate control in a video encoder. Finally, the concept of Region-of-Interest based encoding is introduced.

## 2.1 Hybrid Video Coding

H.264/AVC is one of the most commonly used video coding standard. Figure 2.1 depicts the underlying principle of block-based hybrid video coding used in H.264 [TWL03]. The encoding scheme aims to exploit the spatial and temporal redundancies that exist in a video.

In this coding scheme, the input picture is represented in block-shaped units(16x16 pixels) of associated luma and chroma samples called macroblocks (MBs). The basic source-coding algorithm is a hybrid of inter-picture prediction to exploit temporal statistical redundancies and transform coding of the prediction residual to exploit spatial statistical dependencies. The two types of prediction used are:

**Intra Prediction:** There exists a high similarity among the neighboring blocks in a video frame. In intra prediction, a block is predicted from its neighboring pixels of already coded and reconstructed blocks. H.264 offers nine intra prediction modes (one DC prediction mode and eight directional prediction modes) [TWL03].

**Inter Prediction:** When the frame rate is sufficiently high, there is a great amount of similarity between neighboring frames. It is highly efficient to code the difference between such similar frames than the frames themselves. In inter prediction, block-

based motion estimation is used to predict the motion of macroblocks relative to the previous encoded frames.

The first frame in a video is encoded using intra prediction and transform coding. The transform coefficients are quantized to achieve high compression ratio. These frames which are encoded without any dependencies on the neighboring frames are called key/intra frames. These frames act as reference frames to encode subsequent frames.

Once a reference frame is available, inter prediction can be used to remove the temporal redundancies. During motion estimation, it is usually not possible to find an exact match for the current macroblock. Therefore, the residual error is estimated for the prediction from motion estimation. This is called motion compensation. The residual error is coded using transform coding followed by quantization to achieve higher compression ratio. Due to quantization, it is not possible to recover the exact transform coefficients at decoder end without any loss of information. This introduces distortion in the decoded frame.

These frames which are predicted from other reference frames are called inter frames. The inter frames which use only past frames as reference are called P-frames. In addition to past frames, future frames can also be used as reference. Such frames are called Bi-directional frames or B-frames. Both B-frames and P-frames can be used as reference frames for encoding subsequent frames.

The quantization step used to quantize transform coefficients is specified using quantization parameter (QP). In H.264, QP range of 1 to 51 is allowed which is translated to quantization steps. The magnitude of distortion introduced by quantization depends on the quantization parameter. A lower QP implies low quantization step resulting in high quality output with less compression. The compression ratio increases with increase in QP at the cost of decreased output quality. The output data rate of the encoder is controlled by computing suitable QP, this is the task of bitrate control module, this is described in the following section. A detailed overview of individual steps involved in H.264 coding can be found in [TWL03].

## 2.2 Bitrate Control

The bitrate control module is responsible for controlling the bit-consumption of the encoder to guarantee smooth playback. Bitrate control is not specific to a video coding standard and hence operates independent of any chosen video coding standard. There are various flavors of bitrate control like Constant Bitrate (CBR), Variable Bitrate (VBR) and Average Bitrate (ABR). In this work, CBR type of bitrate control is considered since it is most commonly used in video conferencing and other real-time streaming applications.

Figure 2.2 illustrates the functionality of the bitrate control module. The main purpose of bitrate control module is to ensure smooth playback of the encoded video under given

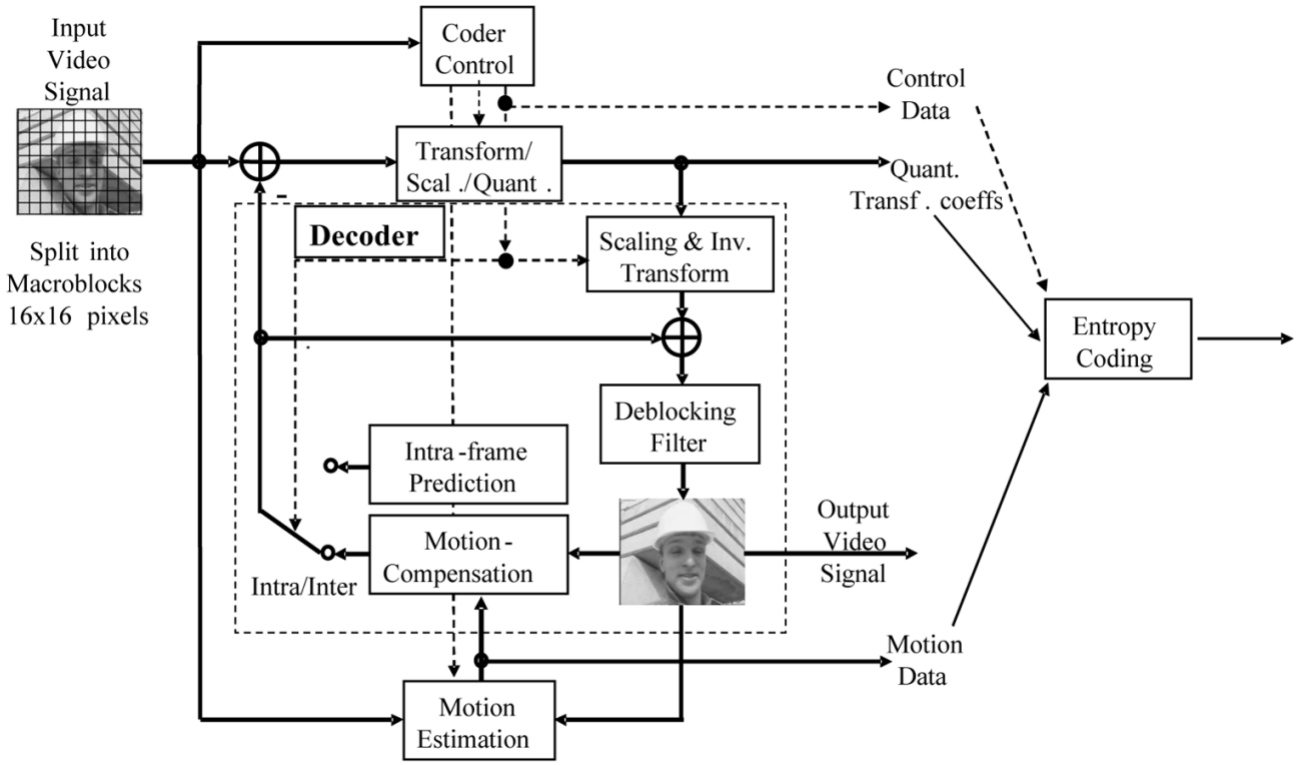



Figure 2.1: Block-based hybrid video coding

bandwidth and delay constraints. It estimates the video bitrate based on the available network bandwidth, ensures that the coded bitstream can be transmitted within the specified delay and makes full use of the limited bandwidth [ZWW11]. It achieves this by controlling the quantization parameter (QP) used during the encoding. The quantization parameter is computed considering the input bitrate, framerate, input complexity (spatial and temporal activity) and acceptable delay of the system. The module also receives regular feedback from the encoder to make better QP adaptation. The feedback from the encoder gives information about the complexity of the input video and helps the bitrate control to compute optimum QP for a given bit-budget.

The functionality of the bitrate control module can be illustrated with the help of the leaky-bucket model [ZWW11]. The output data rate of a video encoder varies depending on the input complexity of the video (motion in the frame). It also depends on the picture type of the encoded frame. The key/I-frames consume a lot of bits compared to inter pictures (P-frames and B-frames). In a video streaming scenario considering a constant bitrate channel, the throughput is maximum when data-rate is constant and equal to the available bandwidth. Therefore, the output data of the encoder is smoothened using a theoretical buffer called Video Buffer Verifier (VBV). The VBV is a virtual buffer modeled by the bitrate control module to ensure that the video stream can be correctly buffered and played back at the decoder end. This is equivalent to the leaky-bucket model as shown

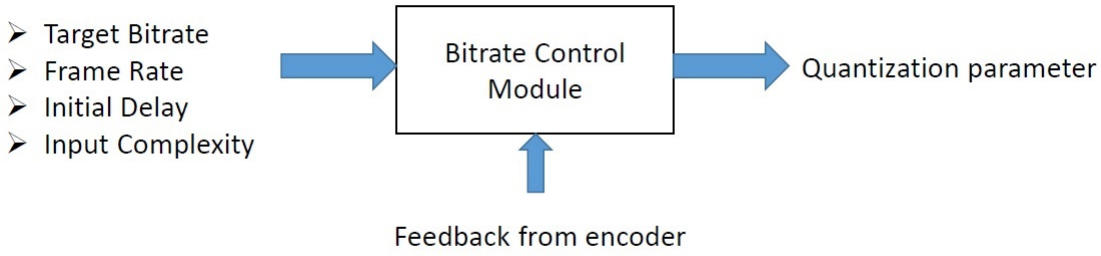



Figure 2.2: Bitrate Control Module Functionality

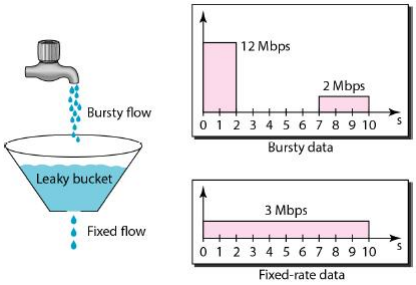



Figure 2.3: Leaky Bucket Model

in Figure 2.3 [Lea], where the output of the encoder with a variable rate (bursty output) is stored in a buffer (leaky bucket) which is draining at a constant rate. Any underflow or overflow of this buffer causes glitch in the video streaming. To ensure that there is no VBV overflow or underflow, encoder's quantization parameter is adapted on a macroblock level so that the maximum allowed bit-count for an encoded frame is not exceeded. A detailed description of the low-delay bitrate control module used in this work is given in chapter 4.

## 2.3 ROI-based Coding

In conventional video coding, all regions of a frame are considered equally important to the viewer. It is assumed that all regions contribute equally to the perceptual quality. However, the study of Human Visual System (HVS) shows that human eyes can only focus on one area in a frame at any given point in time which is called region of interest. For example, it has been found out in [Wan95] that humans normally perceive clearly a small region of 2°-5°of the visual angle. This corresponds to a very small region in a video frame.

In video coding, the compression gain from spatial and temporal prediction is reaching saturation level. Further compression from these techniques demand exponential growth in

computational capabilities. The next generation video coding standards like HEVC/H.265 claims to offer two fold increase in compression ratio over H.264. This increase in compression comes with multi-fold increase in computation complexity during encoding. Therefore, perceptual coding can provide an alternative solution towards lower bitrate video coding. The technique of encoding regions with higher importance to perceptual quality at a higher quality at the cost of degradation of quality in non-ROI parts is called ROI-based coding. The loss of quality in non-ROI is not perceived by the viewer leading to increased perceptual quality at the same bitrate.

The ROI-based coding is not a common practice in video coding because it is very hard to automatically detect important regions in generic contents that contribute the most to the perceptual quality. There are many ways of detecting region of interest, most of which are application specific. The most common approach is usage of difference image based moving object detector. In these systems any moving object is considered as ROI. A typical use-case for such a system is video surveillance. In addition to difference image based motion detection, global motion estimation is used in ROI detection [HM11] in applications like aerial surveillance.

In generic video content, region of interest changes constantly depending on the context. For instance, in a movie, the ROI can depend on context of the scene. Developing generic techniques for detection of ROI in such videos is very difficult. There have been attempts to use eye tracker to record the foveation points of a human observer on the receiver which was used to apply foveation filter in video coding of the sender. An advancement over such approach is proposed in [SL03] which optimizes rate control to maximize foveal visual quality metric. These generic ROI detectors are very hard to implement due to uncommon availability of eye tracking mechanism at the receiver.

The region of interest in the video conferencing scenario is going to be the face region predominantly. Due to recent improvements in face detection algorithms it is possible to detect the face with good accuracy. The study in [MB13] shows that boosting quality of the face regions can improve the overall perceived quality of the video. This work aims to study possible ways of improving perceptual quality of the video by detecting face region and coding it with higher quality than rest of the frame.

In conventional video coding, the bitrate control allocates bits at every macroblock and adjusts the QP accordingly. In a simple approach, every macroblock is considered equally important to perceptual quality hence the available bits are distributed evenly across all macroblocks within a frame. Since low-latency and load efficiency are of high importance it is not advisable to do multi-pass encoding for optimal bitrate allocation. Therefore, over-allocation in one macroblock has to be compensated by under-allocation (using a higher QP) for neighboring macroblocks, regardless of the image content. However, a more intelligent allocation strategy should take the image content into account. In this work, parts of the image with higher importance (ROI) is given a higher share of the overall bit-count resulting in higher visual quality. The additional bits spent in coding ROI is compensated by allocating a lower proportion of the bit-count to the background regions

(non-ROI).

The knowledge of region of interest in input video can be used for many other purposes in addition to its use in video coding to improve perceptual quality. For instance, ROI information can be used in developing techniques for smart thumbnail displays in group video conferencing solutions. In a group video conference, an active person is detected based on the source of the voice and is displayed on the main screen and other participants are displayed on the smaller windows with down-scaling of the entire video. If the face coordinates of the participant is transmitted along with the bit-stream as meta-data, a cropped version of video can be displayed to show only the face in thumbnail display. This can improve the overall user-experience.

# Chapter 3

# Related Work

The concept of region of interest based encoding has been around for a while. In this chapter, some of the papers which are relevant to this thesis work are reviewed. The different approaches of using ROI information to improve perceptual quality can be broadly classified into following categories based on which stage of processing the ROI information is used.

- Preprocessing - Example: blurring.
- Video Coding - Example: during the RDO, bitrate control.

The relevant techniques proposed under these categories are discussed in the following sections.

## 3.1 Preprocessing

The input video stream can be directly altered based on the ROI information. Many pre-processing approaches are also used to directly reduce unimportant information by applying a non-uniform distortion filter in a scene. For instance, the image is divided into foreground (ROI) and background (non-ROI) and the non-ROI parts are blurred to save bits during encoding [AC05]. The work in [HM16] is an example for codec independent ROI encoding. It can work with any codec and arbitrary ROI detector. In this approach, input video stream is modified only to contain relevant information. The non-ROI pixels are replaced such that these regions can be very efficiently compressed by the encoder. The non-ROI macroblocks are either replaced by corresponding blocks from the previous frame or black blocks. The non-ROI regions are reconstructed with post-processing assuming a zero motion vector. This approach is used in scenarios where non-ROI regions are mostly discarded completely at the receiver end.

## 3.2 Video Coding

In video encoding, multiple approaches exist to preferentially code ROI macroblocks with a higher quality at the cost of degrading non-ROI parts. The ROI information can be directly used in rate-distortion optimization (RDO) during video encoding to alter the quality of ROI. One such approach is presented in [FL16] in which non-ROI macroblocks are encoded using only AMC-based modes (Active MB Concealment). The non-ROI macroblocks are encoded with only motion vector but no residual information. This creates a bias to choose larger distortion for non-ROI blocks to save bits during Lagrangian optimization. The bits saved during encoding of non-ROI macroblocks is used to encode ROI blocks with a higher quality

The resources available during encoding like computational power is preferentially allocated to ROI in [YLS08b]. The H.264 coding standard offers many methods to enhance its compression performance such as variable block size ME, quarter-sample-accurate ME, and multiple reference frames ME [TWL03]. The complexity of H.264 encoder is significantly increased due to employment of these new methods. The computation need for non-ROI is reduced significantly at the cost of compression efficiency by adaptively adjusting the coding parameters as follows,

- Choose only a subset of macroblock partition and prediction modes offered by H.264 standard for non-ROI during rate-distortion optimization to find the best mode. This is similar to usage of only AMC-based prediction modes for non-ROI proposed in [FL16].
- The number of reference frames is reduced to use only immediate neighboring frames for non-ROI macroblocks. The ROI MBs are allowed to reference from multiple frames.
- The accuracy of subpixel accurate Motion estimation is set to quarter-pixel for ROI and half-pixel for non-ROI MBs.
- The motion estimation search range is reduced significantly for non-ROI.

These modifications in coding parameters result in a higher quality for ROI even when the available bit-budget is uniformly distributed across ROI and non-ROI parts.

A more common way of ROI-based encoding is by altering the rate control module to allocate above average bits to ROI macroblocks. In addition to preferential allocation of computation power to ROI, ROI-based rate control is proposed in [YLS08b]. The modifications to the rate control scheme for low-delay video communication of H.264 [YLS08a] has been proposed to allocate more bits (smaller QP) to the ROI macroblocks. This is done by assigning weights to every macroblock based on its importance to the human visual system. A linear R-Q model is proposed to optimize QP calculation to provide a ROI-based rate control at the MB level.

The work in [GLW12] proposes a bit-allocation and rate control scheme for enhancing regional perceptual quality using structural similarity(SSIM) index as the quality metric for distortion-quantization modeling. Statistical analysis is adopted to obtain the relation between SSIM of reconstructed MBs and corresponding QP (from 20 to 51 in this paper) after standard video coding. A target SSIM is set for the ROI and corresponding QP is determined with the help of the SSIM-QP model. The proposed algorithm has following steps.

- Target bits for each basic unit is estimated and allocated based on the bit-budget of ROI or non-ROI parts.
- A preliminary QP for each BU is computed based on R-Q model.
- The predicted QP is used for rate-distortion optimization mode decision of the encoding flow to obtain the best prediction macroblock.
- The QP for MBs in ROI used for final encoding is altered to achieve the target SSIM quality based on the SSIM-Q model

A ROI-based encoding scheme specific to video conferencing is proposed in [LT05] for H.263. This work proposes an algorithm to track the face using motion-vector information. Once the ROI is detected, it proposes modification of bit-allocation for both CBR and VBR mode of rate control. The QP for these blocks is predicted from the rate modeling of ROI and non-ROI blocks. The work described in [MX14] goes a step further in face detection based ROI encoding schemes by enhancing finer facial feature to improve the perceptual quality in high-resolution HEVC encoding. In this approach, different weights are assigned to the background, face, eyes, mouth and nose regions which are in-turn used to alter quality by ROI-based adaptive CTU (Coding Tree Unit) partition structure for HEVC.

## 3.3 Inference

The preprocessing based ROI encoding approaches described in section 3.1 [AC05, HM16] offers codec independent ways to enchance quality of ROI. However, these approaches are not suitable for video conferencing since they can cause a very high degree of degradation in the background regions (non-ROI) due to preprocessing. An excessively degraded background can also reduce the perceptual quality. It also needs an additional stage of preprocessing which adds to the complexity of the system making it hard to process in real-time.

The proposed bitrate control schemes for ROI-encoding [LT05, YLS08a] assign arbitrarily large weights to ROI macroblocks compared to non-ROI macroblocks during bit-allocation. These weights do not take into account the characteristics of the content of the input video stream. These works propose effective methods to create a quality difference between ROI and non-ROI. They do not throw sufficient light on determining the optimal quality

difference to achieve best perceptual quality. The approach of targeting a SSIM value for the ROI macroblock [GLW12] offers a simple way to guarantee minimum quality for the ROI macroblocks. However, it is difficult to come up with a target SSIM value for different contents and bitrates that yields best perceptual quality.

Most of the previous work concerned with modification of rate control to achieve better quality in ROI macroblocks deals with altering bit-allocation module. Some of the commonly used bitrate control schemes [SML02] do not perform bit-allocation at the macroblock level. The modifications to the rate control proposed in these works cannot be adopted in such bitrate control module.

In this thesis work, some of the shortcomings of the previous works are addressed. This work proposes ways of achieving ROI encoding for bitrate control modules that do not perform an explicit bit-allocation at the macroblock level. This work proposes multiple approaches for ROI-based bitrate control which vary in terms of ease of implementation. The characteristics of content in both ROI and non-ROI parts is taken into account to vary the weights used during bit-allocation. This results in a superior ROI-based bitrate control which yields better perceptual quality across wide rage of input video streams.

# Chapter 4

# Low-delay Bitrate Control

This section gives an overview of the low-delay bitrate control module used in this work. The need for extremely low end-to-end delay in video telephony puts the following additional constraints on video coding which results in compromise of video quality.

**No B-Frames:** During the low-delay video encoding, tools like bi-directional prediction (B-frames) are disabled. The usage of B-frames needs buffering of at least one frame. This adds on to the overall latency of the system which is highly undesirable in video conferencing.

**Reduced buffer size:** The tolerable delay in video encoding is a direct measure of the Video Buffer Verifier(VBV) size. When the size of the VBV is very low (due to low delay), there is less room to accommodate the variation in bitrate of the encoder. This implies that there can be minimum variation in the frame size across different frames irrespective of the content.

**Increased dropped frames:** Any wrong prediction of QP by the bitrate control module can have bad impact since there is no additional time available to re-encode the content with a corrected QP. For instance, in case of over-consumption of bits by a frame which can lead to dropped frames cannot be corrected by re-encoding the frame which a higher QP.

The bitrate control module used in this work is a modified version of [SML02]. The bitrate control does a frame level bit-allocation based on fullness of the VBV, followed by adapting QP at the macroblock level based on the structural complexity of the macroblock. The functionality of the bitrate control module can be divided into two stages:

- Bit Allocation
- QP Prediction

These two stages are described in detail in the following sections.

## 4.1 Bit Allocation

The low-delay encoding mode does not favor usage of key frames at regular intervals and B-frames. Therefore, in steady state only P-frames are used in encoding video conferencing content. This makes frame level bit-allocation simpler since there is no need to consider relative complexity between different types of frames during frame level bit-allocation. The key-frame at the beginning is handled using special cases.

As depicted in Figure 2.2, one of the inputs for the bitrate control module is a delay/latency parameter (L). This is defined as the maximum permissible delay allowed between the encoder and the decoder assuming zero transmission delay. In other words, the delay parameter is the maximum allowed time for any encoded frame to be transmitted completely through a constant bandwidth channel of per-defined bitrate. In this work, delay parameter (L) is configured as,

$$L_0 = 165ms \text{ and } L = \frac{1.5 * 1000}{framerate}. \tag{4.1}$$

Initially a delay of $L_0 = 165ms$ is allowed for the key-frame. This allows, allocation of higher than average bits for the key-frame. However, the delay parameter (L) for the P frames in steady state is only 1.5 times the frame sampling delay. For instance, if the input video is sampled at 30 frames/sec, then the time interval between two consecutive frames is 33ms (frame sampling delay), the permissible delay(L) for frames in steady state is approximately 49ms. The usage of different delay values for the first key-frame and steady state P-frames is handled by changing the delay value gradually. The large key-frame at the beginning results in huge delay (165ms), this delay is gradually reduced by using less than average bit-count for subsequent few frames (half of average bit-count per frame). Once the over-consumption of the first-key frame is compensated, the steady state delay of $49ms$ is maintained for the rest of the sequence.

It should be noted that the initial delay in the system can only be reduced by displaying the initial few P-frames at shorter intervals than the time interval in which they were captured. This results in momentary increase in playback speed. In practical implementations, usage of above an average bit-count for the I-frame results in a few dropped frames subsequently even if the I-frame over-consumes marginally. All the above artifacts are considered as an acceptable trade-off to achieve good initial spatial quality by allocating a huge amount of bits to the first key-frame.

The bit-allocation module uses VBV fullness and delay parameter (L) to compute the bits allocated for the current frame to be encoded. The VBV fullness ($d_0^n$) before encoding the $nth$ frame is calculated based on the size of the previously encoded $(n-1)th$ frame in bits ($FrameSize_{n-1}$) as follows,

$$d_0^n = d_0^{n-1} + (FrameSize_{n-1} - AvgBitsPerFrame), \tag{4.2}$$

$$d_0^n = max(d_0^n, 0),$$

where

$$AvgBitsPerFrame = \frac{bitrate}{framerate}.$$

The allocated bits for the $nth$ frame is the maximum amount of bits that can be transferred along with the residual bits in the VBV in the duration L ($49ms$ in the above example). The maximum acceptable delay in ms (L) is translated to bits($L_{bits}$) using the following equation,

$$L_{bits} = \frac{L * birate}{1000}.$$

Therefore, the amount of allocated bits for the current frame ($B_{alloc}$) is given by,

$$B_{alloc} = L_{bits} - d_0^n. \tag{4.3}$$

In practice, rate control QP predictions are not very accurate to exactly consume the bits that were allocated to the frame ($B_{alloc}$). If a frame consumes more bits than $B_{alloc}$, it violates the delay conditions. The encoded frame will be unable to reach the decoder in time with the available bitrate. Hence, the frame is not added to the bitstream. These frames which are encoded but not part of the output of the encoder are called *dropped frames*. Such dropped frames must be avoided since they cause jerky playback. A small room for inaccuracy of the QP prediction is considered at the end of the bit-allocation stage to avoid dropped frames. In practice, the amount of target bits used for QP prediction is slightly smaller than $B_{alloc}$ to avoid dropped frames in case of marginal over-consumption of bits.

## 4.2 QP Prediction

Due to low VBV size, the bitrate control needs to have very quick reaction to any deviation in the bitrate to avoid dropped frames. The bitrate control algorithm computes the QP for every macroblock. The two factors considered while computing QP for a macroblock are:

- Macroblock Complexity
- Bitrate Deviation

The macroblock complexity is used for adapting the QP according to the structural complexity of a macroblock. The deviation in bitrate at the macroblock level is computed based on the feedback from the encoder to achieve target bitrate with higher precision. Each of these factors are discussed in the following sections.

### 4.2.1 Macroblock Complexity

The structural complexity of the macroblock is used to compute the delta QP (dq) which is used to adapt QP at the macroblock level. The activity of the macroblock is a measure of complexity of the macroblock and hence indicates the amount of bits required to encode the macroblock. After motion compensation with the selected coding mode and motion vectors, the activity of the $mth$ macroblock ($act_m$) with original pixel value $s_m(i,j)$ and predicted pixel value $c_m(i,j)$ is calculated using (4.4).

$$act_m = \sum_{i,j} \mid s_m(i,j) - c_m(i,j) \mid, \quad i,j = 1,2,....16. \tag{4.4}$$

The relative complexity of the macroblock with respect to the entire frame complexity is used in QP adaptation. The ratio of activity of the current macroblock and the average activity of the entire frame is used to calculate the delta QP (4.5).

$$dq = \begin{cases} -floor(\frac{avg\_act}{act_m} - 1), & 0 < \frac{act_m}{avg\_act} <= 1/2. \\ 0, & 1/2 < \frac{act_m}{avg\_act} <= 2. \\ floor(\frac{act_m}{avg\_act}) - 1, & \frac{act_m}{avg\_act} >= 2 \end{cases} \tag{4.5}$$

Where, $avg\_act$ is the average activity across all the macroblocks of a frame. As depicted in the above equation, a positive dq is used when the current macroblock is relatively complex compared to average frame complexity. This indicates that for a relatively complex macroblocks within a frame, a higher QP is used. Such activity based QP adaptation results in uneven quality within a frame. The peak signal to noise ratio (PSNR) of the simple or static region is higher than that of regions with high motion (foreground). In practice, the average frame activity of the entire frame is unavailable until the last macroblock of the frame has been encoded. Therefore, previous frame average activity is used as current frame activity since the two adjacent frames in a video are likely to remain similar.

The activity metric used in (4.5) is a measure of complexity of the macroblock, hence it can be replaced by similar metrics depicting the complexity of the block. Other metrics like SATD (Sum of Absolute Difference in Transform Domain) and cost of the macroblock (J) can be used instead of the activity. In this work, cost of the macroblock computed during rate-distortion optimization (4.6) is used as the complexity metric.

$$J = D + \lambda R, \tag{4.6}$$

Where, the distortion D represents the residual error after prediction which is measured as the sum of absolute differences (SAD) between the original block and the reconstructed block, is weighed against the number of bits R associated with the motion information using the Lagrange multiplier $\lambda$. The least cost of all the evaluated modes is considered as the complexity of the block. The cost of the macroblock factors in both the amount of residual information to be encoded after motion compensation and bits used for signaling the mode and the motion vector. This makes it more accurate in terms of reflecting the complexity of the block compared to the activity computed in (4.5).

### 4.2.2 Bitrate Deviation

The delta QP computed in (4.5) is added to the QP calculated based on the deviation in the bitrate reflected by instantaneous VBV fullness. The buffer fullness corresponds to fullness of the VBV discussed in the context of leaky bucket model in section 2.2. Any deviation in the bitrate will be reflected in the occupancy of the buffer. For example, a higher level of the buffer indicates over-consumption of bits. The VBV buffer occupancy ($d_0^n$) is calculated only after an entire frame is encoded. In order to account for the deviation in bitrate at the macroblock level, a global deviation factor is computed at the macroblock level. The global deviation is computed based on the deviation in frame level bit-consumption of past frames and the size of the macroblocks encoded in the current frame. The global deviation factor ($D_m^{n'}$) when encoding the $mth$ macroblock of $nth$ frame is calculated using (4.7).

$$D_m^{n'} = d_0^{n'} + CurFrameBitCount - \frac{B_{alloc} * m}{M}. \tag{4.7}$$

Where, M is the total number of macroblocks in a frame, $B_{alloc}$ is the bits allocated to the frame by bit-allocation module (4.3) and hence remains a constant for the given frame. CurFrameBitCount is the bit-consumption of the current frame until the last encoded macroblock. The term $d_0^{n'}$ in (4.7) is accumulated frame level bit-deviation which is computed similar to VBV fullness ($d_0^n$). Since the VBV fullness ($d_0^n$) is computed only after fully encoding the frame, the factor $d_0^{n'}$ also remains constant for a given frame. The two terms $d_0^{n'}$ and $d_0^n$ differ with initialization values at beginning of the encoding [SML02]. The frame level deviation factor ($d_0^{n'}$) is additionally subjected to clipping as shown in (4.9) after encoding every frame.

The global deviation factor ($D_m^{n'}$) accounts for the deviation in bitrate of the encoded video until the last encoded macroblock. The global deviation factor is used to calculate the QP for the $mth$ macroblock ($Q_m$),

$$Q_m = \frac{D_m^{n'} * 31}{r} + dq, \tag{4.8}$$

where,

$$r = i * bitrate/framerate.$$

The factor $r$, is called the reaction factor. This factor indicates the number of frames over which the deviation in bitrate is to be compensated. The bitrate control module in this work uses $i = 1$.

The working of bitrate control can be understood by analyzing the deviation factor $D_m^{n'}$ in (4.8). At the beginning of the frame (m = 0),

$$D_m^{n'} = d_0^{n'}.$$

The allocated QP($Q_m$) solely depends on $d_0^{n'}$ if the deviation in macroblock level bit-consumption and activity based delta QP(dq) are ignored. The init value of $d_0^{n'}$ is chosen

heuristically at the start of encoding based on the most commonly used configuration. Therefore, the QP calculated for the first frame is not content dependent. Once the first frame is encoded, the bit-consumption usually differs from allocated bits by a large extent. For instance, if the content was more complex than average, the initial QP allocation will result in large bit consumption, increasing the value of frame level bit deviation ($d_0^{n'}$) which results in large global deviation ($D_m^{n'}$). This will result in larger QP value for next frame to be encoded according to (4.8). Therefore, the value of $d_0^{n'}$ oscillates for the first few frames. In steady state it takes optimum value to keep the deviation low eventually helping in achieving the target bitrate. The frame level bit-deviation $d_0^{n'}$ is clipped between pre-computed maximum and minimum value to limit the QP to a suitable range.

$$d_0^{n'} = clip(1000, \frac{40 * r}{31}) \tag{4.9}$$

The QP output by the rate control ($Q_m$) is clipped between valid range of QP allowed in H.264 encoding. In addition to these limits, the QP computed in (4.8) is subjected to swing restrictions. Since the QP is modulated based on the activity of the macroblock, a upper limit of maximum QP ($QP_{max}$),

$$QP_{max} = QP_{avg} + 5, \tag{4.10}$$

is set to make sure the high activity regions are not excessively penalized with higher quantization. Here, $QP_{avg}$ corresponds to the average QP of all the blocks in the previous encoded frame.

# Chapter 5

# Study Setup

This section describes the setup used in this work. It describes the configuration of the encoder used to evaluate different algorithms. It also describes the metrics and other aspects used to evaluate the proposed ROI-based bitrate control and to compare its performance with the state of the art bitrate control algorithm.

## 5.1 Sample Input

This section describes the sample video sequences chosen for evaluating the algorithms used for ROI-based encoding. A typical video conferencing scenario is considered in this work. The list of inputs and their characteristics are tabulated in Table 5.1.

| Name | Resolution | Frame rate (Frames/sec) | Total Frames |
|---|---|---|---|
| Paul640x480 | 640x480 | 30 | 1000 |
| Chet640x480 | 640x480 | 30 | 490 |
| Johny1280x720 | 1280x720 | 30 | 300 |

Table 5.1: Sample input sequence

The three chosen sample video sequences have different spatial and temporal complexities. The sample input videos are chosen to cover most of the typical video conferencing environments. The snapshots of the chosen sample video sequence is shown in Figure 5.1. Table 5.2 lists the relative spatial and temporal complexities across different content. The table also lists the relative area of the face (ROI) within a frame throughout the given sequence. The relative ROI area ($A_{roi}$) is defined as,

$$A_{roi} = \frac{M_{roi}}{M},$$

where, $M_{roi}$ is the number of ROI macroblocks and M is the total number of macroblocks in a given frame. The area of face region in chet640x480 changes significantly within the sequence, other input sequence have almost constant $A_{roi}$ across all the frames.

| Name | Relative Complexity | | $\mathbf{A_{roi}}$ (approx) |
|---|---|---|---|
| | Spatial | Temporal | |
| Paul640x480 | Medium | Low | 0.05 |
| Chet640x480 | Medium | High | 0.04 - 0.4 |
| Johny1280x720 | Low | Low | 0.10 |

Table 5.2: The relative spatial and temporal complexity comparison for the sample input videos

## 5.2 Encoder Configuration

This work uses Citrix H.264 video codec for studying different ROI-based encoding approaches. The encoder is configured in low-delay mode suitable for video conferencing and other real-time applications. The bitrate control described in section 4 is used to control the output data-rate of the encoder. The encoder is configured to use IPPP mode with key/I-frame used only at the beginning of the sequence followed by only uni-directional P frames. Due to the low-delay requirements there is no provision to re-encode the frame in case of buffer overflow. The frames are dropped entirely in case of a buffer overflow to maintain a constant low-delay. The configuration for encoding each of the sample input

| Name | Bitrate (Kbps) | VBV Size (Bits) |
|---|---|---|
| Paul640x480 | 250 | 12500 |
| Chet640x480 | 250 | 12500 |
| Johny1280x720 | 350 | 17500 |

Table 5.3: Encoder configuration

sequence listed in the above section is shown in Table 5.3. The bitrates for different input sequence is configured considering the resolution and the complexity of the content. The chosen bitrates are relatively low yielding low visual quality with conventional encoding. This makes it easy to assess the gain in perceptual quality with ROI-based encoding. Figure 7.13 shows the results of conventional encoding for the configuration shown in Table 5.3. Due to usage of low bitrate, blocky artifacts can be noticed in the face regions.

(a) (b)

(c)

Figure 5.1: Snapshot of test sequence (a)chet640x480, (b)Paul640x480 ,(c)Johnny1280x720.

## 5.3 Evaluation Criteria

One of the crucial aspects in this study is the metric used to evaluate various approaches in order to choose the best approach. The goal of this study is to improve the quality of the ROI macroblocks at the cost of degrading the non-ROI macroblocks. Since the whole approach is to measure the gain in perceptual quality, using frame level PSNR alone as a

(a) (b)

(c)

Figure 5.2: Result of Conventional Encoding (a)chet640x480, (b)Paul640x480 ,(c)Johnny1280x720.

metric could be misleading.

In this study, the difference in average PSNR of frames and average ROI PSNR is used as one of the metrics to evaluate different approaches. The expectation is to see an improvement in ROI PSNR, with degradation in PSNR of the non-ROI regions. The shift in quality of ROI and non-ROI parts should be achieved keeping the bitrate unchanged. Therefore, the second aspect is to measure the deviation in bitrate behavior with ROI-

based encoding compared to conventional encoding without using any ROI information. Third aspect involves analysis of PSNR and QP variation within a frame. The QP and PSNR distributions within a frame are studied to ensure that there is visible improvement in the quality of ROI without badly degrading the quality of the non-ROI.

To measure all these behaviors the following metrics are considered. These are discussed in detail in the following sections.

- Quality metrics - PSNR of ROI and non-ROI regions.
- PSNR and QP maps - The map of variation of QP and PSNR within a frame.
- Delay plot - Plots time delay for every frame.

### 5.3.1 Quality Metrics

The increased bit-allocation to ROI macroblocks in ROI-based encoding results in increase in PSNR of the ROI part. This also results in a drop in non-ROI PSNR. Finding the desirable magnitude of improvement in PSNR of ROI along with acceptable drop in PSNR of the non-ROI is tricky. The idea here is to find the right balance between quality improvement in the ROI and the degradation of non-ROI as to achieve maximum perceptual quality. The PSNR computed specific to regions within a frame gives insight into magnitude of transfer of bits from ROI to non-ROI during ROI-based encoding. This measure will also indicate the aggressiveness of an algorithm which is measured in terms of magnitude of objective quality difference forced between ROI and non-ROI parts.

The values in Table 5.4 show the PSNR values for the conventional video coding for the sample inputs listed in section 5.1. It is clear that for most of the sample inputs, the PSNR of ROI is much lower compared to that of the overall frame PSNR. This is not desirable since the regions that matter the most to the perceptual quality have a lower PSNR on an average.

| Content | PSNR Avg (dB) | PSNR ROI (dB) |
|---|---|---|
| Paul640x480, 250kbps | 39.15 | 37.72 |
| Chet640x480, 250kbps | 30.80 | 32.59 |
| Johny1280x720 350kbps | 37.93 | 36.08 |

Table 5.4: PSNR values for conventional encoding

The PSNR is calculated using the weighted sum of PSNR of individual components per picture ($PSNR_Y$, $PSNR_U$ and $PSNR_V$) [JOGJSW12].

$$PSNR_{YUV} = (6 \times PSNR_Y + PSNR_U + PSNR_V)/8, \tag{5.1}$$

where, individual components are computed as,

$$PSNR = 10 \times log_{10}((2^B - 1)^2/MSE), \tag{5.2}$$

where B = 8 is the number of bits per sample (bit-depth) of the video and MSE is the mean squared error.

The change in PSNR of the ROI and non-ROI parts is measured as an average of PSNR of the entire frame and average of PSNR of ROI of all the frames in the sequence. The PSNR for the entire video sequence can be computed in two ways.

- Average of frame PSNR - This is the average of PSNR of all the frames in the sequence.
- Average MSE-based PSNR - The PSNR of average MSE of all the frames in the sequence. This is computed by accumulating MSE over entire sequence and then computing PSNR.

The average of PSNR metric is preferred over average MSE based PSNR since the latter metric was found to be heavily influenced by the outliers. For instance, when the sequence has very few extremely low-quality frames but has good quality on an average, the average MSE based PSNR was found to be very low due to presence of very few extremely low-quality frames.

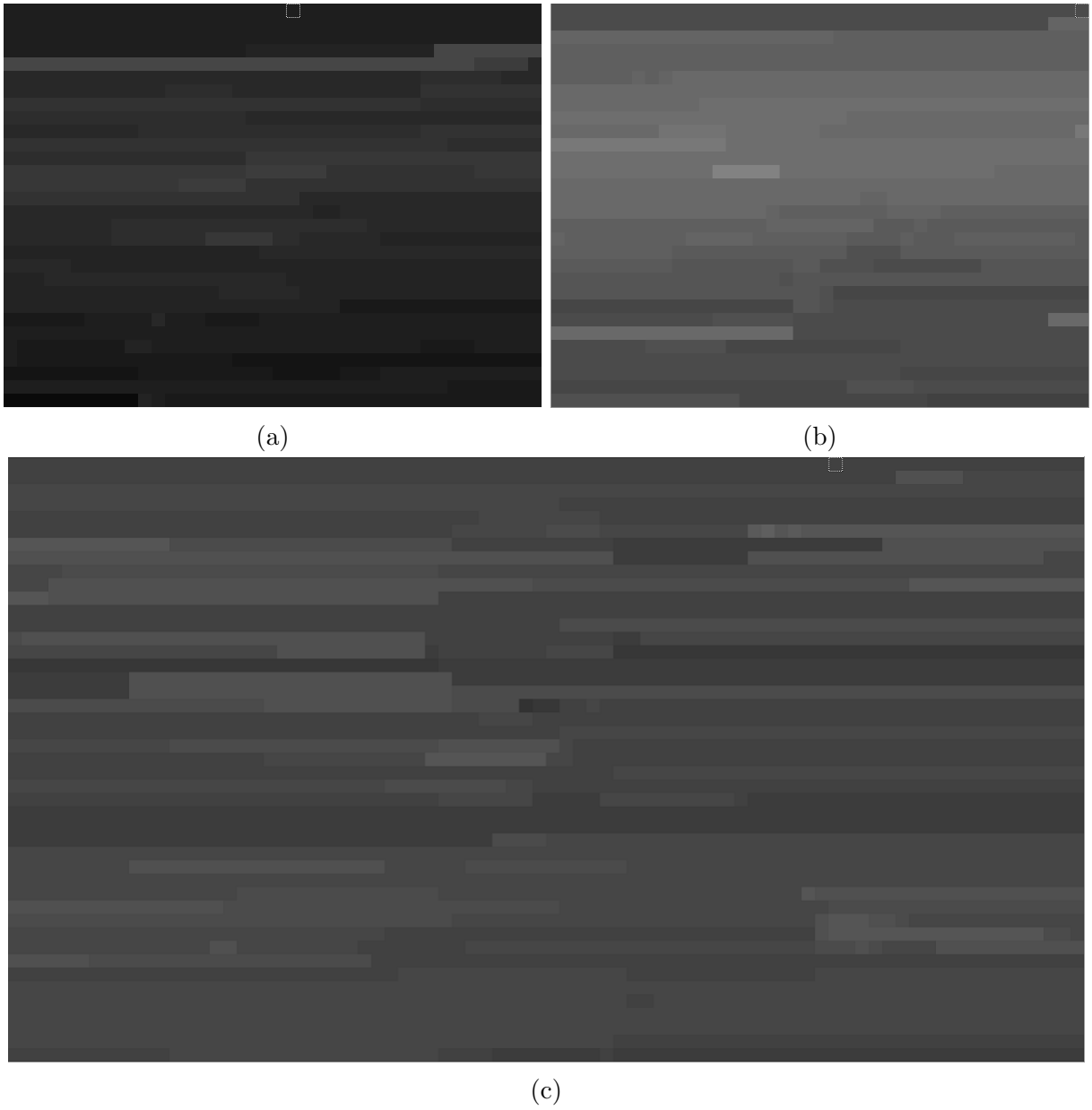
### 5.3.2 PSNR and QP Maps

The study of PSNR and QP distribution within a frame is important to understand the effect of movement of bits from ROI to non-ROI parts. The PSNR and QP is extracted at the macroblock level. It is then stored in raster scan order which can be used to display as an image to compare the structure with that of the video frame. These values are illustrated in a gray scale image.

In a QP scale map, the darker regions in a frame indicate higher quantization. Even though the quantization parameter used for encoding a block is closely related to the PSNR of the block, it is not the only determinative factor. The PSNR can also vary depending on the content. Generally, the lower frequency regions have better PSNR even when encoded with a higher QP. The static regions of the frame also tend to have better PSNR even when higher QP is used because of less new information to be encoded. The PSNR map helps in visualizing the effect of movement of bits from non-ROI to ROI parts.

The images in Figure 5.3 shows the quantization parameters used during conventional encoding for the frames in Figure 7.13. A darker shade of gray in this map indicate usage of higher quantization parameter compared to the regions with brighter shade of gray. The QP range of 1-51 is mapped to gray scale value between 0-255 in the quant maps. It can be noticed that since no information about region of interest is used while encoding the frame, the pattern of quantization appears almost random. The shape of the original content is not recognizable from the quantization map.

The images in Figure 5.4 shows the PSNR distribution for the frames in Figure 7.13 which



(a) (b)

(c)

Figure 5.3: Quantization maps for conventional encoding (a)chet640x480, (b)Paul640x480 ,(c)Johnny1280x720.

are encoded using conventional encoding. Similar to the quantization map, a brighter shade of gray represents the regions with higher PSNR, the darker regions indicate lower PSNR and worse quality. A PSNR range between 25dB-45dB is linearly mapped to gray scale value of 0-255 in the PSNR maps.

It is evident that the structure of the original content is preserved in the PSNR map. In most of the contents (Johnny1280x720 and Paul640x480), the background regions have

Figure 5.4: PSNR maps for conventional encoding (a)chet640x480, (b)Paul640x480 ,(c)Johnny1280x720.

better PSNR, the foreground has worse quality and the difference in quality is quite huge. The difference in the quality is due to the fact that background in a video conferencing scenario is mostly static and hence gets encoded better with every frame. On the other hand, the foreground has motion and new data to be encoded, hence it cannot achieve the same quality as the background. Since the focus of attention during video telephony is foreground or the face region, improving the face region must help in improving overall

perceptual quality. The PSNR maps help to visualize the effect of such preferential encoding. The goal of ROI-based encoding is to reduce the PSNR difference between foreground and background and to boost the quality of the foreground(face regions) to same level as the background or even better.

### 5.3.3 Delay Plot

The fundamental idea in this work is to efficiently use the bits within a frame to encode the region of interest with better quality. The algorithms used to achieve this should not alter the behavior of the encoder in terms of frame level bit-consumption. As mentioned in previous sections, in case of VBV overflow, the encoder drops the frame in order to maintain strict VBV compliance. The dropped frames result in jerky playback and hence should be avoided. The ROI-based bitrate control scheme should not contribute to an increase in the number of dropped frames due to changes in the bit-consumption pattern.

In contant bitrate control, the size of an encoded frame decides the buffering delay of each frame. Therefore, any change in bit-consumption pattern at the frame level gets reflected in the buffering delay of corresponding frames. The delay due to buffering of each frame in the sequence is plotted to analyze the bit-consumption behavior. Figure 5.5 is the delay plot of the sample video sequences encoded with the configuration shown in Table 5.3. Every point in the delay plot specifies the time taken by the corresponding frame (marked on the x-axis) to reach the decoder assuming zero transmission delay. It can be noticed the delay is almost constant for Johnny1280x720 and Paul640x480. However, there is slight variation for chet640x480 due to high temporal complexity.

The curve in delay plots appear mostly smooth except for few sudden drops (zero values). These zero-valued points indicate dropped frames. Since these frames are not included in the final bitstream and hence not transmitted, the delay is indicated as zero. Ideally, The ROI-based encoding approaches should not alter the shape of this plot. The ROI-based bitrate control should re-distribute the bits within a frame with minimum error carried to the next frame. It is also not desirable to have any increase in the number of dropped frames.

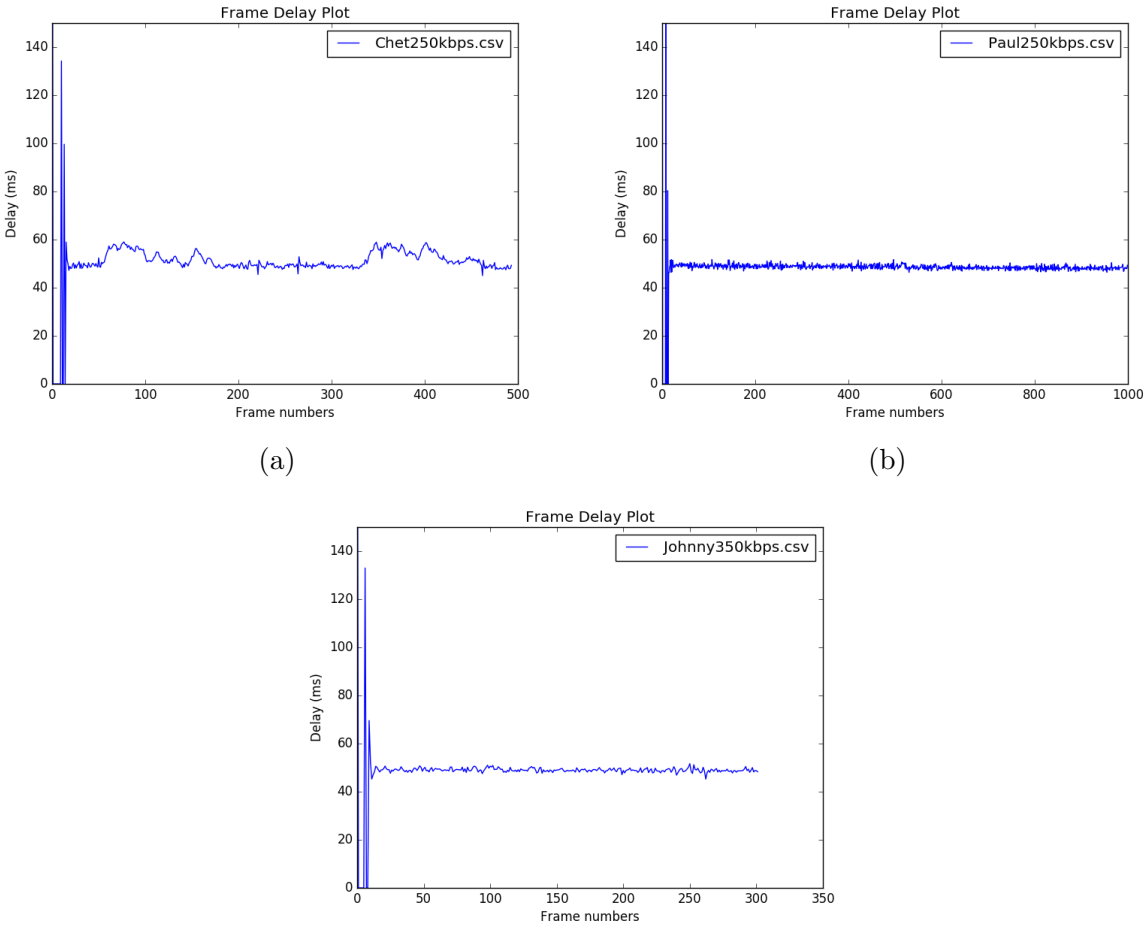



Figure 5.5: Delay plot for conventional encoding (a)chet640x480, (b)Paul640x480 ,(c)Johnny1280x720.

# Chapter 6

# Face Detection

Face detection algorithms are used to mark the region of interest in the current frame. All the algorithms considered for intelligent bit allocation involve improving the region of interest at the cost of rest of the frame. Therefore, it is very important to have high reliability with face detection. Any false detection will lead to degradation of the actual region of interest compared to normal encoding, this should be avoided in all scenarios. The damage caused by false detection is higher than the loss due to not detecting any face. Therefore, a high threshold must be used to declare any region of the frame as face.

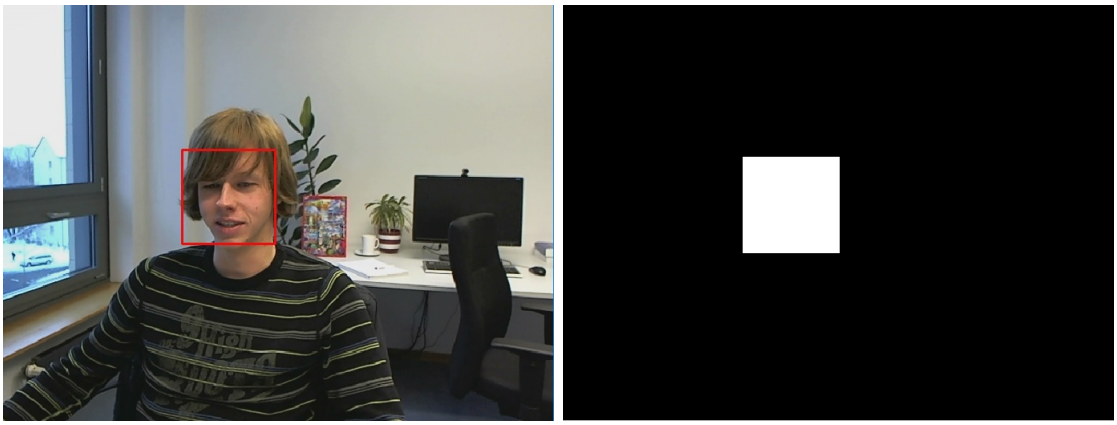


Figure 6.1: Face Detection binary map

The face detection module itself shall not be a part of the encoder, but the output from the face detection is a binary file with face regions marked is used as input by the encoder. In this work, the face detection module uses input YUV to the encoder and marks the region of interest at the macro block level. Each byte value in the output file of face detection module represents a macro-block scanned in raster scan order. A value of 0xff signifies macro-block being part of the face or region of interest and 0x00 represents a normal macro-block. Figure 6.1 represents the face map generated for the frame shown in

Figure **??**. The region in white is considered as region of interest, this information is used inside the bitrate control module of the encoder to perform intelligent bit allocation.

Different approaches are used to detect the face region in the video. There is always a trade-off between accuracy of face detection algorithm and its complexity. The work presented here is mostly relevant to real time systems. Any added complexity due to additional module of face detection will cause significant delay which is totally unacceptable in such systems. Therefore, the algorithm chosen for face detection must be light weight and reasonably accurate in all lighting conditions.

## 6.1 Spatial Domain Face Detection

In this approach, the face detection algorithms work directly on the pixel values. This approach is simple in terms of implementation. Many open-source solutions like OpenCV offers a ready to use solution that can be integrated with the codec library. It has large set of trained classifiers considering many types of faces and viewing angles. However, this is a computation intensive approach and almost impractical to use in the final solution.

## 6.2 Compressed domain Face Detection

Most available face detection algorithms work on the pixel domain. These algorithms provide good level of detection accuracy. The main drawback of this approach is that they are computationally intensive. As discussed earlier, the use case considered in this work has very less room for additional computations. Therefore, in this work ways of compressed domain face detection is explored to detect faces with less computational requirements. Many works have been published TBD LATER

# Chapter 7

# ROI-based Bitrate Control

As discussed in the previous sections there are many ways in which ROI information can be used to improve the perceptual quality of the video. This section describes the following two major approaches of ROI-based bitrate control to enhance the quality of the ROI macroblocks.

- ROI QP Offset
- ROI Bit-Allocation

These approaches are designed considering the bitrate control module presented in section 4. However, the underlying principles of ROI-based encoding to create an optimal quality difference between ROI and non-ROI are applicable to generic bitrate control modules. The principles presented in this work can be used to modify other standard low-delay bitrate control algorithms to produce equivalent results. The different approaches presented here vary in terms of ease of implementation, complexity and output quality. This offers flexibility to choose a suitable approach according to the specific requirements. The following subsections describe each of the ROI-based bitrate control approaches.

## 7.1 ROI QP Offset

A simple way of creating a bias in the quality between ROI and non-ROI is by using a QP offset for the macroblocks belonging to the ROI. A negative QP offset ($dq_{roi}$)is added to the QP allocated by the bitrate control module ($Q_m$) for ROI macroblocks. The QP for a macroblock assigned by the bitrate control is modified before using it for the final encoding as shown below.

$$Q_m^{roi} = Q_m - dq_{roi},$$
$$Q_m^{nroi} = Q_m.$$

Where, $Q_m^{roi}$ and $Q_m^{nroi}$ are the QP used for encoding a macroblock belonging to ROI and non-ROI parts respectively. The QP offset is used outside the bitrate control module, hence this approach requires minimum or no modifications to the bitrate control module.

The effects on encoding of non-ROI macroblocks due to usage of external QP offset for ROI macroblocks are:

- A lower QP for ROI results in ROI macroblocks consuming more bits.
- The rate control module obtains feedback from the encoder regarding bit-consumption at the macroblock level (Section 4). This feedback signals over-consumption of bits by ROI macroblocks resulting in a higher deviation factor ($D_m^{n'}$).
- The bitrate control reacts to the usage of reduced QP for ROI macroblocks by increasing the QP of non-ROI blocks to compensate for the additional bits used by ROI macroblocks.
- This results in the frame level bit-consumption very similar to conventional encoding without using any ROI-based QP offsets. Therefore, the effect of ROI-based QP offset is mostly neutralized within the frame.
- Any error in bit consumption at the frame level is compensated in the subsequent frames.

The feedback from the encoder to the bitrate control module will ensure that target bitrate is successfully achieved. The bit-allocation based on buffer level (section 4.1) will ensure that a constant delay is maintained even after using the QP offset. This approach offers the simplest way of implementing quality bias without breaking the core functionality of the bitrate control module.

### 7.1.1 Conventional and ROI-based encoding Comparison

The images in Figure 7.1 shows the comparison between conventional encoding and ROI-based encoding using QP offset and their corresponding attributes like PSNR map and quantization map. A QP offset ($dq_{roi}$) of 4 was used during ROI-based encoding. The image in Figure 7.1b looks better in the face regions compared to Figure 7.1a due to usage of negative QP offset for ROI. This improves the overall perceptual quality in a video. Figure 7.1e-7.1f shows the effect of QP offset on the final QP used for encoding the macroblock. The macroblocks in the face region of 7.1f have lower QP which is seen as macroblocks with a lighter shade of gray compared to the other macroblocks in the frame.

The PSNR maps in Figure 7.1c-7.1d shows that the PSNR of the face is closer to the background region with ROI-based encoding using QP offset. There is no major difference in the PSNR of the background region (non-ROI). The face appears much sharper due to the additional boost in quality from reduced QP. The overall bitrate of both the compared

(a) (b)

(c) (d)

(e) (f)

Figure 7.1: The Comparison of conventional encoding and ROI-based encoding with QP offset of 4 for ROI. The figures (a), (c), (e) and (b), (d), (f) correspond to conventional encoding and QP offset based ROI-encoding approaches respectively. (a), (b) are snapshots from the encoded output of Chet640x480. (c), (d) are PSNR maps and (e), (f) are Quantization maps corresponding to the frames in (a) and (b).

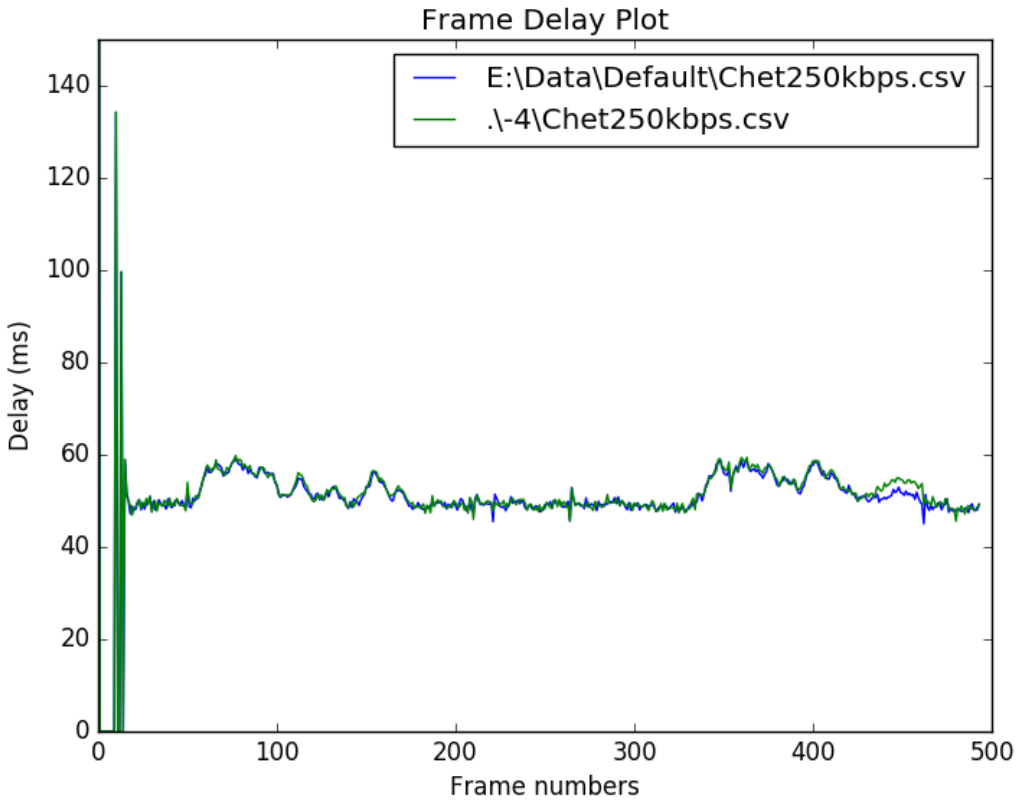



Figure 7.2: Delay plot for conventional encoding and ROI-based encoding with QP offset of 4 for ROI (Purple - Conventional encoding, Greem - QP offset based ROI encoding)

bitstreams remained almost equal. The difference in the perceptual quality is only due to the movement of bits from non-ROI to ROI macroblocks.

Figure 7.2 shows the delay plot described in section 5.3.3 for both conventional and ROI-based encoding using QP offset. The visibility of a single color predominantly (due to overlapping) shows that there is no significant change in the bit-consumption at the frame level. This implies that the additional bits consumed by the ROI macroblocks are compensated in non-ROI blocks within a given frame. There is very less difference in bit-consumption at the frame level that is carried over to the next frame. There is no significant change in the number of dropped frames which are represented as zero points in the delay plot. The number of dropped frames with conventional and ROI-based encoding remains the same. The dropped frames also appear at the same time interval in both output bitstreams.

The PSNR of ROI-based encoding with QP offset is tabulated in Table 7.1. There is a considerable improvement in the PSNR of ROI. There is a corresponding drop in PSNR of non-ROI which is reflected in reduced overall frame PSNR. The drop is not significant leading to a overall increased perceptual quality.

| QP Offset | Content | PSNR Avg (dB) | PSNR ROI (dB) |
|---|---|---|---|
| -4 | Chet640x480, 250kbps | 31.22 | 34.27 |
| -8 | Chet640x480, 250kbps | 30.87 | 35.51 |
| -12 | Chet640x480, 250kbps | 30.39 | 36.54 |

Table 7.1: PSNR Comparison for for QP offset based ROI encoding

The boost in ROI PSNR is dependent on the magnitude of QP offset used for the ROI macroblocks. A QP offset of -4 might not be optimum QP offset for all the contents. Therefore, It is necessary to consider following aspects during QP offset computation to achieve optimal PSNR for ROI and non-ROI parts resulting in improved perceptual quality.

- Tuning QP offset - Understanding the effect of using different QP offsets.
- Area of ROI - The relative area of ROI and non-ROI parts in a frame.
- Bi-direction QP offset - A positive QP offset is used for non-ROI blocks along with negative QP offset for ROI.

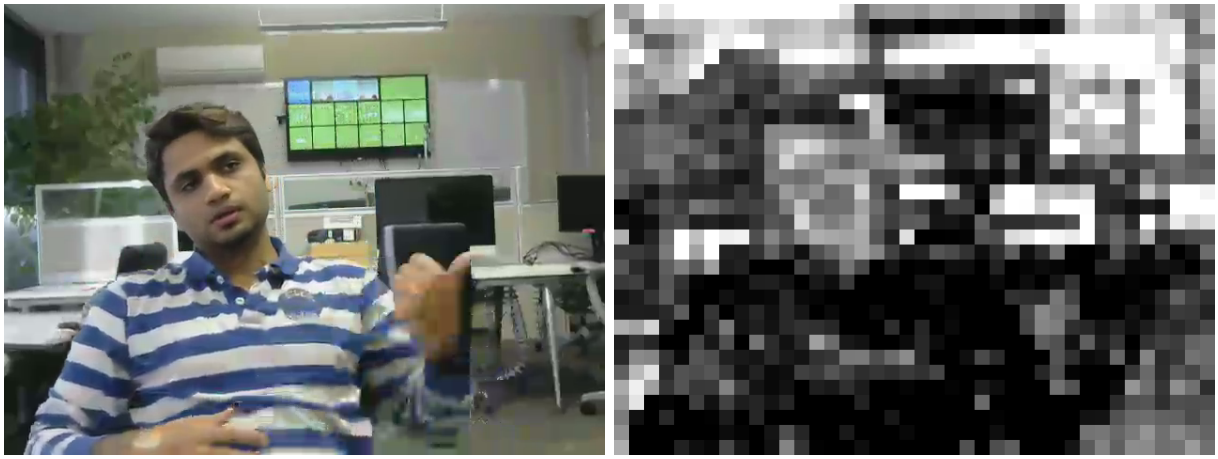
The following section discusses each of these aspects in detail to form a generic QP offset computation approach applicable to most of the contents.

### 7.1.2 Tuning QP Offset

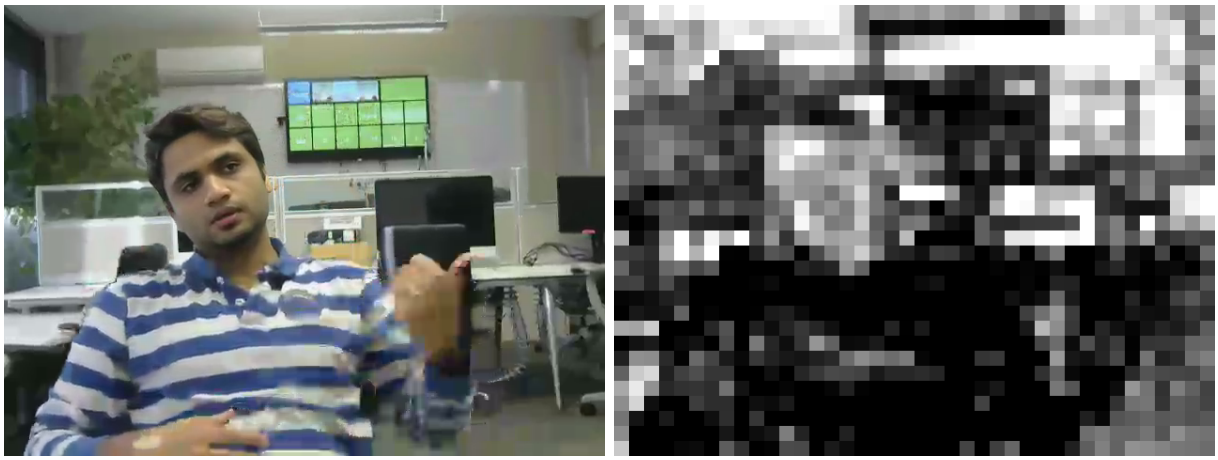
As mentioned earlier, the ROI-based encoding approaches discussed in this work only aim to re-distribute the bits within a frame based on the region of interest. The magnitude of re-distribution should be carefully chosen to avoid degradation of the background to an extent that artifacts become noticeable to the viewer even though those regions are not of primary importance to the viewer. An ideal redistribution of bits will make sure that there is a maximum transfer of bits from non-ROI part to ROI part without creating any visible artifacts in the non-ROI parts of the image.

In this work, various offsets were used to study the effect of magnitude of QP offset on perceptual quality. The results of ROI-based encoding with QP offset of -4 shows favorable results with increased perceptual quality. This section examines the impact of increasing the magnitude of QP offset further.

The snapshot of sample input encoded with QP offset of -8 shown in Figure 7.3a has sharper face region compared to same fame encoded with QP offset of -4 shown in Figure 7.1b. This is expected since the face region is coded with lower QP due to a larger negative QP offset. The blockiness in the non-ROI, specifically around the arm region has increased. The increase in sharpness in face region masks this blockiness to some extent. However, with further increase in the magnitude of QP offset the blockiness in the background becomes more prominent. The image in Figure 7.3b is encoded with QP offset of -12. At this



(a) QP offset = -8



(b) QP offset = -12

Figure 7.3: The snapshot of ROI-encoded Chet640x480 (Left) and its corresponding absolute PSNR map (Right) for different QP offsets.

point, the increase in sharpness in the face region is masked by extreme blockiness in the non-ROI, specifically around the arm region. It is clear from the images in Figure 7.3. The corresponding PSNR plots show extreme low PSNR in the non-ROI with increasing QP offsets. Therefore, a very high QP offset can have adverse effect the perceptual quality.

### QP Swing Restriction

It is possible to preserve the quality of the non-ROI macroblocks even with high QP offsets by using QP swing restriction discussed in section 4.2.2. The maximum value of macroblock QP ($QP_{max}$) given by (4.10), avoids the excessive degradation of the non-ROI macroblocks due to the usage of large QP offset. The QP of non-ROI blocks is not allowed to go very high despite over-consumption of bits by ROI blocks. The side-effects of increased QP

| | Encoding Mode | |
|---|---|---|
| **ROI QP Offset** | **Swing Restriction = on** | **Swing Restriction = off** |
| 0 | 16 | 12 |
| -4 | 19 | 12 |
| -8 | 31 | 12 |
| -12 | 45 | 14 |

Table 7.2: Number of dropped frames with different QP offsets for ROI for test sequence with 1000 frames.

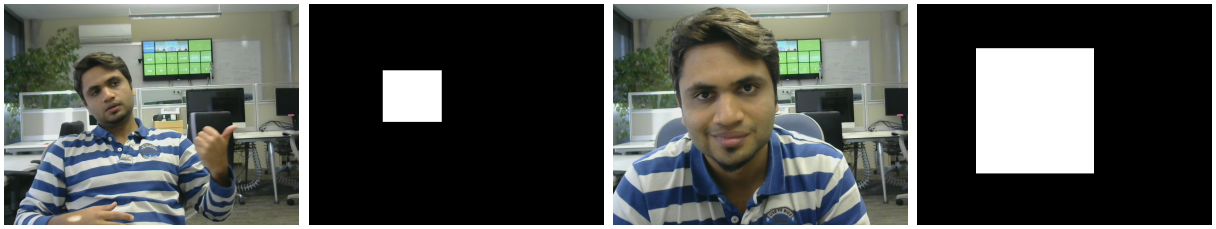
offset shows up in the form of an increase in the number of dropped frames. Table 7.2 shows the number of dropped frames in the encoded video with 400 input frames. There is a drastic increase in the number of dropped frames with QP offset of -8 and -12. This increase in dropped frames reduces the smoothness of the playback which is annoying to the viewer.

When the QP swing restriction was turned off, the number of dropped frames with increased QP offset reduced drastically as shown in Table 7.2. There was no considerable increase in dropped frames compared to the output of conventional encoding without using any ROI information. However, the quality of non-ROI blocks dropped significantly without QP swing restrictions as shown in Figure 7.3. The extreme blockiness in the non-ROI regions decreased the perceptual quality. Therefore, using large QP offsets can lead to less perceptual quality either due to increased dropped frames or excessive blockiness in the background depending on the configuration of QP swing restriction. Therefore, the QP offset for the ROI should be tuned not only considering the degradation of the background quality but also by assessing any other side-effects like an increase in the number of dropped frames.

### 7.1.3 Area of Region of Interest

This section discusses the importance of considering the relative area of ROI with respect to the non-ROI parts in computing the QP offsets. The variation in QP offsets discussed in the above section corresponds to a frame with relatively smaller ROI area compared to the whole frame. In this frame, 64 macroblocks out of a total of 1200 macroblocks belong to the face region(ROI). Therefore, the ROI is less than 5 percent of the entire video frame. Due to the small percentage of ROI blocks, it is possible to use large QP offset (very low ROI QP) since there are a large number of non-ROI macroblocks to compensate for the over-consumption of bits by ROI macroblocks.

However, in a video conferencing scenario, based on the focal length of the camera and distance of the participant from the camera, area of the face in the video frame can change significantly. The number of ROI macroblocks can also change within a given input video



(a) (b) (c) (d)

Figure 7.4: The snapshots and corresponding face map of frame number 119 (a),(b) with $A_{roi} = 0.675$ and frame number 446 (c),(d) with $A_{roi} = 0.35$ for the content chet640x480.

sequence. The images in Figure 7.4 shows the variation in size of ROI within a video sequence. This demands continuous adaptation of the QP offset to avoid visual artifacts due to the usage of wrong QP offset. For instance, when the area of ROI is half of the entire frame, usage of higher QP offsets will cause severe degradation in the quality of non-ROI macroblocks. In order to avoid severe degradation of the background, the magnitude of the QP offset should be inversely proportional to the ratio of the number of ROI blocks to non-ROI blocks.

The algorithm implemented to adapt QP offset uses relative area of ROI. The QP offset is calculated using linear relationship described in (7.1). This is a heuristic approximation which was found to yield best perceptual quality across many video contents. The scaled QP offset is then clipped to a value of -6 to avoid large QP offsets which can lead to side-effects discussed in the previous section.

$$\begin{aligned} dq_{roi'} &= -round\Big(\frac{M}{M_{roi} * 3}\Big), \\ dq_{roi} &= clip(dq_{roi'}, -1, -6), \end{aligned} \tag{7.1}$$

where, $dq_{roi}$ is the offset used for ROI blocks, $M$ is the total number of macroblocks in the frame and $M_{roi}$ is the total number of macroblocks marked as region-of-interest. The negative sign in the equation implies that the calculated offset is negative, which results in QP lower than non-ROI blocks. It is also evident from (7.1) that there is no QP offset for ROI region if the ROI covers more than two-third of the whole frame. The QP offset increases linearly with subsequent decrease in the ROI area.

### 7.1.4 Bi-direction QP-offset

The QP offset computed so far is only applied to the ROI macroblocks. The bitrate control module is held responsible for compensating the additional bits used in encoding the ROI blocks by increasing the QP of non-ROI blocks.

As explained in section 4.2.2, the bitrate control uses feedback from the encoder to constantly react to any deviation in the bitrate at the macroblock level. This feedback loop is

not aware of the QP offset which is applied externally to QP computed by bitrate control ($Q_m$). Due to this, the bitrate control reacts to over-consumption of bits by increasing the QP of non-ROI blocks only after encoding the ROI macroblocks. This effect is clearly seen in the quant map of ROI-based encoding in Figure 7.1f.

The macroblocks encoded immediately after ROI have larger QP (depicted as a darker shade of gray). Therefore, this approach does not increase the QP of non-ROI blocks uniformly. The non-ROI blocks encoded before ROI blocks have no increase in QP since reaction by rate control to increased deviation in bitrate is not triggered until the ROI macroblocks are encoded. The non-ROI macroblocks encoded after ROI blocks tend to have lower quality than the non-ROI blocks encoded before the ROI blocks. This non-uniform loss of quality in non-ROI blocks is not desirable for good perceptual quality.

The non-uniform increase in non-ROI QP also depends on the position of the ROI within the video frame. For instance, consider a content where the ROI part in the frame falls in the bottom right corner of the frame. Assuming that macroblocks are encoded in the raster-scan order, the over-consumption of bits by the ROI part cannot be compensated within that frame. The error is carried over to the next frame. This alters the frame level bit-consumption behavior and can lead to increased number of dropped frames in extreme cases.

A Bi-direction QP offset is used to avoid the behavior described above. The non-ROI blocks are assigned with a positive QP offset, which can compensate the over-consumption in ROI blocks. The non-ROI blocks are encoded with a higher QP even before ROI blocks are encoded. Since the non-ROI blocks from the start of the frame are encoded with higher QP, there will be a surplus of bits already which can be used in encoding ROI blocks. In an ideal scenario, the negative and positive QP offsets must negate each other's effect resulting in frame level bits being unchanged had the frame been encoded without any offset.

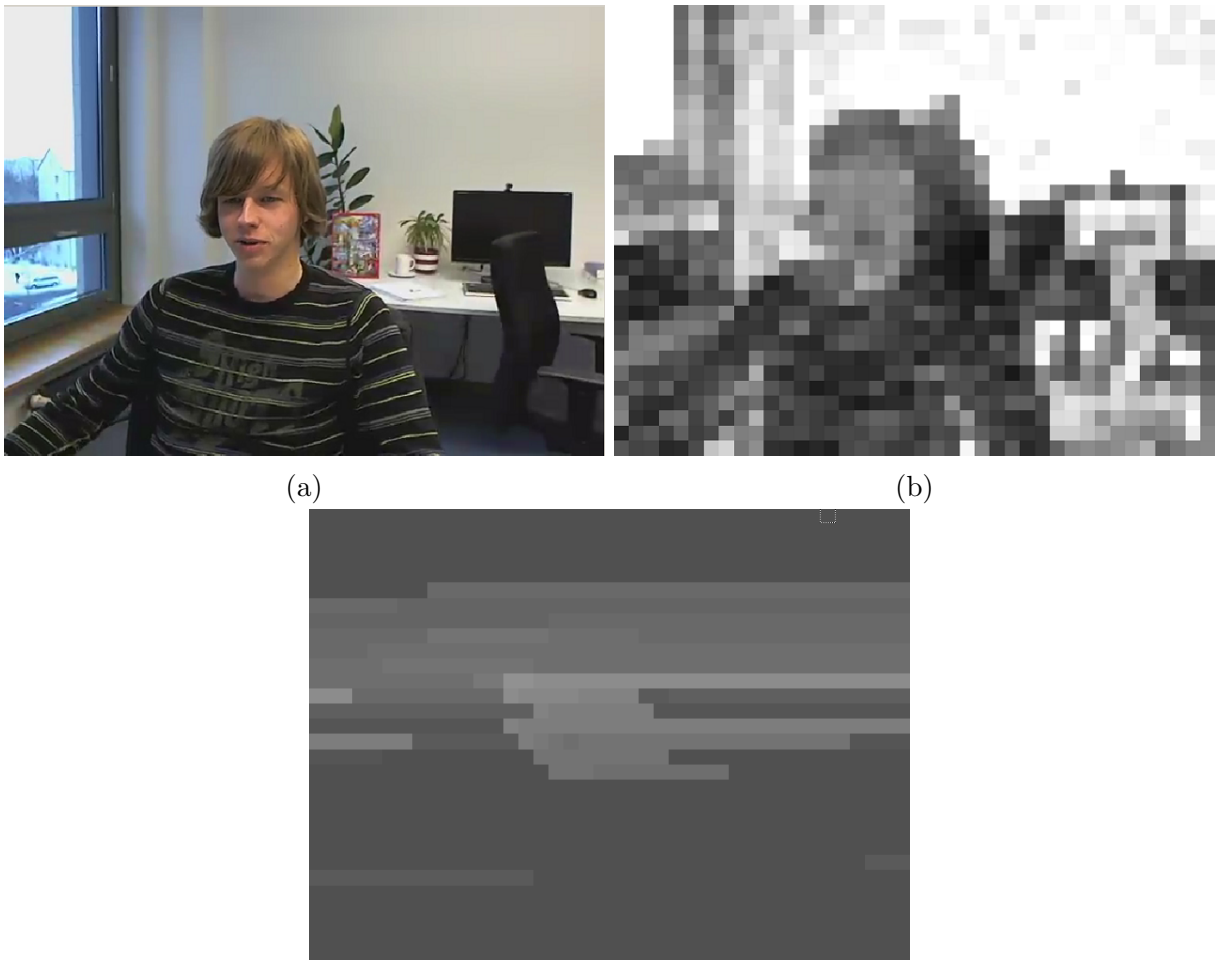
The study in [MB13] suggests that for the frame level bitcount to be constant, the average QP of the frame must remain unchanged before and after adding the offsets. This is an observation made after multiple experiments. The QP offset for non-ROI macroblocks used to compensate for negative QP offset used by ROI is given by,

$$dq_{nroi} = \frac{M_{roi} * dq_{roi}}{M - M_{roi}} \tag{7.2}$$

where, $dq_{nroi}$ is a positive QP offset used for non-ROI blocks when negative QP offset of $dq_{roi}$ is used for ROI blocks (7.1).

### 7.1.5 Results

The results of improved QP offset based ROI encoding is shown in Figure 7.5. The improvements include using of relative area based QP offset and Bi-direction QP offset discussed in the previous sections. The quantization map shown in Figure 7.5c has less increase in the



(a) (b)

(c)

Figure 7.5: The results of ROI-based encoding with Bi-direction QP offset.(a) Snapshot of encoded video, (b) absolute PSNR map (range 25dB - 50dB), (c) quantization map

non-ROI QP compared to approach without bi-direction QP offset shown in Figure 7.1f. This approach also works with any size of ROI since the QP offsets are scaled according to the relative area of the ROI.

This section analyses the advantages and disadvantages of ROI-based encoding using QP offsets. Since the QP offset is applied outside the bitrate control, there is less possibility of other factors like buffer fullness and global deviation ($D_m^{n'}$) overriding the QP offset. Such guaranteed QP offsets will ensure boost in the ROI quality at all circumstances. However, this approach has many side effects due to the forced QP offset. The main disadvantage of this approach is not considering the content to decide the QP offset. Any given input with a given ratio of the number of ROI blocks to the number of non-ROI blocks will have the same QP offsets irrespective of the content. The QP offset computation is empirical which might work for most of the contents. However, it is not guaranteed to give optimal results for all contents.

The advantages and disadvantages of this approach can be summarized as follows,

## Advantages

- Easy to implement with minimal changes required to the encoder.
- QP offset is forced irrespective of buffer conditions, therefore guarantees quality difference between ROI and non-ROI.

## Disadvantages

- Since buffer conditions are not considered at macroblock level, there is increased possibility of dropped frames.
- The QP offsets chosen does not take into account any of the input content characteristics. Same QP offset is used for all the contents for a given ROI to non-ROI ratio.
- The deviation control mechanism described in section 4.2.2 is not in sync with QP offsets.

The next section introduces an alternative approach for ROI-based encoding using region based bit-allocation. Some of the disadvantages of the QP offset approach are addressed in the new approach.

## 7.2 ROI based Bit-Allocation

This section introduces an approach with modifications to the bit-allocation module for improved ROI-based encoding. The main drawback of the QP offset based approach is that the input video content characteristics are not considered for determining the QP offset. The difference in complexity of ROI and non-ROI macroblocks determines the magnitude of quality difference required for good perceptual quality. This magnitude of optimal quality difference varies across different contents depending on the variation in complexity within a frame. For instance, consider a sample input video with moving background regions (non-ROI) but a simple static foreground (ROI). The usage of heuristic QP offsets which assumes background to be mostly static will result in annoying blocky artifacts in the background. This might decrease the perceptual quality to a level below that of conventional encoding. Therefore, it is important to consider the relative complexities of ROI and non-ROI parts to compute the desired quality difference between these regions.

In conventional encoding, the bit-consumption is not uniform across all the macroblocks even though all regions of a frame are considered equally important to the viewer. This is due to variation in complexities across different macroblocks. The image in Figure 7.6 shows relative macroblock size in bits for a conventionally encoded frame. Similar to quant and psnr maps, the brighter shade of gray represents larger bit-consumption. It is clear that background regions consume least amount of bits and face region consumes an above average amount of bits. A bit-allocation strategy for ROI-based encoding which allocates higher than average amount of bits to ROI based on arbitrary weights might not allocate any additional bits compared to the conventional encoding. For ROI-based encoding, the bit-allocation strategy should take into account the relative complexities of ROI and non-ROI and allocate bits with a bias over the complexity based bit-allocation. The fundamental idea behind this approach is to allocate bits to ROI and non-ROI proportional to their complexities with an additional bias for ROI.

In this approach, the bits allocated at frame level ($B_{alloc}$) is split between ROI and non-ROI parts based on the relative complexities of these regions. The cost of a macroblock computed during rate-distortion optimization (4.6) is accumulated for ROI and non-ROI regions to get the relative complexity ($C_r$),

$$C_r = \frac{\sum_{i=0}^{M_{roi}} J_i}{\sum_{i=0}^{M-M_{roi}} J_i},$$

where, $J_i$ is the RDO cost of the macroblock $i$. The cost of a region determines the bits consumed by the corresponding region. The relative bit consumption of ROI and non-ROI is predicted assuming that,

$$C_r \propto B_R, \tag{7.3}$$

where,

$$B_R = \frac{B_{roi}}{B_{nroi}}.$$

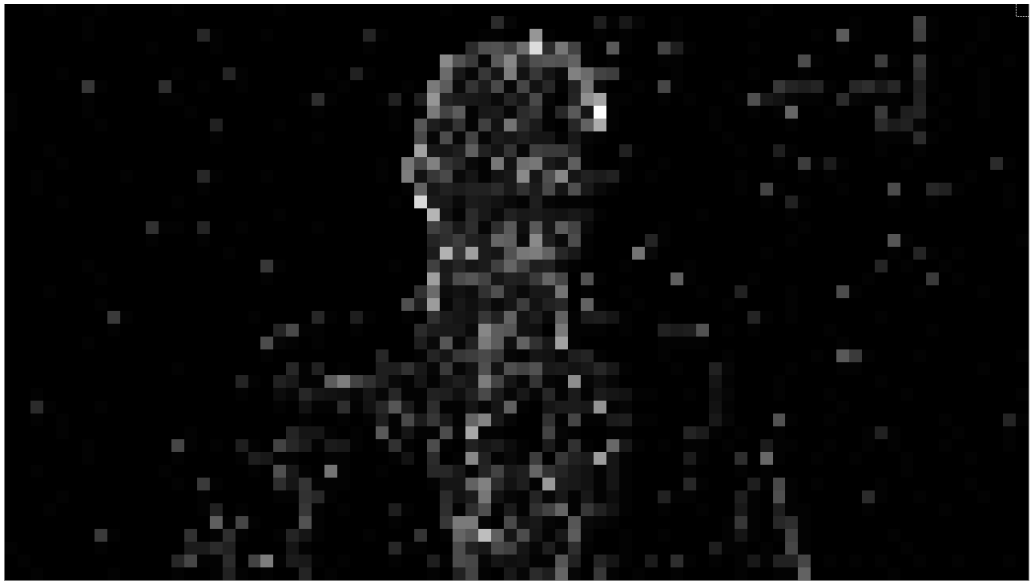


Figure 7.6: Macroblock level bit-consumption for Johnny1280x720 (Conventional encoding).

Here $B_{roi}$ and $B_{nroi}$ are the accumulated bit consumption of $M_{roi}$ macroblocks belonging to ROI and $M_{nroi}$ non-ROI macroblocks respectively. The above equations hold true for any two regions in a given frame given that these two regions are encoded with equal importance (no ROI-encoding). Therefore, $B_{roi}$ and $B_{nroi}$ corresponds to bit consumption of the two regions if the frame was encoded conventionally without using any ROI information. The idea is to predict the bit-consumption in conventional encoding and allocate bits for ROI and non-ROI regions according to the estimate but with an additional bias for the ROI part.

The ROI-based bit-allocation involves following major steps.

- *Relative Bit-consumption Prediction* - Predict relative bit-consumption of ROI and non-ROI ($B_R$) in conventional encoding.
- *ROI and non-ROI Bit Allocation* - Split allocated frame level bits ($B_{alloc}$) between ROI and non-ROI parts according to ($B_R$) with an additional bias for ROI parts.
- *Region-based bitrate control* - Independent bitrate control for ROI and non-ROI regions.

The following sections discuss each of the above steps in detail.

### 7.2.1 Relative Bit-consumption Prediction

The first step in ROI based bit-allocation is to determine the relative bit-consumption of the ROI and non-ROI parts ($B_R$) if the frame was encoded normally without any bias for ROI regions (conventional encoding). There are two possible ways of achieving this,

1. Encode a frame using conventional encoding without any bias for ROI, measure the bit-consumption for ROI and non-ROI. The frame can be re-encoded with a bias in bit-allocation for ROI.
2. Estimate the complexities of ROI and non-ROI parts and use a cost-bits model to predict relative bit-consumption.

The first approach is very inefficient and not suitable for real-time encoding scenario. Therefore, a cost-bits model is used to predict $B_R$.

In this work, the relationship between the complexity of a macroblock and its bit-consumption during conventional encoding is studied. The RDO cost (4.6) is chosen as the complexity metric for prediction of $B_R$. The computation of a new complexity metric for predicting bit-consumption is not desirable due to increased complexity. The RDO cost which is computed by the encoder during rate-distortion optimization stage is reused to predict $B_R$.

#### Effect of RDO Cost on bit-consumption

The cost of a macroblock affects its bit-consumption in two opposite ways listed below.

- The cost of a macroblock reflects the complexity of the macroblock. A macroblock with a higher cost consumes more bits (7.3).
- The RDO cost is also used in the bitrate control for delta QP prediction as discussed in section 4.2. The delta QP computation (4.5) allocates a higher QP to the macroblock with a higher cost for spatial and temporal masking considering HVS. This will marginally reduce the bit-consumption of the macroblock.

These conflicting influence of cost on bit-consumption makes its prediction more complex. In this work, the effect of delta QP on bit-consumption is ignored to for simplicity. A cost-bits model is developed assuming no variation in QP across the macroblocks.

#### Cost-Bits Model

A model to relate cost of a macroblock and its bit-consumption is developed using data from multiple video conferencing sequences. Figure 7.7 shows the plot of cost vs bits at the macroblock level. Every point on this plot corresponds to bit-consumption of a macroblock and its corresponding RDO cost computed during the RDO stage. The plot corresponds

to all the macroblocks in the encoded video sequences with 50 frames each from four different video conferencing samples. The sample input sequences listed in section 5.1 are not included for developing the model for a fair evaluation. The input video sequence used for developing the model and the ones used for evaluating the model are different. The encoding for this plot was done in constant QP mode with QP=35 for all the macroblocks. The data of the first key frame is not included in this plot since the first key frame is encoded without using any ROI information. It can be noticed that there is a linear relationship between predicted macroblock cost and the corresponding bit-consumption.

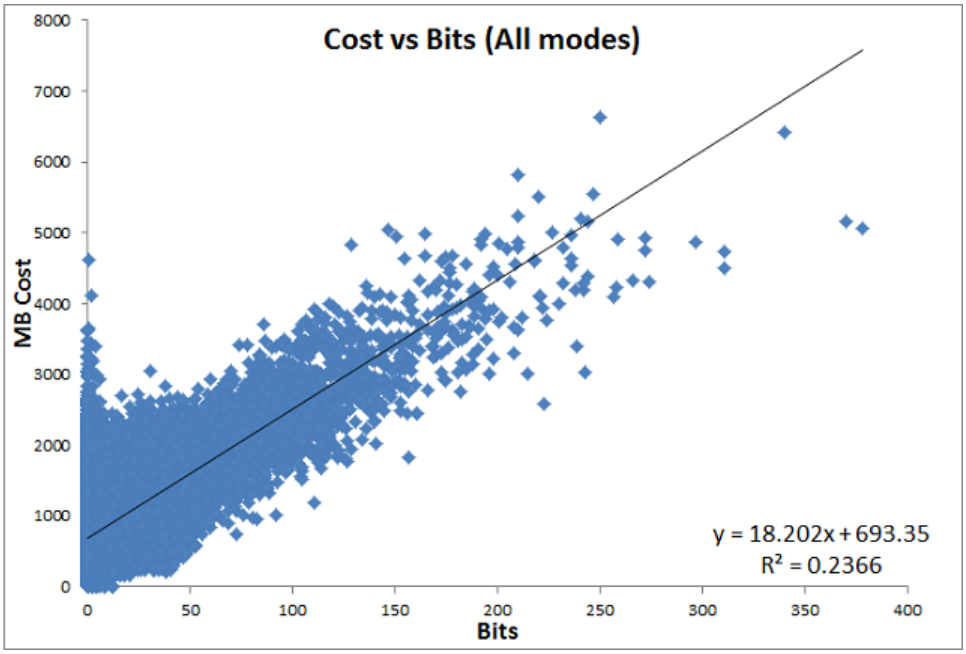



Figure 7.7: Macroblocks cost vs bits plot for all the macroblocks of multiple video sequences

The linear relationship between cost of a macroblock (J) and its corresponding bit-consumption (b) at a constant QP is given by,

$$J = 18.2.b + 693. \tag{7.4}$$

The above equation is obtained by performing linear regression on the entire data set shown in Figure 7.7. It has a corresponding $R^2$ value of 0.2366. The $R^2$ value (also displayed on the plots) is a measure of correlation between actual data and the predicted data using the linear relationship. A value of $R^2 = 1$ indicates perfect correlation. The $R^2$ value in this case is lower to form an efficient prediction model. The correlation between macroblock cost and its bit-consumption is increased further by considering the prediction mode of the macroblock.

### Cost-Bits Model - Prediction Modes

The data in Figure 7.7 includes data for macroblocks of all types of prediction modes like skip, intra and inter prediction. It was observed that the predictability of bit-consumption using RDO cost was heavily dependent on the prediction mode used for encoding the macroblock.

The analysis of the macroblock cost and its corresponding bit-consumption specific to a prediction mode is shown in Figure 7.8. The plots in Figure 7.8 are generated using the same data as shown in Figure 7.7. Different plots are generated for skip, inter and intra prediction modes. It can be noticed that the correlation between the macroblock cost and its bit-consumption is maximum for intra-prediction ($R^2 = 0.8757$) followed by inter-prediction ($R^2 = 0.2995$). However, the bit-consumption pattern was found to be almost independent of the cost for skip-mode macroblocks ($R^2 = 0.0412$). This is because skip-mode decision happens in a later stage of encoding when there are no transform coefficients generated for a macroblock encoded with intra or inter-prediction. The cost associated with skip macroblock corresponds to the encoding mode chosen before the macroblock was decided to be encoded as skip macroblock.

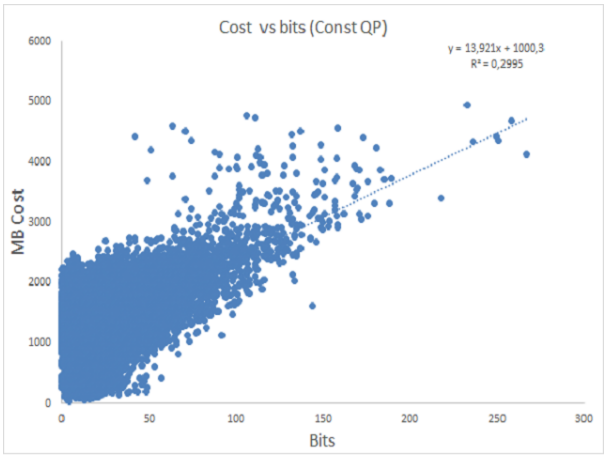
| **Prediction Mode** | **Occurrence Probability** | $R^2$ **value** |
|---|---|---|
| Intra | 0.233 | 0.8757 |
| Inter | 0.28 | 0.2995 |
| Skip | 0.69 | 0.0412 |

Table 7.3: The probability of prediction modes in constant QP (QP = 35) and its corresponding $R^2$ value.

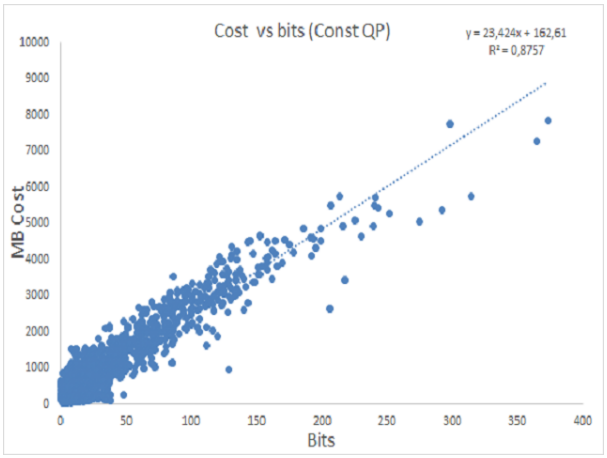
The data in Table 7.3 shows the probability of choosing a given prediction mode. It is clear that skip mode in the most commonly used prediction mode for the encoder configuration discussed in section 5.2. This is due to the low target bitrate and mostly static background in video conferencing videos. Therefore, it is very important to have good bits vs cost prediction model for skip macroblocks to compute $B_R$ with reasonable accuracy.

Fortunately, the extremely low correlation between macroblock cost and its bit-consumption for skip mode can be easily handled by assuming a constant number of bits for skip mode and ignoring its cost. The average amount of bits consumed by the skip macroblocks over the entire data-set was found to be 0.458 bits. This is extremely low bit-consumption compared to the intra and inter mode macroblocks. Therefore, a constant bit-estimate (average bits) is used in (7.3) to compute relative bit-consumption of ROI and non-ROI based on the number of skip macroblocks in each of these regions. The $R^2$ value of the prediction model is improved considerably by this assumption for the skip macroblocks.

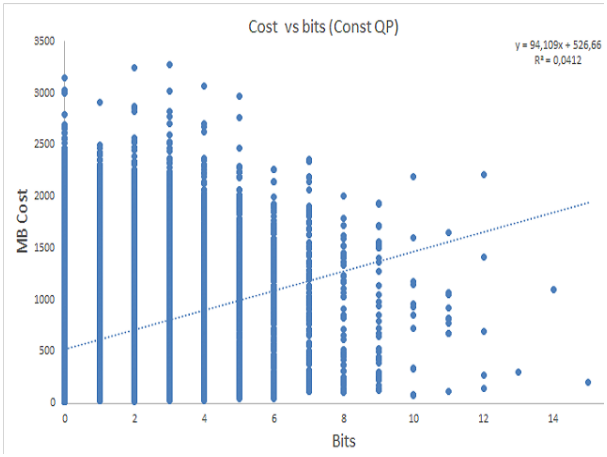
The steps involved in computing relative bit-consumption of ROI and non-ROI region ($B_R$) are summarized below.




(a) Inter Prediction mode




(b) Intra Prediction mode




(c) Skip mode

Figure 7.8: The Cost vs Bits plots of macroblocks specific to mode of encoding. (a) inter prediction mode, (b) intra prediction mode, (c) skip mode

- Identify the ROI and non-ROI macroblocks in the input frame.
- Get the number of intra macroblocks, inter-macroblocks and skip macroblocks in the ROI and non-ROI region separately. This can be done using data from the previous frame.
- Compute relative complexity of the ROI and non-ROI parts by accumulating costs of macroblocks in these regions.
- The relative complexity ($C_R$) is translated to relative bit-consumption ($B_R$) between ROI and non-ROI using prediction models shown in Figure 7.8. A pre-computed value is used for the skip macroblocks to account for bits consumed by skip macroblocks.

The procedure involved in using ($B_R$) to allocate bits to ROI and non-ROI parts with a bias for ROI parts to perform ROI-based encoding is discussed in the subsequent sections.

### 7.2.2 ROI and non-ROI Bit Allocation

This section gives an overview of the process involved in splitting the allocated frame level bits ($B_{alloc}$) computed in (4.3) between ROI($B^{roi}_{alloc}$) and non-ROI($B^{nroi}_{alloc}$) parts. This is done using relative bit-consumption factor ($B_R$) computed in the previous section. The idea behind ROI-based bit allocation is to estimate relative bit-consumption between ROI and non-ROI in conventional encoding, perform bit-allocation according to the estimate with a bias for ROI. The relative bit-consumption between ROI and non-ROI is given by $B_R$ computed in the previous sections. For ROI-based encoding, a bias factor k is introduced to create bias in bit-allocation. The bias factor signifies the importance of ROI over non-ROI. The bits allocated for ROI region is given by

$$B^{roi}_{alloc} = k \times B_R \times B^{nroi}_{alloc} \tag{7.5}$$

and

$$B^{nroi}_{alloc} = B_{alloc} - B^{roi}_{alloc}$$

Here k is the ROI bias factor. In conventional encoding where all the parts of the frame are considered equally important, $k = 1$. For ROI-based encoding, a value $k > 1$ is used to bias bit-allocation to allocate more bits to ROI considering its importance to the perceptual quality. The bias factor($k$) gives the flexibility of tuning the bias for ROI based on the importance of the ROI.

As mentioned earlier, the ROI-based bit-allocation is independent of the procedure involved in determining the relative bit-consumption factor ($B_R$). The approach presented in section 7.2.1 needs computation of relative complexities of ROI and non-ROI regions. A simpler approach is to assume that the frame is uniformly complex and splitting the allocated frame level bits ($B_{alloc}$) uniformly between ROI and non-ROI with bias factor specified in (7.5). Therefore, if the content information is not factored in for simplicity, the bits allocated for ROI and non-ROI is computed using (7.5) with,

$$B_R = \frac{M_{roi}}{M - M_{roi}}.$$

The bias, factor k determines the additional bits allocated to the ROI macroblocks. The main advantage of encoding with ROI-based bit-allocation compared to ROI-encoding based on QP offset is that it gives more control to guarantee minimum quality for the background macroblocks (non-ROI). For instance, the value of ROI-bias factor (k) can be bounded in such a way that non-ROI macroblocks are allocated at least half the amount of bits allocated during normal encoding. This ensures that the quality of non-ROI does not deteriorate to an extent that blockiness in the background reduces the perceptual quality.

In this work, the value of k is chosen to first allocate non-ROI half of the bits it would have consumed in conventional encoding. The excess is allocated to ROI parts. The maximum excess bits allocated to ROI is three times its bit-consumption in conventional encoding. The excess is allocated back to the non-ROI parts. Therefore, the value of k is changes depending on the complexities of ROI and non-ROI regions.

### 7.2.3 Region based Bitrate Control

In the previous stages, the procedure for splitting the allocated frame level bits ($B_{alloc}$) between ROI and non-ROI is described. The input to this stage is bits allocated for ROI ($B_{alloc}^{roi}$) and non-ROI ($B_{alloc}^{nroi}$) parts. The target is to achieve bit-consumption according to the split between ROI and non-ROI parts with minimal error.

The bitrate control module described in section 4 performs only frame level bit-allocation. There is no macroblock level bit-allocation. The task of the bitrate control module was to reduce the error between allocated frame bits $B_{alloc}$ and the frame level bit-consumption without worrying about bit-distribution within a frame.

In the ROI-based bit-allocation, the bit-allocation is not limited to frame level. The input frame is divided into two regions based on the ROI information and target bits are specified for each of these regions separately. The task of the bitrate control module is to meet bitrate for different regions independently (ROI and non-ROI) with minimal error. This section describes the modifications to the bitrate control module to accomplish this task.

The bitrate control module studied in section 4 computes a deviation factor ($D_m^{n'}$) based on accumulated error in frame level bit-consumption after encoding every frame as shown in (4.2). In steady-state, the deviation factor ($D_m^{n'}$) takes a value such that the error in frame level bit-consumption is minimum.

The following steps summarize the approach to implement region-based bitrate control using the exisiting bitrate control described in section 4.

- Compute the deviation factor for ROI ($Droi_m^{n'}$) and non-ROI ($Dnroi_m^{n'}$) regions independently. The deviation factors are updated using feedback data from ROI and non-ROI parts independently. This can be visualized as having two instances of rate control running for ROI and non-ROI regions treating ROI and non-ROI regions as separate frames.
- A single instance of macroblock level bitrate error is maintained as shown in (4.2). This ties the two instances of bitrate control together, so that frame level bit-error is minimized. The error in allocation and consumption of bits in the ROI and non-ROI is compensated throughout the frame resulting in less overall frame level error.
- The VBV compliance works with frame level bit-allocation and therefore remains unchanged. There will be no increase in number of dropped frames as long as frame

level deviation in bit-consumption is kept under check.

The major problem with two instances of global deviation for ROI and non-ROI, which are simultaneously updated at the end of encoding of a frame is that it can lead to severe oscillations in the frame QP. The error in allocated bits and consumed bits for a given region have opposite effects on two instance of global deviation which results in oscillations of average frame level QP. Figure 7.9a shows the variation in encoded frame size with the usage of two different instances of deviation factor corresponding to ROI and non-ROI parts. The alternative frames differ in size almost by a factor of 2. Such oscillations were observed in steady-state (frame number 400-450 in the example shown). This variation in quality across alternate frames affects the perceptual quality adversely

The oscillation of the frame average QP was avoided by keeping the update of two instances of deviation factors separate. The deviation factor for ROI and non-ROI is updated at the end of alternate frames. Since, the two updates are not happening simultaneously, there is increased stability resulting in near constant frame level bit consumption as shown in 7.9b.

| Methodology | Content | PSNR Avg (dB) | PSNR ROI (dB) |
|---|---|---|---|
| QP Offset | Paul640x480, 250kbps | 38.90 | 38.88 |
| | Johny1280x720 750kbps | 40.49 | 40.22 |
| Reaction Factor | Paul640x480, 250kbps | 38.22 | 39.50 |
| | Johny1280x720 750kbps | 39.71 | 40.74 |
| ROI-based Bit-allocation | Paul640x480, 250kbps | 38.86 | 39.70 |
| | Johny1280x720 750kbps | - | - |

Table 7.4: PSNR Comparison for different approaches of ROI encoding

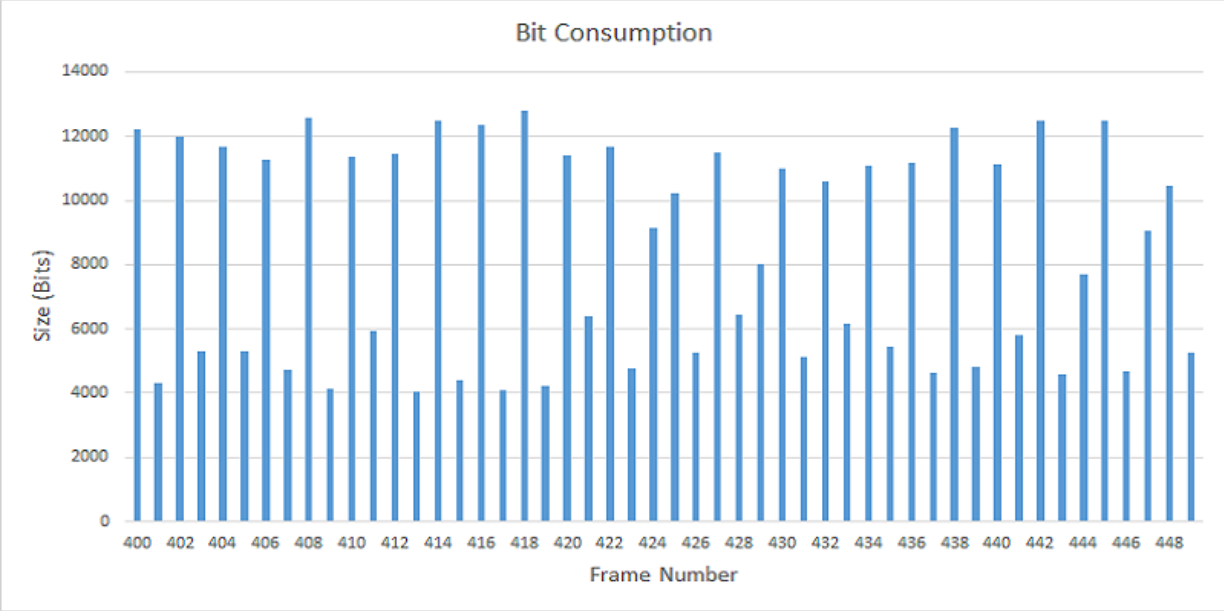
### 7.2.4 Results

The bit-allocation based ROI encoding considers input content characteristics to perform the bit-allocation for ROI blocks. Therefore, this approach is applicable to generic contents and hence can be expected to be more robust across multiple contents. The results of this approach is shown in Figure

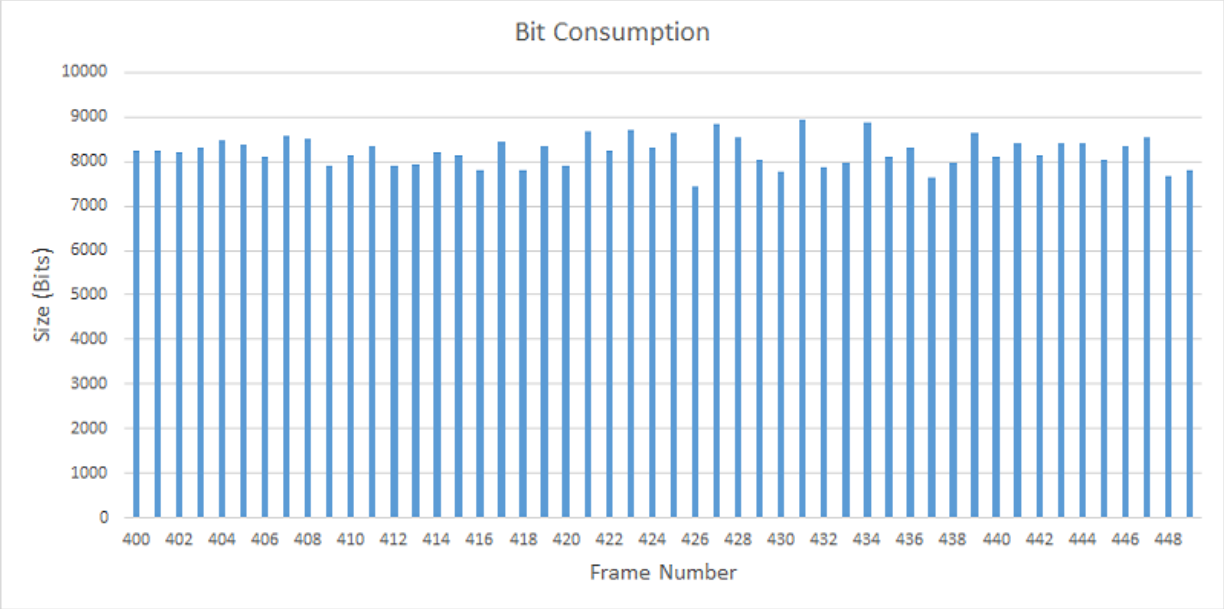
The advantages and disadvantages of this approach compared to QP offset approach can be summarized as follows.

## Advantages

- The consideration of input content characteristics for ROI-based encoding makes this approach work well for all types of contents




(a)




(b)

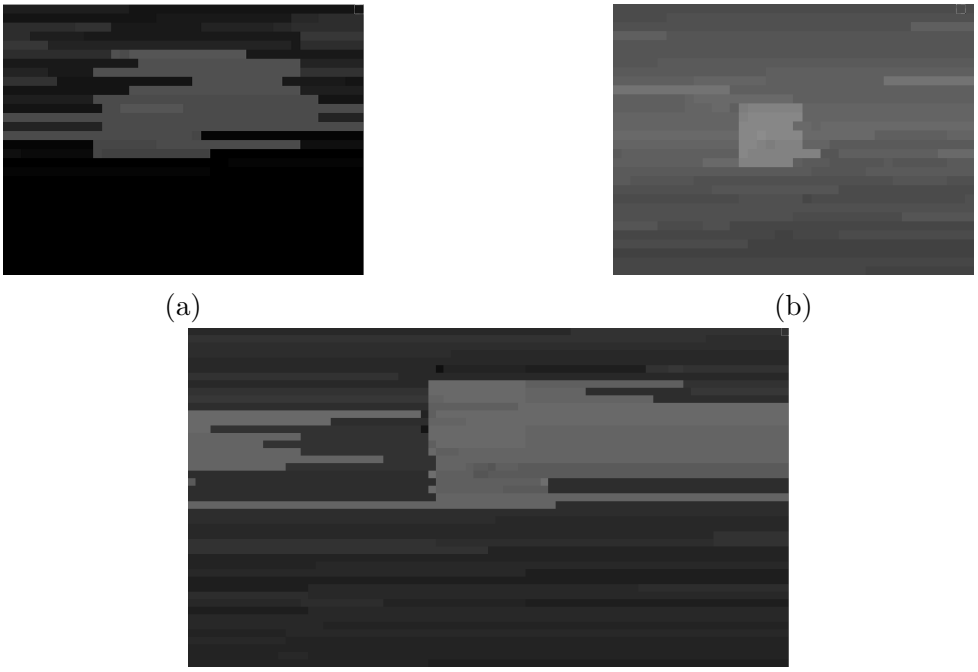
Figure 7.9: A plot of frame size in bits for 50 frames from frame number 400 to 450 for the encoded Paul250kbps content with Bit-allocation based ROI encoding. The two figures vary in approach used to update the deviation factors for ROI and non-ROI regions (a) Simultaneous update (b) Alternate Frame Update

(a) (b)

(c)

Figure 7.10: Result of ROI-based bit allocation encoding for the encoding configuration discussed in section 5.2 (a)chet640x480, (b)Paul640x480 ,(c)Johnny1280x720.

- This approach can guarantee minimum quality for the background region by having sufficient bits allocated to the non-ROI to avoid blockiness.
- The variable importance of ROI can be factored in easily by altering the bias factor (k) to vary the magnitude of bit movement from ROI to non-ROI.



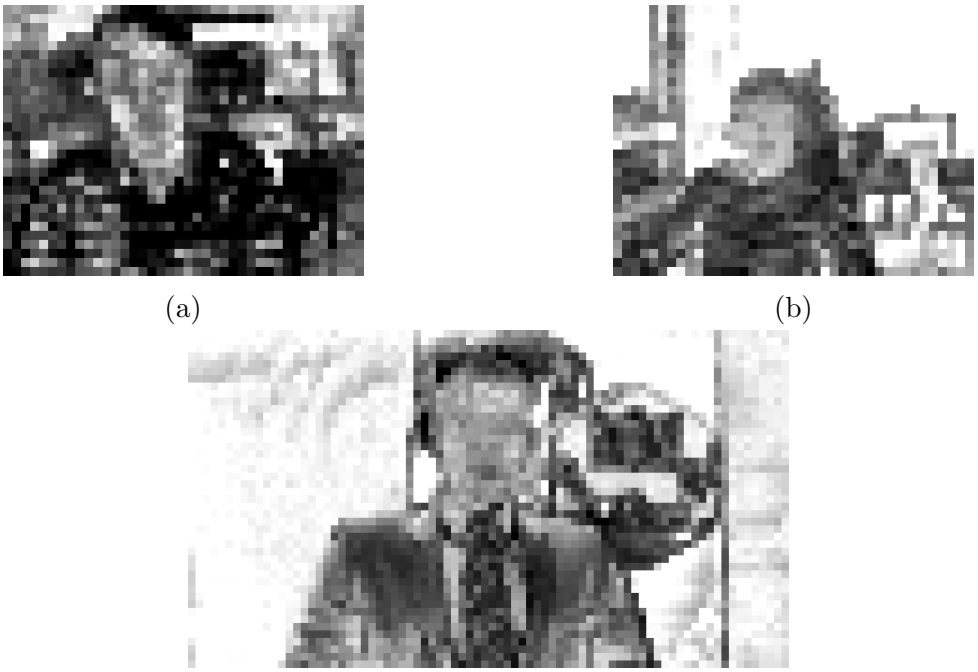
(a) (b)

(c)

Figure 7.11: Result of ROI-based bit allocation encoding (a)chet640x480, (b)Paul640x480 ,(c)Johnny1280x720.

## Disadvantages

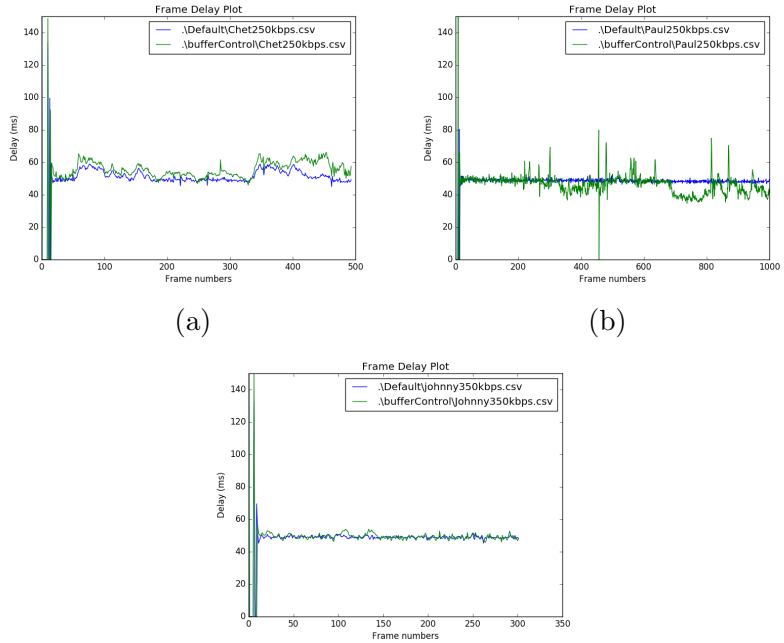
- This approach requires changes to the bit-allocation module and hence more complex to implement.
- The determination of relative complexity factor ($C_R$) needs trained models to predict bit-consumption of ROI and non-ROI parts during normal encoding without using ROI information. This prediction model has bad accuracy for macroblocks encoded with inter prediction.



(a) (b)

(c)

Figure 7.12: Result of ROI-based bit allocation encoding (a)chet640x480, (b)Paul640x480 ,(c)Johnny1280x720.




(a) (b)

(c)

Figure 7.13: Result of ROI-based bit allocation encoding (a)chet640x480, (b)Paul640x480 ,(c)Johnny1280x720.

# Chapter 8

# Zusammenfassung

Am Schluß werden noch einmal alle wesentlichen Ergebnisse zusammengefaßt. Hier können auch gemachte Erfahrungen beschrieben werden. Am Ende der Zusammenfassung kann auch ein Ausblick folgen, der die zukünftige Entwicklung der behandelten Thematik aus der Sicht des Autors darstellt.

# List of Figures

# List of Tables

# Bibliography


[AC05] T. Ebrahimi. A. Cavallaro, O. Steiger. Semantic video analysis for adaptive content delivery and automatic description. *IEEE Trans. Circuits Syst. Video Technol.*, 15(10):1200–1209, 2005.

[FL16] Na Li. Fan Li. Region-of-interest based rate control algorithm for H.264/ AVC video coding. *Multimed Tools Appl*, 75:4163–4186, 2016.

[GLW12] S.-Y. Chien. G.-L. Wu, Y.-J. Fu. Region-based perceptual quality regulable bit allocation and rate control for video coding applications. *VCIP*, 2012.

[HM11] Jorn Ostermann. Holger Meuel, Marco Munderloh. Low Bit Rate ROI Based Video Coding for HDTV Aerial Surveillance Video Sequences. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition - Workshops (CVPRW)*, pages 13–20, 2011.

[HM16] Jorn Ostermann. Holger Meuel, Florian Kluger. Codec independent region of interest video coding using a joint pre- and postprocessing framework. *IEEE International Conference on Multimedia and Expo (ICME)*, 2016.

[JOGJSW12] Thiow Keng Tan J. Ohm. G. J. Sullivan, H. Schwarz and T. Wiegand. Comparison of the Coding Efficiency of Video Coding Standards-Including High Efficiency Video Coding (HEVC). *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12), 2012.

[Lea] `http://www.eenadupratibha.net/pratibha/engineering/content_three_tra_layer_u6.html`.

[LT05] K.R. Rao. Lin Tong. Region of interest based H.263 compatible codec and its rate control for low bit rate video conferencing. *Intelligent Signal Processing and Communication Systems*, 2005.

[MB13] Manzur Murshed and James Brown. High Quality Region-of-Interest Coding for Video Conferencing based Remote General Practitioner Training. *The Fifth International Conference on eHealth, Telemedicine, and Social Medicine*, 2013.

[MX14] Shengxi Li. Mai Xu, Xin Deng. Region-of-Interest Based Conversational HEVC Coding with Hierarchical Perception Model of Face. *IEEE Journal of Selected Topics in Signal Processing*, 8(3), 2014.

[SL03] A. C. Bovik. S. Lee. Fast algorithms for foveated video processing. *IEEE Trans. Circuits Syst. Video Technol.*, 13(2):149–161, 2003.

[SML02] Yan Lu Siwei Ma, Wen Gao and Hanqing Lu. Proposed draft description of rate control on JVT standard. Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) 6h Meeting, 2002.

[TWL03] Gary J. Sullivan Thomas Wiegand and Ajay Luthra. Overview of the H.264/AVC Video Coding Standard. *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*, 13(7), 2003.

[Wan95] B. Wandell. Foundations of Vision. Sinauer, 1995.

[YLS08a] Z. G. Li Y. Liu and Y. C. Soh. A novel rate control scheme for low delay video communication of H.264/AVC standard. *IEEE Trans. Circuits Syst. Video Technol*, 17(1):68–78, 2008.

[YLS08b] Z. G. Li Y. Liu and Y. C. Soh. Region-of-Interest Based Resource Allocation for Conversational Video Communication of H.264/AVC. *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*, 18(1), 2008.

[ZWW11] Kexin Zhang Zongze Wu, Shengli Xie1 and Rong Wu. Rate Control in Video Coding. `http://www.intechopen.com/books/recent-advances-onvideo-coding/rate-control-in-video-coding`, 2011.