2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)

Foveated Video Coding for Real-Time Streaming Applications

Oliver Wiedemann, Vlad Hosu, Hanhe Lin and Dietmar Saupe
Department of Computer and Information Science
University of Konstanz, Germany
Email: {oliver.wiedemann, vlad.hosu, hanhe.lin, dietmar.saupe}@uni-konstanz.de

***Abstract*—Video streaming under real-time constraints is an increasingly widespread application. Many recent video encoders are unsuitable for this scenario due to theoretical limitations or run time requirements. In this paper, we present a framework for the perceptual evaluation of foveated video coding schemes. Foveation describes the process of adapting a visual stimulus according to the acuity of the human eye. In contrast to traditional region-of-interest coding, where certain areas are statically encoded at a higher quality, we utilize feedback from an eye-tracker to spatially steer the bit allocation scheme in real-time. We evaluate the performance of an H.264 based foveated coding scheme in a lab environment by comparing the bitrates at the point of just noticeable distortion (JND). Furthermore, we identify perceptually optimal codec parameterizations. In our trials, we achieve an average bitrate savings of 63.24% at the JND in comparison to the unfoveated baseline.**

I. Introduction

Video streaming is ubiquitous and imposes ever-growing demands on content and network providers [1]. Increasing resolutions coupled with bandwidth limitations motivate ongoing research on sophisticated video codecs and compression algorithms. In this paper, we are concerned with the emerging subclass of real-time video streaming applications. Besides transmitting and rendering a video sufficiently fast, these additionally require to encode it under strict latency constraints.

Video codecs rely on the analysis and exploitation of correlations between pixel color values to achieve high visual quality at low bitrates. Recently proposed approaches combine a multitude of incremental and often marginal improvements to coding techniques [2], on which real-time requirements impose additional feasibility criteria. On the one hand, coding methods have practical limitations, which can be solved often by fast, hardware-supported implementations of certain mathematical operations. On the other hand, theoretical requirements enforce strict constraints on methodology choices. For example, bi-directional inter-prediction schemes that use past and future keyframes during interpolation, are usually not applicable in real-time scenarios, in particular when the required temporal buffering already violates latency constraints.

In addition to the conventional approach of improving the coding efficiency over the entire frame, region of interest (ROI) coding allows focusing on rendering parts of the frame that are more relevant to the viewer at higher quality settings. However, conventional video applications allow making only a narrow range of assumptions about the spatial relevance of the content within a given frame. This limits the potential improvement due to ROI-coding. The limitation stems primarily from not knowing a priori what region an observer will be interested in. Content-based predictions of ROIs are generally speculative and inaccurate.

We investigate possible improvements for video coding by adding gaze information to the encoding process by using an eye-tracker. We expect this approach to be much more accessible in the near future due to the prevalence of large high-resolution screens in combination with the advent of eye-tracking devices in consumer hardware. The approach allows us to make strong assumptions about the active region of interest in each frame, devise an appropriate encoding scheme that boosts the quality around gaze fixation-points, and design subjective experiments to study the performance of our software framework. Our approach goes beyond the traditional concept of ROI coding by gently degrading the quality of the video outside the ROI. This is the process of foveation, which adapts the video rendering according to the acuity of an observer's eyes, in real-time. Possible application scenarios include video telephony and streaming, and novel technologies that could be bolstered by our approach, e.g., human assistance in steering semi-autonomous vehicles, medical or industrial robots and drones, and more.

In summary, we present a modular software for the assessment of foveated video coding schemes, including a reference implementation based on x264 [3]. Furthermore, we evaluate the performance of our approach in a lab study and quantify the bitrate gains at the point of just noticeable distortion. Our experiments show that our foveated codec achieves a 63.24% average bitrate savings in comparison to the unfoveated baseline.

II. Related Work

The relevant literature on this subject matter has broad coverage, ranging from psychophysical aspects of human perception to contributions in the image and video coding domain. We need to understand the limitations of human vision, with regard to attended regions during video playback, in order to devise an efficient foveated codec.

The core assumption of our work is centered on the concept of foveation in human vision. Due to the limited bandwidth

978-1-7281-5965-2/20/$31.00 ©2020 IEEE

of the optic nerve, the human eye encodes spatial information coming from the environment unevenly. The fovea, the central piece of the retina, has the highest visual acuity, whereas the acuity rapidly drops towards the periphery. An exhaustive depiction of the human visual system (HVS) is found in [4], of which the chapter on the retinal representation is seminal to this work. Further insight into the distribution of retinal cells, the foveal and peripheral density of cone receptors, is given in [5]; the latter also broaches on implications for the resolution capacity of the human eye from a theoretical perspective.

As the human eye perceives less information from the visual periphery, videos do not need to render all details in such regions. Adapting a visual stimulus relative to the acuity of the human eye is not a novel idea. Girod [6] discussed compression algorithms that exploit this peculiarity already in the late 80s. He considered the "usefulness of the approach [as] limited" due to encoding and transmission latencies, which may not be the case anymore three decades later. Wang et al. published an approach to a region of interest (ROI) coding in the context of embedded zerotree wavelet coding [7]. This methodology has seen successful applications for still images already [8]. When reviewing older contributions, one has to carefully distinguish between ROI coding and foveation, as the nomenclature is partially overlapping and ambiguous. For the purpose of this paper, foveation describes the process of adapting a visual stimulus according to the acuity of an observer's eyes, in real-time.

Foveation is currently popular in the domain of virtual reality [9]. The approach generally benefits from increasing screen sizes, as these allow to present a larger area in the peripheral regions of an observer's field of view.

The work of Illahi et al. [10] is closely related to ours. They present a streaming framework that utilizes foveation in the context of cloud gaming and evaluate its performance for different genres of video games. We aim to investigate a similar approach, but for the live streaming of natural videos. Thereby we hope to reduce subjective influences due to the less interactive nature of videos in comparison to computer games.

Arndt et al. [11] present a study on the perceptual impact of overlaying a low quality background stream with suitably cropped parts of a high quality video. However, the authors didn't address the difficulties imposed by compression: the cropping is happening on the client side, which has the HQ video readily available. There is also no mention of the influence on the bitrate, which is one of major reasons such a coding procedure would be applied in practice.

III. The `FFoveated` Software Suite

The general idea of a software suite for subjective quality assessment of foveated video opens many engineering questions. Video data is seldom stored raw, due to the sheer file size, but instead as a compressed stream, together with audio tracks and possibly other metadata in multiplexed container formats. Furthermore, the goal of this project is not the development of a novel codec per se, but the extension, utilization, and evaluation of existing methods. We aim for the ability to interface data formats and codec implementations in a flexible way. In an effort to keep our program extensible, we hence decided to build upon FFmpeg's [12] library collection. The source code of our implementation, which we call `FFoveated`, will be released to the community.

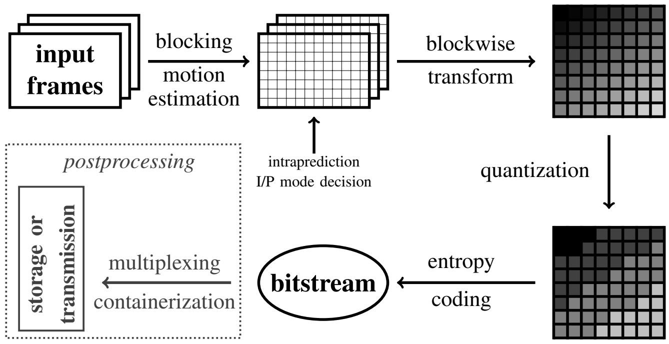

Fig. 1. General schematic of a transform-domain video encoder.

From a coarse perspective, `FFoveated` provides functionality to re-encode and display a given video in a foveated fashion. For this purpose, an eye-tracker permanently updates information about the user's gaze. `FFoveated` is codec-agnostic in the sense that it passes all required information to the wrapper in `libavcodec` on a per-frame basis, where implementation-dependent handling has to happen. This is provided by a patch for FFmpeg, that allows utilizing the functionality provided around the `AVFrameSideData` struct. In order to reduce the intrinsic latency of the pipeline to a minimum, we implemented the reading and demultiplexing of a container file, decoding of the source video, foveated re-encoding, decoding of the foveated video and finally displaying the frames in independent threads. We utilize SDL [13] to keep rendering of `AVFrames` to the screen as well as event handling abstract and platform-independent. Our implementation was tested on Linux and MinGW [14].

A. Codec Implementation: Foveation Using `x264`

Following the codec-independent part, we will now describe how to steer the compression rate for a specific codec spatially. For this purpose, we chose x264 [3], an established implementation conforming to the H.264 [15] standard. The modus operandi of a lossy transform-domain encoder is outlined in Fig. 1. A large share of the bitrate reduction is realized through block quantization, which maps the frequency coefficients from a continuous domain to a discrete set of possible values. In the case of x264, the granularity of this process is governed by a quantization parameter qp, which ranges from 0 to 51. A higher qp results in a coarser quantization, thus a lower bitrate and reduced visual quality. This value is usually determined for each block through content-based heuristics, which assign, for instance, a lower bit budget to blocks with large motion vector magnitudes, as the eye is less sensitive to details in fast-

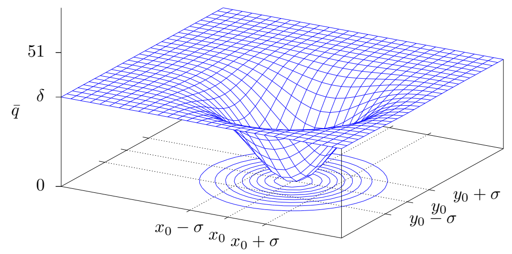

Fig. 2. Spatial distribution of the quality offset function $\bar{q}$.

moving parts of a video. We incorporate eye-tracking data by supplying an offset $\bar{q}$ to qp.

$$\bar{q}(x, y) = \delta\left(1 - \exp\left(-\frac{(x - x_0)^2 + (y - y_0)^2}{2\sigma^2}\right)\right) \quad (1)$$

This offset can be calculated for a macroblock at position (x, y) given a fixation point (x_0, y_0), a standard deviation σ and a scale factor δ. A two dimensional Gaussian is a natural choice to model the spatial quality distribution given the circular shape of the fovea centralis [4]. As depicted in Fig. 2, our function of choice does not inflict a quantization penalty on the fixation point at (x_0, y_0), where $\bar{q} = 0$.

Since the acuity of the human eye sharply drops around the fovea centralis [16], we chose to set σ to $2.5°$ of visual angle. The choice is based on the characteristics of the retina, but still an approximation, as the perceptual impact of $\bar{q}$ is content dependent. Under these premises, the question is how to choose δ such that the bitrate is minimized without introducing overly-disturbing artifacts in the peripheral regions.

We adapted the default settings of x264 in the following way. The group of picture (GOP) size limit is set to only three frames in comparison to the default of 250. This enforces frequent I-frames, which prevents quantization artifacts from remaining visible for an extended period of time in the event of saccades. Our strategy is similar to the one used by Lungaro et al. [17], who experimented with frequent I-frames in the context of quality switching problems occurring in head mounted virtual reality streaming. Additionally, it serves as mitigation to packet-loss-induced stalling during network streaming. We are aware that this choice influences the visual quality in the peripheral regions, as it increases the frequency at which flickering at GOP transitions occurs. The effect on perceptual quality is probably positive. This is a tradeoff, in which the bitrate penalty for I frames has to be balanced with the perceptual effects of larger GOPs.

We chose the *ultrafast* encoder preset and enabled the *zerolatency* tuning option. Adaptive quantization has to be enabled to utilize foveation, and we set the *variance* aq-mode. Notably, this implementation does not break the H.264 specification and that the resulting videos can hence be displayed with any standard-conforming player.

IV. Perceptual Quality in Foveated Coding

The human eye with the complementary neuro-visual system is the final recipient of visual media. Therefore, performance improvements in coding have to be measured in terms of bitrate gains relative to the impact on subjectively perceived quality. Automated assessment of image and video quality is an active field of research. Current developments in this domain are focused on the creation of databases, e.g., [18]–[22] and the utilization of machine-learning algorithms [23]–[26] to predict subjective scores for images and videos. While the challenges in this field are far from being solved for traditional video codecs, foveation adds yet another layer of complexity to the problem. The deliberate introduction of degradations in certain regions of a video requires a spatial weighing of local quality scores with regard to their significance for the observer.

To a certain degree, foveation reduces the repeatability of quality assessment studies, as it introduces a feedback loop that modifies the media item under inspection. In contrast to non-foveated coding, it is not currently possible to run large-scale studies on crowdsourcing platforms, as eye-tracking hardware is not yet common enough in consumer hardware. Therefore, the practical feasibility of lab studies is another factor when choosing a study methodology.

The notion of just noticeable difference [27] is relevant in the context of comparative multi-stimuli experiments, but can be adapted to single-stimulus scenarios in the sense of a *just noticeable distortion* (JND) [28]. A study participant is required to indicate whether noticeable distortions are present in a displayed video. This approach is promising for the quality assessment of foveated videos in general, and the definition of a just-noticeable maximal quality difference δ in particular.

We hypothesize that our algorithm performs well for scenes where the predominant salient region is confined to a small area surrounded by a tranquil background. However, videos with regions in an observer's peripheral vision that suddenly attract attention might pose a challenge, as frequent saccadic eye movement over larger spatial distances could expose severe compression artifacts to the observer.

V. Subjective Study

We conducted a subjective study in a controlled lab environment in order to evaluate the performance of our implementation. For this purpose, we installed a color-calibrated HP Z31x screen in a room with solely artificial illumination to avoid inconsistencies caused by daylight changes. Gaze data was gathered with an SMI Red250mobile eye-tracker.

We utilized the undistorted source files taken from the VQEG JEG Hybrid [29] dataset for our experiments. The 10 videos have a length of 10 seconds each and have been recorded at a resolution of 1920×1080 pixels with a framerate of 25 fps. The dataset was originally intended for the development of video quality assessment algorithms but provided

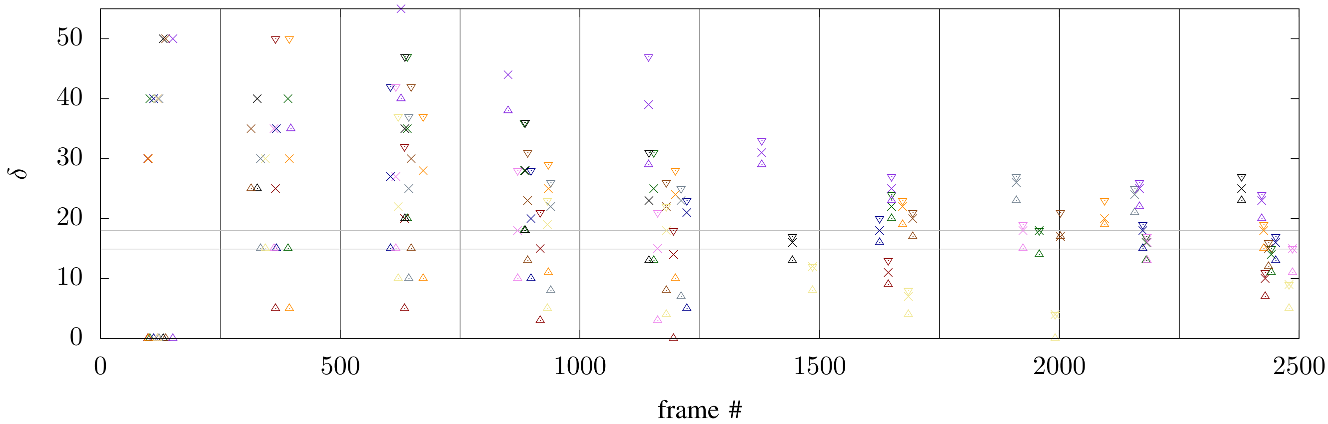

Fig. 3. Interactions for source #3: The horizontal axis enumerates all possibly displayed frames. Each vertical bar indicates the beginning of a new repetition. Colors encode unique participants. An $\times$-marker is placed for each *interaction*. Upward and downward triangles denote the δ-range per user, from the lowest to the highest possible value within a repetition. The grey lines indicate the .25 and the .1 JND for this source over all participants and repetitions.

a sufficiently suitable test set for our application scenario. It contains diverse scene types that might affect gaze behavior in various ways, thus exposing the benefits and shortcomings of our implementation.

The experimental procedure is as follows. After an introduction to the task and the required setup of the eye-tracker, we present each source video for a total of 10 *repetitions*. During each *repetition*, we increase δ by adding $\delta_\uparrow$ every Δf frames. Initially, at the beginning of the first *repetition* of each *source*, δ is set to 0, which means that no foveation is happening.

rep	1	2	3	4	5	6	7	8	9	10
$\delta_\uparrow$	10	5	3	2	2	1	1	1	1	1
$\delta_\downarrow$	25	20	17	15	10	8	8	5	5	5
Δf	25	25	25	25	25	50	50	50	50	50

TABLE I
PARAMETERIZATION OF THE ASSESSMENT PROCEDURE

The only possible *interaction* a participant can perform is a button press that indicates that she perceived a visual distortion in the video. Following an *interaction* by the participant, the current *repetition* is canceled and δ is reduced by $\delta_\downarrow$. After displaying a black screen for a second, either the next repetition of the same video is started, or the program advances to the next source. The value of δ is carried on in between repetitions of the same source video, and the values of $\delta_\uparrow, \delta_\downarrow$ and Δf are updated according to Table I. If a participant does not interact during a repetition, δ is simply increased further according to the parameterization of the next repetition.

This parameter choice enforces a rapid introduction of degradations during the first few repetitions of each source video. As $\delta_\uparrow$ and Δf are adapted over time, the participants can spend attention to increasingly minute distortions. The idea is to approach the individual δ quickly in the beginning, and then ever-more slowly, in order to obtain more precise results.

We conducted this study with 10 participants, totaling 1000 displayed repetitions, in which we observed 734 interactions.

VI. Results and Discussion

During our experiments, we gathered the fixation points and the δ values that were used to encode each frame and the information on participant interactions. The results presented in Table II are calculated at the .25 JND, which is the distortion level δ, at which 25% of the participants expressed that a noticeable distortion is present. Fig. 3 indicates that the rapid initial increases in δ, likely in combination with a certain reaction time required by the participants to recognize distortions, leads to a bias due to which the initial participant interaction reports are over-exaggerated.

Throughout the repetitions, the reported δ values shift towards lower, more plausible values, and the participants agree up to a certain difference in subjective sensitivity.

src	*int.*	JND (δ)	br_0	br_{fov}	Saving
#1	72	20	6411999.51	2167756.07	66.19%
#2	75	25	4406088.82	1381735.77	68.64%
#3	80	18	5082276.75	1564011.82	69.22%
#4	70	20	8181685.61	2663371.61	67.44%
#5	75	20	8749157.74	1945557.09	77.76%
#6	74	18.25	12931029.08	4619756.35	64.27%
#7	72	17.25	3926922.22	1303707.70	66.80%
#8	75	14.5	5486313.74	1545896.75	71.82%
#9	71	19	8721752.96	2700076.71	69.05%
#10	70	20	5900066.11	1909775.83	67.63%
avg	73.5	19.2	6979729.25	2180162.42	68.88%

TABLE II
FOR EACH SOURCE, WE LIST THE TOTAL NUMBER OF PARTICIPANT INTERVENTIONS OVER ALL REPETITIONS, δ AT THE .25 JND, THE BITRATE br_0 OF THE UNFOVEATED REFERENCE AND THE BITRATE br_{FOV} OF THE FOVEATED VERSION AT THE .25 JND, AS WELL AS THE RELATIVE BITRATE SAVINGS.

Even though the .25 JND is widely used [28], this choice likely overestimates the performance of our method due to high δ values in early repetitions. For reference, we also list the bitrate differences at the .1 JND, at which only 10% of the participants expressed that noticeable distortions were present.

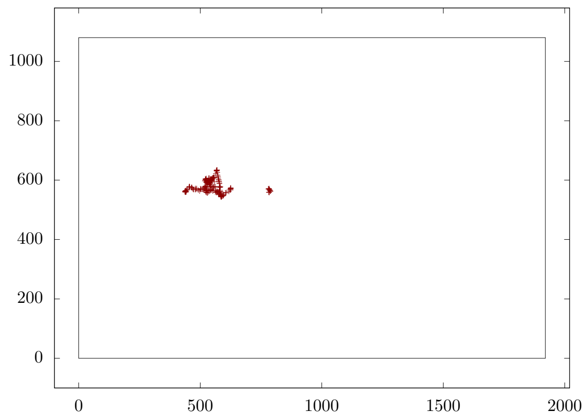

Fig. 4. Pixel fixation points of the whole repetition, corresponding to Fig. 5.

Fig. 5. Frame sampled from a video where severe distortions in the periphery went unnoticed. Extracted from participant #4, source #2, repetition #4.

src	JND (δ)	br_0	br_{fov}	Saving
#1	18	6411999.51	2343415.84	63.45%
#2	22	4406088.82	1484050.98	66.41%
#3	14.90	5082276.75	2207136.72	56.57%
#4	18	8181685.61	2861446.75	65.02%
#5	15	8749157.74	2528454.37	71.10%
#6	12	12931029.08	6066398.65	54.08%
#7	14	3926922.22	1475211.66	62.43%
#8	10.4	5486313.74	1885710.74	65.62%
#9	14	8721752.96	3456326.30	60.37%
#10	16	5900066.11	2208880.07	62.56%
avg	15.43	6979729.25	2651704.20	62.55%

TABLE III
THE RESULTS OF TABLE II AT THE .1 JND

We utilize all recorded gaze paths for comparison, resulting in a whole 1000 foveated videos that are compressed at their JND level. Partial gaze paths that were recorded up to the n-th frame and then discontinued due to participant interaction are compared against an unfoveated compression of the same source, which was also cropped at the n-th frame. When utilizing the average δ at the 0.1 JND for all these comparisons instead of the per-video δ, we achieve an *average bitrate saving of* 63.24% in comparison to the baseline.

VII. FUTURE WORK

The good experimental results motivate further work on research into foveated real-time video coding. A follow-up task would be to verify our findings in larger-scale studies, in which aim to deduce more generally valid recommendations on the parameter choices, as our current experiment is limited to rather few contents. Higher resolutions and screen sizes should, as mentioned earlier, further improve bitrate gains, since it is possible to reduce the quality in a larger share of each frame. We expect the limiting factor to be the increase in encoding time; this can be alleviated with task-specific hardware implementations.

A straightforward improvement for our exemplary x264 backend would be the dynamic introduction of I frames when large saccades are occurring. The current approach with frequent I frames ameliorates certain issues in this context, but raises other difficulties, mainly the high bitrate required for I frames in contrast to P frames. The reaction speed in terms of quality adjustment on longer saccades is a general problem in foveated video coding. This might be mitigated by higher framerates, which desired anyways in certain scenarios, such as video game streaming.

As the human eye is more sensitive to abrupt contrast changes and motion in peripheral vision [4], we assume that the discernibility of changes in coarsely quantized blocks can be reduced by simple post-processing such as blurring.

ACKNOWLEDGMENT

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 251654672 – TRR 161 (Project A05).

REFERENCES

[1] "Cisco Visual Networking Index: Forecast and Trends, 2017–2022," *Cisco Systems*, pp. 1–7, 2018.

[2] Y. Chen, D. Murherjee, J. Han, A. Grange, Y. Xu, Z. Liu, S. Parker, C. Chen, H. Su, U. Joshi *et al.*, "An overview of core coding tools in the AV1 video codec," in *Picture Coding Symposium (PCS)*. IEEE, 2018, pp. 41–45.

[3] L. Merritt and R. Vanam, "x264: A high performance H. 264/AVC encoder," 2006. [Online]. Available: http://neuron2.net/library/avc/overview_x264_v8_5.pdf

[4] B. A. Wandell, *Foundations of Vision*. Sinauer Associates, 1995. [Online]. Available: https://foundationsofvision.stanford.edu

[5] J. Hirsch and C. A. Curcio, "The spatial resolution capacity of human foveal retina," *Vision research*, vol. 29, no. 9, pp. 1095–1101, 1989.

[6] B. Girod, "Eye movements and coding of video sequences," in *Visual Communications and Image Processing: Third in a Series*, vol. 1001. International Society for Optics and Photonics, 1988, pp. 398–405.

[7] Z. Wang and A. C. Bovik, "Embedded foveation image coding," *Transactions on Image Processing (TIP)*, vol. 10, no. 10, pp. 1397–1410, IEEE, 2001.

[8] V. Hosu, F. Hahn, O. Wiedemann, S.-H. Jung, and D. Saupe, "Saliency-driven image coding improves overall perceived JPEG quality," in *Picture Coding Symposium (PCS)*. IEEE, 2016, pp. 1–5.

[9] A. Patney, M. Salvi, J. Kim, A. Kaplanyan, C. Wyman, N. Benty, D. Luebke, and A. Lefohn, "Towards foveated rendering for gaze-tracked virtual reality," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 6, p. 179, 2016.

[10] G. Illahi, M. Siekkinen, and E. Masala, "Foveated video streaming for cloud gaming," in *19th International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2017, pp. 1–6.

[11] S. Arndt and J.-N. Antons, "Enhancing video streaming using real-time gaze tracking," in *in Proc. ISCA/DEGA Workshop on Perceptual Quality of Systems*, 2016.

[12] "FFmpeg." [Online]. Available: https://ffmpeg.org

[13] "SDL: Simple direct media layer." [Online]. Available: https://libsdl.org

[14] "Minimalist GNU for windows." [Online]. Available: http://mingw.org

[15] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on circuits and systems for video technology*, vol. 13, no. 7, pp. 560–576, 2003.

[16] H.-W. Hunziker, "Im auge des lesers: foveale und periphere wahrnehmung-vom buchstabieren zur lesefreude," *The Eye of the Reader: Foveal and Peripheral Perception-from Letter Recognition to the Joy of Reading). Zurich: Transmedia*, 2006.

[17] P. Lungaro, R. Sjöberg, A. J. F. Valero, A. Mittal, and K. Tollmar, "Gaze-aware streaming solutions for the next generation of mobile vr experiences," *IEEE transactions on visualization and computer graphics*, vol. 24, no. 4, pp. 1535–1544, 2018.

[18] H. Sheikh, "Live image quality assessment database release 2," *http://live.ece.utexas.edu/research/quality*, 2005.

[19] X. Liu, M. Pedersen, and J. Y. Hardeberg, "CID: IQ–a new image quality database," in *International Conference on Image and Signal Processing*. Springer, 2014, pp. 193–202.

[20] N. Ponomarenko, V. Lukin, J. Astola, and K. Egiazarian, "Analysis of HVS-metrics' properties using color image database TID2013," in *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, 2015, pp. 613–624.

[21] V. Hosu, H. Lin, T. Sziranyi, and D. Saupe, "KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment," *arXiv preprint arXiv:1910.06180*, 2019.

[22] H. Lin, V. Hosu, and D. Saupe, "KADID-10k: A large-scale artificially distorted IQA database," in *Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2019, pp. 1–3.

[23] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *Transactions on Image Processing (TIP)*, vol. 27, no. 1, pp. 206–219, 2017.

[24] S. Bianco, L. Celona, P. Napoletano, and R. Schettini, "On the use of deep learning for blind image quality assessment," *Signal, Image and Video Processing*, vol. 12, no. 2, pp. 355–362, 2018.

[25] O. Wiedemann, V. Hosu, H. Lin, and D. Saupe, "Disregarding the big picture: Towards local image quality assessment," in *Tenth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2018, pp. 1–6.

[26] C. G. Bampis, A. C. Bovik, and Z. Li, "A simple prediction fusion improves data-driven full-reference video quality assessment models," in *Picture Coding Symposium (PCS)*. IEEE, 2018, pp. 298–302.

[27] B. Keelan, *Handbook of image quality: characterization and prediction*. CRC Press, 2002.

[28] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.

[29] M. Barkowsky, N. Staelens, L. Janowski, Y. Koudota, M. Leszczuk, M. Urvoy, P. Hummelbrunner, I. Sedano, and K. Brunnström, "Subjective experiment dataset for joint development of hybrid video quality measurement algorithms," in *Third Workshop on Quality of Experience for Multimedia Content Sharing QoEMCS*, 2012, pp. 1–4.