

Face Tracking for Optimized Bitrate Control in Low Delay Video Encoding

Chethan Ningaraju

September 6, 2016

1 Introduction

1.1 Low Delay Bitrate Control

In recent years there is increasing demand for high quality video conferencing solutions. To address this growing need there has been constant improvement in low delay video coding techniques. The need for extremely low end to end delay in video telephony puts additional constraints on video coding which results in compromise of video quality.

The bitrate control module is responsible for controlling the bit-consumption of the encoder to guarantee smooth playback. Bitrate control module is not codec specific and operates independent of any chosen codec. The main purpose of the bitrate control module is to ensure efficient playback of the encoded video. Figure 1 shows the functionality the bitrate control module. It achieves this by controlling the quantization parameter using during the encoding. The decision of quantization parameter is done considering the input bitrate, framerate, input complexity (spatial and temporal activity), acceptable input delay (a measure of VBV buffer size). The module also takes the feedback from the encoder regularly to make better decision. During the low delay video encoding, tools like bi-directional prediction are disabled.

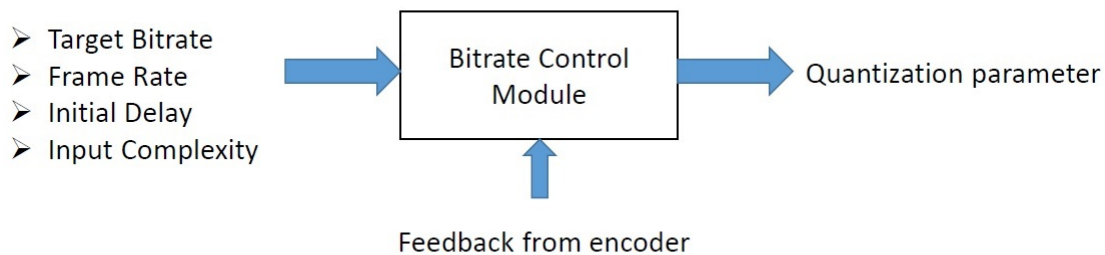


Figure 1: Bitrate Control Module Functionality

1.2 ROI based coding

In conventional video coding, all regions of the frame is considered equally important. It is assumed all regions contribute equally for perceptual quality. Some video codecs use the fact that high frequency components are less important to human visual system and do you preferential coding based on spatial frequency. However, such preferential coding does not take into account the contents the frame to be encoded.

Region of interest based coding is not a common practice in video coding because it is very hard to automatically detect important regions that contribute the most to perceptual quality. However, region of interest in video conferencing content is going to be face region predominantly. Due to recent improvements in face detection algorithms it is possible to detect face with good accuracy. This work aims to study possible ways of improving preceptual quality of the video by detecting face regions and coding it with higher quality than rest of the frame.

2 Operating Conditions

This work uses Citrix h264 video codec for the study. The encoder is configured in low delay mode sutiable for video conferencing. The encoder is configured to use IPPP mode with intra/key frames encoded only at the beginning of the sequence.