Quality-Constant Per-Shot Encoding by Two-Pass Learning-based Rate Factor Prediction

Chunlei Cai, Yi Wang, Xiaobo Li, Tianxiao Ye
Bilibili Inc., Shanghai, China
{caichunlei, wangyi, lixiaobo, yetianxiao}@bilibili.com

Abstract—Providing quality-constant streams can simultaneously guarantee user experience and prevent wasting bit-rate. In this paper, we propose a novel deep learning based two-pass encoder parameter prediction framework to decide rate factor (RF), with which encoder can output streams with constant quality. In first-pass, an RF is predicted based on spatial-temporal and pre-coding features of video segment. Then video segment is encoded using the predicted RF and then its VMAF is measured. If first pass VMAF doesn’t meet target quality, a second pass prediction is performed using another model, in which results of first pass is added to features. Experiments show the proposed method requires only 1.55 times encoding complexity on average, meanwhile the accuracy, that the compressed video’s actual VMAF is within ±1 around the target VMAF, reaches 98.88%. Compared with average rate mode, this method can both improve visual quality and save ~10% bit-rate, as shown in demos.

Index Terms—Video Encoding, Quality-Constant, Deep Learning, Per-Shot Encoding, VMAF

I. INTRODUCTION

Compared with other methods, such as average rate mode, quality-constant compression can simultaneously guarantee quality and prevent wasting bit-rate as illustrated in Fig. 1.

![Fig. 1. Per-shot quality constant coding mode is adaptive to video scene’s complexity. It can both guarantee quality and prevent wasting bit-rate.](image)

Realizing quality-constant compression requires accurately deciding appropriate coding parameters which is corresponding to a target quality [1]. As shown in Fig. 2, this paper proposes a novel deep learning based two-pass prediction framework to decide Rate Factor (RF), with which an encoder can output streams with required quality at high accuracy.

![Fig. 2. The proposed deep learning based two pass RF prediction framework, which reaches 98.88% accuracy to compress one-shot video segment at target quality at only 1.55 times of encoding complexity on average.](image)

II. TWO-PASS RATE FACTOR PREDICTION

The proposed method follows per-shot coding framework. For each one-shot segment in a video, we firstly extract spatial (Gray-level Co-occurrence Matrix), temporal (Normalized correlation Coefficient) and pre-coding features by an ultra fast pre-process. Based on these features, an RF parameter is predicted by a deep neural network. Video encoder uses the RF to compress segment as the first encoding pass. Then VMAF quality of the first pass encoding is measured. If the quality doesn’t meet target, a second pass prediction and coding will be performed. With the help of first pass predicted RF and corresponding actual quality as effective feedback, the second pass prediction will be highly accurate.

Deep networks within this framework are trained on 200 thousands videos collected from bilibili.com. Ground truth RF label of each one-shot scene segment is generated by performing several times encoding and finding the best one which VMAF equals to a quality target, such as VMAF_target = 91.

III. EXPERIMENTS

We build a comparative average rate encoding method, which encodes one whole video using one fixed target rate. At this rate, the averaged VMAF is controlled around 91, but each one-shot scene’s quality is fluctuating. The results on 10 thousands test videos are shown as Table I, which verifies the effectiveness of the proposed two-pass prediction method. Moreover, deploying this method in our H.265 based cloud transcoding system has helped us save about 10% bit-rate.

![Table I](image)

REFERENCES

[1] H. Xing, Z. Zhou, J. Wang, H. Shen, D. He, and F. Li, “Predicting rate control target through a learning based content adaptive model,” in 2019 Picture Coding Symposium (PCS). IEEE, 2019, pp. 1–5.

Subjective comparison videos can be downloaded at https://drive.google.com/drive/folders/1BJNZ5HssxFaKbcXMenEdiolJ77hnZ-vx?usp=sharing.