Eye Controlled Region of Interest HEVC Encoding

Joose Sainio, Arttu Ylä-Outinen, Marko Viitanen, Jarno Vanne, Timo D. Hämäläinen
Laboratory of Pervasive Computing
Tampere University of Technology
Tampere, Finland
{joose.sainio, arttu.yla-outinen, marko.viitanen, jarno.vanne, timo.d.hamalainen}@tut.fi

***Abstract*—This paper presents a demonstrator setup for real-time HEVC encoding with a gaze-based region of interest (ROI) detection. This proof-of-concept system is built on Kvazaar open-source HEVC encoder and Pupil eye tracking glasses. The gaze data is used to extract the ROI from live video and the ROI is encoded with higher quality than non-ROI regions. This demonstration illustrates that performing HEVC encoding with non-uniform quality reduces bit rate by 40-90% and complexity by 10-35% over that of the conventional approaches with negligible to minor deterioration in perceived quality.**

Keywords—high efficiency video coding (HEVC), region of interest (ROI), eye tracking, Kvazaar HEVC encoder, open source

I. Introduction

High Efficiency Video Coding (*HEVC/H.265*) [1] is currently the state-of-the-art video coding standard. It is targeted to reduce bit rate by 40% over the preceding AVC/H.264 standard for the same objective visual quality [2]. Further bit rate savings can be obtained with *Region of Interest* (*ROI*) coding that seeks to improve the perceived quality at the same bit rate, or reduce the bit rate with the same quality by encoding the ROI with higher quality than its surroundings [3]. Moving from uniform to non-uniform quality coding calls for appropriate ROI based detection schemes. In this work, the ROI is extracted with eye tracking.

Eye tracking glasses usually consists of a scene camera and one or two eye cameras. The eye cameras are used to detect the direction of the gaze, which is then transformed into coordinates relative to the image captured by the scene camera. Currently, there are multiple providers for eye tracking glasses on the market such as Tobii [4], SMI [5], and Pupil Labs [6]. For this demonstration, the *Pupil Mobile Eye Tracking Headset* (*PMETHS*) with binocular eye cameras is selected because of its competitive price and open-source software stack.

Methods to achieve non-uniform video quality on a single video frame can be categorized into pre-processing and in-encoder processing. The most common pre-processing approach is to apply a non-uniform strength Gaussian blur to the image before encoding [3]. However, the effects of pre-processing are limited so it is not used very often.

The existing approaches for in-encoder processing commonly implement non-uniform quality with a custom rate control or non-uniform *delta quantization parameter* (*DQP*) maps that increment the ordinary QP value as a function of the ROI across the video frame [3]. Rate control methods are only suitable for scenarios where a constant bit rate is the goal whereas the DQP map approach will provide a variable bit rate but more constant visual quality. For this demonstration, DQP maps are used because a constant bit rate is not required and the implementation is more straightforward.

For HEVC encoding, there are two real-time capable open-source solutions: x265 [7] and Kvazaar [8], [9]. x265 is probably the best-known HEVC encoder, but unlike Kvazaar, it does not support DQP maps. Therefore, Kvazaar is chosen for this demonstration.

For the time being, a couple of gaze-based solutions for ROI coding have been presented, but the existing approaches are focused on pre-encoded video [10], make use of older standards such as H.263 [11], or are limited to specific conditions due to a custom compression method [12]. To the best of our knowledge, this is the first open-source work on real-time HEVC encoding with gaze-based ROI detection.

II. Heat Map Generation for ROI

In this work, the DQP maps are generated by calculating a DQP offset for each *coding tree unit* (*CTU*) [1]. The calculation of the offset is based on the logarithmic distance from the center of the CTU to the *gaze center* (*GC*). Logarithmic degradation of quality was chosen instead of linear because it better matches the human visual system and thereby results in higher perceived quality [13]. The steepness of the DQP increases together with the distance from the GC. The slope can be adjusted by changing the *degradation coefficient* (*DC*). Fig. 1 illustrates how the DC affects the bit rate of the encoded video with a base QP value of 27. Zero DC is equivalent to a video encoded with uniform QP.

The eye cameras in the PMETHS operate at four times the frequency of the scene camera so multiple gaze points are mapped to a single frame. In order to minimize the latency, each gaze point is mapped to the following frame instead of the nearest one. Gaze points that cannot be mapped with enough confidence are discarded. The GC equals the mean of the gaze points. If there are no valid gaze points for a frame, e.g., when the user blinks, the GC from the previous frame is used.

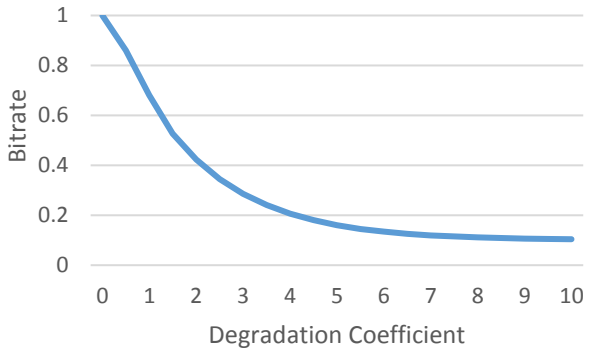

Fig 1. Bit rate of the encoded video relative to DC.

This work was supported in part by the European Celtic-Plus Project 4KREPROSYS and the Academy of Finland (decision no. 301820)

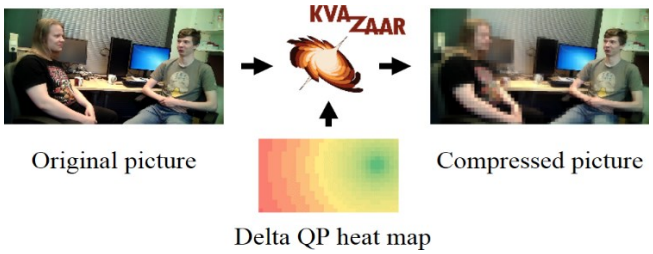

Fig 2. Application of DQP map to a frame.

Fig. 2 depicts how the DQP map is applied to the frame during encoding. On the green area of the heat map, the DQP value is lower and it grows towards the red area.

III. Demonstration Setup

Fig. 3 depicts the individual components of the demonstrator and the data flow between the components. The PMETHS worn by the user is fitted with a 60 degree field-of-view lens. The scene camera of the headset provides 1080p30 YUV422 video that is compressed with *Motion JPEG* (*MJPEG*) on the camera hardware.

The cameras are connected to a computer running the Pupil capture software, which detects the gaze of the user. The YUV422 video is down-sampled into YUV420 format using linear filtering before it is fed into Kvazaar for encoding, along with the gaze data in *JavaScript Object Notation* (*JSON*). The data is transferred through a ZeroMQ socket.

Kvazaar ultrafast preset [8] and a base QP value of 27 at the GC are used in encoding. DC is set to seven to illustrate a noticeable degradation of quality around the GC. Kvazaar writes the GC for each frame to a custom HEVC SEI message.

The encoded video is piped to an FFmpeg instance, which encapsulates it in an MPEG-2 TS container and streams it to a separate laptop over an Ethernet cable. On the laptop, the stream is decoded and displayed by another FFmpeg instance. The GCs reported in the SEI messages are visualized by drawing a red dot with a custom FFmpeg filter.

Compared with traditional HEVC encoding, the proposed system achieves 40-90% reduction in bit rate and 10-35% reduction in complexity with negligible to minor reduction in perceived quality.

Acknowledgment

The authors would like to thank all contributors of Kvazaar open-source project [8].

IV. References

[1] High Efficiency Video Coding, document ITU-T Rec. H.265 and ISO/IEC 23008-2 (HEVC), ITU-T and ISO/IEC, Apr. 2013.

[2] Advanced Video Coding for Generic Audiovisual Services, document ITU-T Rec. H.264 and ISO/IEC 14496-10 (AVC), ITU-T and ISO/IEC, Mar. 2009.

[3] J. Lee and T. Ebrahimi, "Perceptual video compression: A survey," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 6, Oct. 2012, pp. 684-697.

[4] *Tobii* [Online]. Available: https://www.tobii.com

[5] *SMI* [Online]. Available: https://www.smivision.com

[6] *Pupil Labs* [Online]. Available: https://pupil-labs.com/pupil

[7] *x265* [Online]. Available: http://x265.org

[8] *Kvazaar HEVC encoder* [Online]. Available: https://github.com/ultravideo/kvazaar

[9] M. Viitanen, A. Koivula, A. Lemmetti, A. Ylä-Outinen, J. Vanne, and T. D. Hämäläinen, "Kvazaar: open-source HEVC/H.265 encoder," *in Proc. ACM Int. Conf. Multimedia*, Amsterdam, The Netherlands, Oct. 2016.

[10] S. Arndt and J. N. Antons, "Enhancing Video Streaming Using Real-Time Gaze Tracking," *in Proc. ISCA/DEGA Workshop on Perceptual Quality of Systems*, Berlin, Germany, Aug. 2016

[11] S. Lee and A. C. Bovik, "Fast algorithms for foveated video processing," *IEEE Trans. Circuits Syst. for Video Technol.*, vol. 13, no. 2, Feb. 2003, pp. 149-162.

[12] D. Pohl, D. Jungmann, B. Taudul, R. Membarth, H. Hariharan, T. Herfet, and O. Grau, "The next generation of in-home streaming: Light fields, 5K, 10 GbE, and foveated compression," *in Proc. Federated Conf. Computer Science and Information Systems*, Praque, Czech Rebuplic, Sep. 2017.

[13] B. Ciubotaru, G. Ghinea, and G. M. Muntean, "Subjective assessment of region of interest-aware adaptive multimedia streaming quality," *IEEE Trans. Broadcast.*, vol. 60, no. 1, pp. 50-60, Mar. 2014.

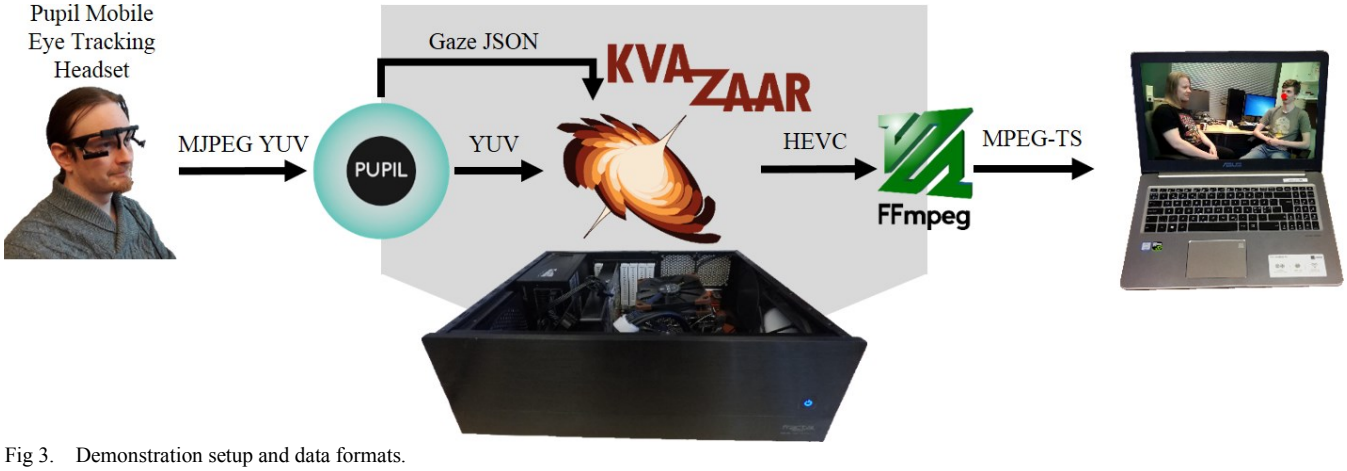

Fig 3. Demonstration setup and data formats.