TECHNISCHE UNIVERSITÄT
ILMENAU

Faculty of Electrical Engineering and Information Technology
Institute for Media Technology
Audio Visual Technology

MASTER THESIS

Netflix Like Encoding Optimization

Submitted by: Vijaykumar Singh Rana

Major: Media Technology

Advisor: Prof. Dr.-Ing. Alexander Raake

Co-Advisor: M.Sc. Steve Göring

Ilmenau, May 13, 2019

Acknowledgments

optional

Zusammenfassung

maximum of 2400 chars; one paragraph

Abstract

maximum of 2400 chars; one paragraph

Contents

Chapter 1

Introduction

1.1 Background

In today's world, media is being offered to the users on a large scale. Numerous platforms like YouTube[], Netflix[], Prime Video from Amazon[] aim to provide the users with internet video on demand. These organizations offer television shows and movies for users to watch over the internet. These kind of organizations aims at providing the users with end product like a television show with highest possible quality depending on the users' affordable bandwidth. Providing such data requires various kinds of encoding schemes that can yield the data with best quality for a given bandwidth.

To achieve such task, the videos are not just encoded using conventional constant bit-rate video encoders. The organizations like Netflix offers the service where they provides the videos by employing various encoding schemes and further optimizing the process in order to obtain the high quality data with lowest bitrate. This then allows them to stream such high quality video with low bitrate and the end-user receives seamless playback of the content he receives. The reason behind achieving this is subjective. The end-user would not like the blurriness or blockiness because of the network bandwidth he affords. This could lead to users miss out their favourite TV shows. This requires the development of algorithms that could deal with the issues of arifacts in videos encoded with low bitrates.

This step also requires to tackle the challenges for encoding the video intelligently. The process of encoding these videos also have to be done for various bitrates. This is done because the mobile users may have fluctuating network which may then lead to varying bandwidth. This can then lead to videos getting stuck in buffering. From subjective point of view. This would be a very annoying experience for an end-user. As this whole process can become quite long or can reach to a bottleneck situation for the system and hence they have to be optimized based on various optimizing techniques. Netflix employs the use of their "Dynamic Optimizer" [] in order to optimize the whole process.

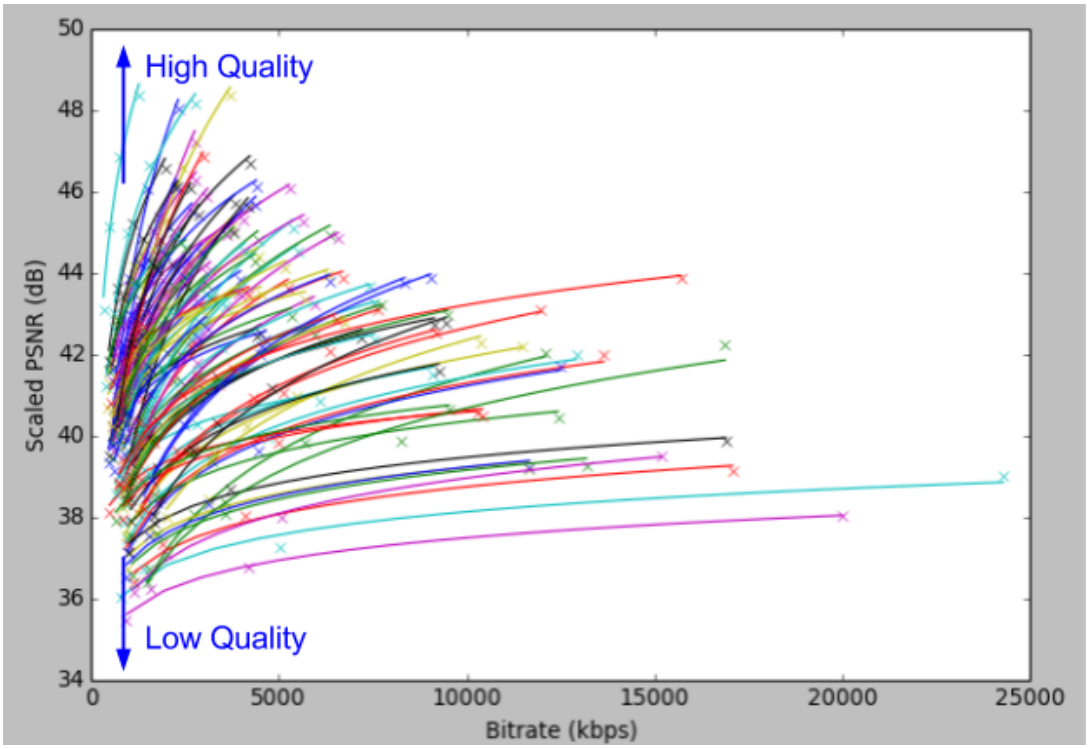

Figure 1.1: 100 randomly sampled sources at 1080p resolution using x264 constant QP (Quantization Parameter) rate control. [**per title**]

YouTube provides videos using approach where a video is segmented into multiple segments. Each of the segments are then encoded with different all the different qualities. When the video is streamed over the internet, the chunk to be transferred is decided upon the bandwidth the end-user can afford. This is achieved by DASH (Dynamic Adaptive Streaming over HTTP) []. This monitors the network bandwidth between server, providing the content to the client who receives it. Based on the bandwidth measured the client fetches the chunk with quality corresponding to the network bandwidth. This is handled from the perspective of networking. The objective here is to provide end-user the content with high quality which he can enjoy with limited bandwidth capabilities.

The content of a given video differs with other videos more or less. Applying a constant bitrate technique to each video does not make sense. Assuming a video that contains dynamic and constant scenes. In constant bitrate technique the dynamic scenes may utilize the bitrate efficiently but constant scenes where the background is not changing too much, the bitrate would be more than required thus wasting leading to over utilization of bits and the worst case scenario would be a scene where it is completely dark. So the bits are completely wasted for such scenes. So the modern approach is to use encoding per title based.[**per title**] The figure 1.1 demonstrates how the quality varies for different video sources. Certain sources never reach high quality even though the bitrate is high enough as 25000 kbps as shown in the figure. 1.1

Also the process of achieving best quality goes further into encoding each of chunks based on the scenes. A scene is a set of frames with similar attributes. A scene could be with fixed characters in it and a constant background or it could be a dynamic action scene. As each of the scenes have different characteristics so the idea is to encode them individually with different bitrates for different resolutions with highest possible quality.

The metrics like PSNR (Peak Signal to Noise Ratio), SSIM (Structural Similarity Metric), VMAF (Video Multi-Method Assessment Fusion) are used for video quality evaluation. The videos encoded for each quality with all possible bitrates are evaluated. Finally those bitrates are chosen which provides high quality amongst the different resolutions on a rate-distortion or rate-quality curve.

1.2 Motivation

As discussed in the previous section 1 about how each video is processed. This pipeline includes segmenting a video first. Next each of these segments are assessed and divided into scenes or shots. These shots are based on the type of content. These shots have relatively similar kind of frames which makes it easier to encode them with low bitrate compared to encoding a video with heterogeneous content. Netflix applies encoding with all possible bitrates for each resolution. **??** Then based on the quality parameters obtained from a rate distortion plot the appropriate bitrate is chosen for each resolution. Netflix applies a brute force technique in order to compute quality parameter for each bitrate for all the resolutions. This requires too much computation power. The idea is to minimize the computation cost and approximate this whole process using optimization techniques.

The goal is to investigate if it is possible to approximate the whole process by reducing computation steps. To have such comparison both the approaches will be implemented; brute force technique of computing quality parameters by computing quality for each bitrate for different resolutions to generate a rate-distortion curve and approximating quality parameters based on select bitrate point to generate a rate-distortion curve.

Chapter 2

Fundamentals

2.1 FFMPEG

FFMPEG is an open source cross platform tool which allows for handling audio and video data by the means on encoding, decoding, transcoding, multiplexing, demultiplexing etc. This tool allows for many options for customizing video data by the means of its capabilities. [Bel]

FFMPEG provides various command line tools, which is described in the table 2.1.

FFMPEG also provides libraries for various applications described in the table 2.2.

Command Tool	**Usage**
ffmpeg	is used for converting audio or video formats
ffplay	is a command line media player that uses Simple Direct-media Layer(SDL) and FFMPEG libraries
ffprobe	displays text based media information in formats like XML, CSV etc

Table 2.1: FFMPEG Command Line Tools

Libraries	Usage
libswresample	provides functions to resample audio
libavresample	provides functions to resample audio from Libav project [tea]
libavcodec	provides native FFmpeg audio/video encoders and decoders
libavformat	provides demuxers and muxers for audio/video container formats
libavutil	includes hash functions like SHA-1, LZO decompressor and Base64 encoder/decoder.
libpostproc	provides functions for old H.263 video postprocessing
libswscale	provides functions for video image scaling and colorspace/pixelformat conversion
libavfilter	is the substitute for vhook which allows the video/audio to be modified or examined between the decoder and the encoder

Table 2.2: FFMPEG Libraries

2.2 Video Quality Metric: Video Multi-Method Assessment Fusion (VMAF)

VMAF was developed by Netflix along with the external researchers which uses machine-learning approach which are fused with existing efficient objective video quality metrics called the elementary metrics.[KG18] [Blo] [Net]

VMAF is a full reference, perceptual video quality metric. Due to compression and scaling several artifacts result in the video. So VMAF uses the existing perceptual quality scores from multiple quality assessment algorithms to estimate the quality of a given video. This is fused with the support vector machine [Ras17]. As it is full referenced metric the original source and encoded/ compressed source both are taken in to consideration. For multiple encoded versions of a source video the comparison is done using a scaled VMAF score [KG18]. The figure 2.1 shows VMAF system diagram.

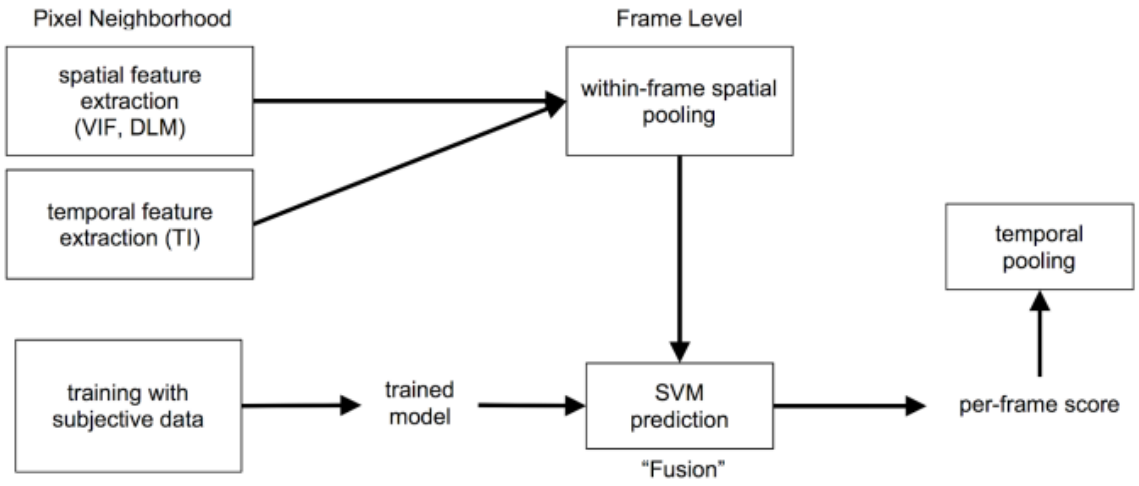

Figure 2.1: VMAF System Diagram [**per title**]

complexity based consistent quality encoding..... Fundamentals / environment and related work: 1/3

- ▷ comment on employed hardware and software
- ▷ describe methods and techniques that build the basis of your work
- ▷ review related work(!)

Chapter 3

Architecture/Implementation

Developed architecture / system design / implementation: 1/3

- ▷ start with a theoretical approach
- ▷ describe the developed system/algorithm/method from a high-level point of view
- ▷ go ahead in presenting your developments in more detail

Chapter 4

Analysis/ Evaluation

Measurement results / analysis / discussion: 1/3

- ▷ whatever you have done, you must comment it, compare it to other systems, evaluate it, using e.g. subjective tests
- ▷ usually, adequate graphs help to show the benefits of your approach
- ▷ caution: each result/graph must be discussed! what's the reason for this peak or why have you ovserved this effect

Chapter 5

Conclusion

Conclusion: 1 page

- ▷ summarize again what your thesis did, but now emphasize more the results, and comparisons
- ▷ write conclusions that can be drawn from the results found and the discussion presented in the paper
- ▷ future work (be very brief, explain what, but not much how)

Bibliography

[Bel] Fabrice Bellard.

[Blo] Netflix Technology Blog. *Toward A Practical Perceptual Video Quality Metric*. https://medium.com/netflix-techblog/toward-a-practical-perceptual-video-quality-metric-653f208b9652.

[KG18] Ioannis Katsavounidis and Liwei Guo. "Video codec comparison using the dynamic optimizer framework". In: *Applications of Digital Image Processing XLI*. Vol. 10752. International Society for Optics and Photonics. 2018, 107520Q.

[Net] Netflix. *Perceptual video quality assessment based on multi-method fusion*. https://github.com/Netflix/vmaf.

[Ras17] Reza Rassool. "VMAF reproducibility: Validating a perceptual practical video quality metric". In: *2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE. 2017, pp. 1–2.

[tea] Libav team. *Libav - Open source audio and video processing tools*. https://www.libav.org/.

List of Figures

List of Tables

Appendix A

AppendixExample

Declaration

I declare that the work is entirely my own and was produced with no assistance from third parties.
I certify that the work has not been submitted in the same or any similar form for assessment to any other examining body and all references, direct and indirect, are indicated as such and have been cited accordingly. Ilmenau,

_____

Todo list