Video Encoding Cloud System for High Performance Scenarios.

Wilmer Moina-Rivera
Universitat de València
Burjassot, Valencia, Spain
wilmoiri@alumni.uv.es

Juan Gutiérrez-Aguado
Universitat de València
Burjassot, Valencia, Spain
juan.gutierrez@uv.es

Miguel García-Pineda
Universitat de València
Burjassot, Valencia, Spain
miguel.garcia-pineda@uv.es

ABSTRACT

Nowadays, there is a huge proliferation of multimedia content, with a wide range of services and applications serving video in many scenarios. The increase of device resolution and new formats such as 360 video demand new alternatives to optimize the preparation these multimedia contents.

With this premise, a new proposal for distributed video encoding in the cloud is being developed in this doctoral work. The aim is to develop an architecture and application to achieve a target quality of the content while reducing the number of bits and the required time to encoding a stream. In addition to making its processes adaptable to the aforementioned types of content.

A first prototype based on containers has been developed and the results show that we can reduce the size of the encoded file, maintaining significant quality levels that correspond to the Mean Opinion Scores (MOS) rating range, compared to a full-video coding scheme with fixed parameters.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

The Adjunct Proceedings of ACM TVX/IMX 2020, Barcelona (Spain), June 2020

© 2020 Copyright held by the owner/author(s).

https://doi.org/https://doi.org/10.6084/m9.figshare.12479348.v1

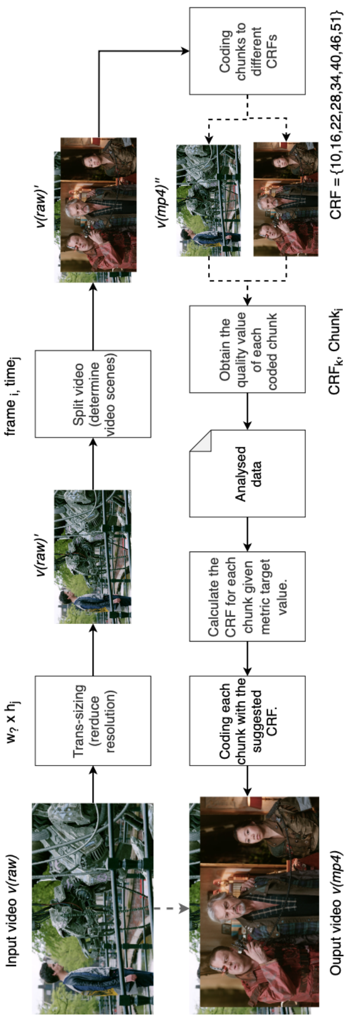

Figure 1: Structure to determine the CRF indicated by each scene in the video.

CCS CONCEPTS

• **Information systems** → **Multimedia content creation**.

KEYWORDS

QoE prediction, Cloud computing, Video coding, Video quality, Virtual reality, Shot detection.

ACM Reference Format:

Wilmer Moina-Rivera, Juan Gutiérrez-Aguado, and Miguel García-Pineda. . Video Encoding Cloud System for High Performance Scenarios.. In *Proceedings of The Adjunct Proceedings of ACM TVX/IMX 2020, Barcelona (Spain), June 2020. Copyright is held by the author/owner(s). (The Adjunct Proceedings of ACM TVX/IMX 2020).* ACM, New York, NY, USA, 4 pages. https://doi.org/https://doi.org/10.6084/m9.figshare.12479348.v1

INTRODUCTION

Nowadays, more than 80% of the Internet traffic is multimedia, mainly [1]. This type of traffic requires special conditions in order to give a good quality of experience (QoE) to the end users. Moreover, it is changing all the time due to the wide variety of services and applications such as live 360 streaming, virtual reality/augmented/mixed (VR/AR/MR), etc. This has led the chain of systems involved (OTT, IPTV, CDN, ISP, etc.) that provide these services, to seek new alternatives to offer quality content to the greatest number of participants with the least use of resources, while maintaining a properly user experience, without saturating the available communication channels.

One of the alternatives is to propose new strategies in encoding technologies, where from a video signal in raw format (yuv, y4m, etc.) as input is generated a compressed video signal as output. The output video can be characterized by its bit rate, frame rate, frame size or by the characteristics of different coding standards. Therefore, this contribution presents a new encoding strategy that accepts a video in raw format and aims to reduce the storage size of the encoded video and to reduce the processing time while maintaining a certain level of QoE. Reducing the processing time is important because pay-per-use systems encode the video in multiple combinations of resolutions and bitrates. The proposed system split the video in scenes and encodes each scene by selecting the coding parameters adaptively in order to achieve a target quality value. This ensures the quality of the encoded video and the user experience (see Figure 1). In addition, it can be deployed in the cloud, as each scene can be coded independently, taking advantage of its elasticity and pay-per-use capabilities.

The document is structured as follows, next section presents the work related to video encoding and transcoding over cloud environments. Thirdly, an overview of the system is given, from the segmentation of the video to the method used to obtain the parameters that will use to encode a scene. Fourth, preliminary results obtained are presented. Finally, conclusions and future work are presented in section five.

Table 1: Video sample settings

Parameter	Value
Name	Tears of Steel
Resolution	720p
Frame Rate	24 fps
Number of frames	17620
Format	RAW (y4m)
Available at:	https://mango.blender.org

Table 2: Video chunks characteristics and encoding settings

Parameter	Value
Number of scenes in video	168
Chroma subsampling	YUV420
Frame rate	24 fps
Resolution	Calculated width x 208
Encoder	FFmpeg
Encoding parameter	CRF (10,16,22,28,34,40,46,51)
Video compression standard	H.265/HEVC (libx265)
Preset	Medium (default)

RELATED WORK

Video coding has been a recurrent topic of study in recent years due to the continuous evolution of multimedia applications/services. Several proposals have raised the use of cloud computing to reduce encoding times. However, determining the coding parameters so that each segment of video processed in a distributed environment has the highest quality, the ability to adapt to the bandwidth of clients and the capabilities of viewing devices with the least use of resources, remains an open research topic.

Jan De Cock [3] proposes two alternatives in distributed video encoding algorithms in the cloud. The first one analyzes the complexity of the title to determine its average bitrate in a range of resolutions. The second, controls the bit rate of each video segment to maintain a constant quality, which would allow for a more efficient use of bandwidth and more consistent video quality. Xiangbo Li [4] presents a cloud-based service architecture for video-on-demand transcoding, with its focus on maintaining quality of service (QoS) in transmissions. Covell Michele [2] presents a neural network to predict the parameters of a model that relates bit rate to various properties of video. Its purpose, to determine a target bit rate for each video segment without requiring multi-step coding.

In this analysis we have observed that the common factor among the proposals is to maintain the highest quality in the videos. Performing two or three pass transcodings, training a neural network to predict the parameters to be used in the transcoding of the video segments; or the massive provisioning of resources in the cloud in order to maintain a consistent quality and user experience.

PROPOSAL TO DETERMINE A CRF PER SCENE.

In this section, the cloud video coding proposal is presented. As shown in Figure 1, the input is a video in RAW format, $v(\text{raw})$, to which a trans-sizing process (dimension reduction) is applied to obtain a video in the same RAW format, $v(\text{raw})'$. The goal is to reduce the processing time of subsequent steps. In this prototype, each scene is determined as the frames between two consecutive I encoded frames. Therefore, the $v(\text{raw})'$ is encoded and the position of the I frames are extracted. Inside each scene, frames are very similar in terms of spatio-temporal information. The next step encodes each $\{v(\text{raw})'_s\}$ scene, where s denotes the scene number, with 8 different CRF values to obtain $\{v(encoded)_s^{\text{CRF}_j}\}$ (see Table 2). This step can be performed in parallel in order to take advantage of benefits of cloud computing (elasticity, latency, pay-per-use, etc.),

For each scene, the quality of the encoded videos for the different encoded CRF values, is obtained. The quality metrics that have been used in this first exploration are: PSNR, SSIM and VMAF. Therefore, for each scene we have quality metric values for the different CRF values and we can perform an inverse interpolation to obtain the CRF value for the desired quality target value. After having characterized each segment with a CRF value, the scenes of the original RAW, $v(\text{raw})$, video are encoded in its original resolution using the CRF values obtained.

Table 3: Total time used to obtain the metrics for the 1344 encoded videos (168 segments encoded using 8 CRF values).

Metric	Time(min)
VMAF	5.5963
PSNR	1.2619
SSIM	1.3059

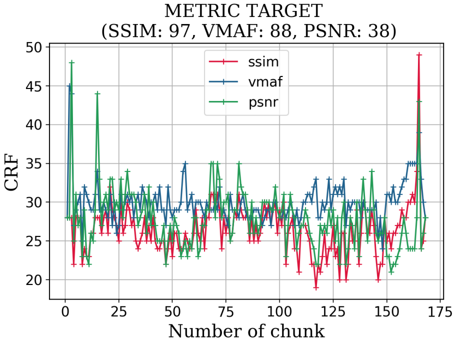

Figure 2: CRF values obtained for each metric.

Table 4: Size of the encoded video as a function of the target metric.

Metric	Target	Size(MB)
VMAF	88	63.54
PSNR	38	129.52
SSIM	97	124.95
CRF default libx265		85.54

EVALUATION METHODOLOGY

In this work, we used the video source presented in Table 1. Video processing tasks such as trans-sizing, encoding, quality analysis (VMAF, PSNR, SSIM, etc.) were done with FFmepg (https://ffmpeg.org). The only encoding parameter that has been adjusted at each scene is the CRF obtained from the previous step. The coding and configuration characteristics used in the segments can be seen in Table 2. The proposed coding system, which is in first phase, uses the H.265/HVEC coding standard with libx265 library. All the tools have been encapsulated in a container and at this point all the processing is done sequentially. The next step in our research will be to provide an suitable architecture to encode each scene in parallel in cloud infrastructures.

Table 3 shows the time needed to calculate the indicated CRF per scene with the VMAF, SSIM, PSNR metrics. In Figure 2 we can observe the recommendations of the CRF parameter for each scene according to each quality metric. Some correlation between them can be observed in the complex parts of the video (high spatio-temporal values), however, when using low CRF values to match the target value, the chunk will use more or less bits. This has a direct impact on the overall size of the video as can be seen in the Table 4.

CONCLUSION

In this article, we have presented a cloud-ready encoding system capable of using a metric value as the quality objective of the encoded video. We can see that if we use the VMAF metric as a target, the processing time is longer than the rest of the metrics (PSNR and SSIM). But we have an important improvement in the video size. In all tests we have selected a high quality target value and the time used by a metric to determine the CRF values per chunk may seem high, but this time will be reduced when the system runs in a distributed cloud environment. In future work, it is expected to extend the analysis to compression systems such as AV1 and VP9 codecs and in other services like VR360.

ACKNOWLEDGEMENTS

This research was supported by the Spanish Research and Development Grant RTI2018-098156-B-C55.

REFERENCES

[1] Cisco and/or its affiliates. 2020. *Cisco Annual Internet Report (2018–2023)*. Technical Report.

[2] Michele Covell, Martín Arjovsky, Yao Chung Lin, and Anil Kokaram. 2016. Optimizing transcoder quality targets using a neural network with an embedded bitrate model. In *IS and T International Symposium on Electronic Imaging Science and Technology*. Society for Imaging Science and Technology. https://doi.org/10.2352/ISSN.2470-1173.2016.2.VIPC-237

[3] J. De Cock, Z. Li, M. Manohara, and A. Aaron. 2016. Complexity-based consistent-quality encoding in the cloud. In *2016 IEEE International Conference on Image Processing (ICIP)*. 1484–1488. https://doi.org/10.1109/ICIP.2016.7532605

[4] X. Li, M. A. Salehi, M. Bayoumi, and R. Buyya. 2016. CVSS: A Cost-Efficient and QoS-Aware Video Streaming Using Cloud Services. In *2016 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)*. 106–115. https://doi.org/10.1109/CCGrid.2016.49