Politecnico di Torino

Politecnico di Torino

Master degree program
in Cinema and Media Engineering

AY 2022-2023
March-April graduation session 2023

Master Thesis Summary

The Dynamic Optimizer Framework

Video encoding, assessment and comparison

Author
Chemin Davide

Supervisor
Masala Enrico

SUMMARY

In 2018, Netflix published a paper introducing a new and innovative video compression technique called Dynamic Optimizer. By that time, the technique had already been tested and used by the company internally for encoding their content. The Dynamic Optimizer is a tool that fine-tunes encoding and compression parameters for each shot of a video sequence with the aim of finding the best combination that meets a specific bitrate or quality target. This technique is compatible with all current and future video codecs and it is suitable in particular for non-real-time encoding of on-demand video content in adaptive streaming applications, which is an ideal scenario for the long sequences of shots typical of Netflix's catalogue. The Dynamic Optimizer takes advantage of three common tools in perceptual video processing, CRF, VMAF, and RD curves.

Constant Rate Factor (CRF) is used in dynamic optimization as the primary parameter for determining the compression level of each shot. That single parameter determines the trade-off between quality and bitrate and allows for efficient perceptual optimization using VMAF. Video Multi-Method Assessment Fusion (VMAF) is an objective video quality metric developed by Netflix. It is preferred over traditional metrics like Peak Signal-to-Noise Ratio (PSNR) because it uses the power of machine learning to take into account our human visual system, making it ideal for visual perceptual optimization. The roots of today's video optimization lie on the classical Rate-Distortion (RD) theory, which represents as a function, the RD curve, the fundamental tradeoff between the bitrate used to encode a video signal, and the distortion introduced when decoding.

According to the reference paper, the dynamic optimization process starts by splitting a long video sequence, a raw uncompressed video file, into single shots or coding units. Each shot gets encoded using multiple CRF values. For example, the H.264 encoder in FFmpeg supports CRF values ranging from 0 to 51, resulting in 52 possible versions of the same coding unit at different compressions. We refer to these encoded shots as elemental encodes, and their number is equal to the product of the number of coding units and the number of CRF values used for each unit. Each elemental encode is characterised by a specific bitrate and quality level. Therefore, its quality is assessed using the VMAF metric. Scores can be stored in a two-dimensional RD pair array and plotted, producing an RD curve. Next, a joint convex hull is constructed for the entire sequence, which selects only a subset of

points from each individual shot. Finally, the convex hull is utilised to generate the optimal coding path that maximises quality for a given target average bitrate, or conversely, minimises bitrate for a specific target average distortion.

As part of their presentation, Netflix explained the functionality of their Dynamic Optimizer, but did not provide any software implementation, and there are no publicly available algorithms up to now. To address this gap, this work was conducted with the goal of identifying the most effective video encoder and optimization strategy to achieve the optimal balance between quality and bitrate. The outcomes include a software implementation of a dynamic optimizer that will be released open-source and which includes three different alternatives optimization techniques.

The first one is the most straightforward, as it directly targets the optimal solution without relying on any approximations, by comparing all possible encoded solutions. For this reason, we refer to it as the brute force method, since it has an exponential complexity in terms of time, computational resources, and encoding. The second method is based on Lagrange theory. According to it, encoding parameters, in our case the CRF for each shot, can be optimised independently of each other. This can be done shot-by-shot, without considering the sequence as a whole, thus simplifying computations significantly. However, compared to the first approach, it returns a suboptimal solution, because it is not designed to consider all possible options. Nonetheless, studies have demonstrated that the Lagrangian solution is a reasonable approximation of the actual one when the points in the convex hull are dense enough. Both methods require the exhaustive encoding of all shots at different distortion levels, as only the points encoded before computation are part of the final solution. For this reason, we are proposing an alternative method, which utilises the lagrangian approach by integrating curve fitting. Curve fitting techniques are used to estimate and reconstruct the RD curve of the shot from a limited number of encoded points, hence allowing a great reduction in the number and the costs for encoding.

A comprehensive set of objective assessment tests were conducted to evaluate the performance of the three methods included into the framework. The tests provide a comparison between the optimised versions generated by the optimisation algorithm and the non-optimized versions produced by the encoding process itself, defined as Fixed CRF encoding. Performances and results turn out to be not as significant as those claimed in the reference paper by Netflix, mainly because test sequences used in the study do not fully represent the Netflix catalogue, and include for example resources with few shots. Nevertheless, albeit smaller,

presented results prove the efficiency of the developed system, leaving room for further optimization and improvements.

Regarding the implementation of the three methods, in all cases the non-optimized Fixed CRF versions performs the worst, followed by the Curve Fitting method which may be limited by its approximation level. The Lagrangian Method performs better than Curve Fitting when the number of encoded points is dense enough. Indeed, Curve Fitting should be preferred when the number of elemental encodes has to be low and encoding costs are limited. In any other case it is recommended to use Lagrange Method, as it is the most versatile and flexible, and because it is able to reach a suboptimal solution very close to the optimal one with a great reduction in the computational complexity. Finally, the BF method is the most accurate but due to its exponential complexity, it can only be used for testing purposes, as the reference for the optimal solution. By comparing the distance between the RD curves of the optimised sequences, it is estimated that Lagrangian Method achieves up to 8% gain in quality over the Fixed CRF, and Curve Fitting achieves up to 6% gain in quality over Fixed CRF. It should be noted that there are other factors besides the optimization method that affect the optimization performance, the main ones we identify are content complexity and used coding standards.