QoE Analysis for Videos with Adaptive Resolution Transcoding

Thalita P. da C. Ribeiro, Eduardo S. S. Coelho, Milena F. Pinto, *Senior Member, IEEE,* and Gabriel F. C. de Queiroz

***Abstract*—Providing high-resolution video streaming services with a high Quality of Experience (QoE) for end users is a challenge today. This is due to limitations in both the hardware of user devices and the channel's bandwidth. In parallel, there is a noticeable decrease in people's attention span, which can make it difficult for them to perceive details on the screen. This paper introduces Transcoding Resolution Induced by Custom Keyframes (TRICK), an adaptive transcoding strategy that reduces the video resolution in frames where this reduction might not be noticed, saving bandwidth and possibly not compromising the QoE, which has a time complexity of $O(N)$. An opinion survey compares videos with TRICK to videos in full 480p and 1080p. Videos with TRICK are liked by more than 84% of people, and offer a similar experience to that of 1080p videos, with savings of up to 18% in bandwidth in these transmissions.**

***Index Terms*—QoE, TRICK, transcoding, video streaming.**

I. Introduction

VIDEO transmissions are the most bandwidth-intensive traffic on the Internet, being present in several network applications today, such as video streaming, social networks, and games [1]–[3]. According to the Cisco Visual Networking Index (VNI), by 2022, videos comprised 82% of the Internet traffic [4]. This surge in video traffic is partially driven by the popularity of Over-The-Top (OTT) services, such as YouTube and Netflix [5]. In fact, by 2024, Netflix subscribers streamed 362 962 hours of content per minute[1].

This increase in video streaming consumption has placed significant strain on network infrastructures, as video transmissions are expected to require up to 20 GB per hour of usage by 2030 [6]. Such overload can lead to prolonged video startup times, freezing during playback [6]–[8], and overall user dissatisfaction, as the user's immersion in the video content is compromised [9]. This negatively impacts the users' Quality of Experience (QoE), a subjective metric defined here as the degree of user satisfaction with the service [10]. User satisfaction arises from a positive interaction with the service and the fulfillment of their expectations, influenced by the utility or enjoyment provided by the service [9], [11], [12].

In parallel, video streaming growth also leads to longer screen exposure times, which may cause long-term concentration loss, inattention, and eyestrain [13]. Average attention span has decreased to 47 s [14]. Moreover, 50% of people acknowledge checking their smartphones during tasks, possibly due to the distractions caused by on-screen notifications [15].

Combining these issues, this paper proposes the Transcoding Resolution Induced by Custom Keyframes (TRICK) strategy, an adaptive transcoding algorithm, which varies the resolution of the video being displayed. TRICK reduces bandwidth consumption by using lower resolutions to decrease the video size, while maintaining a high level of QoE for the user. TRICK searches for sequences of images (frames) that can have resolutions reduced without the user, who may be distracted by the scene's content, noticing a degradation in image quality. This is done by searching for frame sequences (e.g., with high motion or low brightness), which may represent moments when the user is inattentive (distracted or entertained by the scene) or struggles to visually detect a resolution drop.

An opinion survey is conducted, in which transcoded videos (using TRICK) and non-transcoded videos (purely in 480p or 1080p resolutions) are displayed to 90 participants in order to evaluate the Mean Opinion Score (MOS), a metric adopted here to measure QoE. Results show that 84.4% of participants consider the TRICK exhibitions satisfactory. Additionally, 36.7% of participants identify resolution drops even in non-transcoded videos, which can reach up to 70%. The main contributions of this paper are as follows.

- The TRICK strategy is introduced. TRICK is a novel algorithm with a time complexity of $O(N)$, which decides how video frames are transcoded based on their content. This can be done by reducing the resolution at moments when users may not notice it (e.g., a fast-paced scene).
- A database with 17 280 frames is created for this study[2].
- An opinion survey is carried out to evaluate the intuition behind TRICK's rationale, that is, that reducing the resolution based on the frame content (e.g., scenes of close-up, darkness, and motion) may go unnoticed by distracted participants and not hinder their QoE.

T. P. da C. Ribeiro (e-mail: thalita.padua@aluno.cefet-rj.br, ORCID: 0009-0003-0116-0356), E. S. S. Coelho (e-mail: eduardo.coelho@aluno.cefet-rj.br, ORCID: 0009-0005-8628-6373), and G. F. C. de Queiroz (e-mail: gabriel.queiroz@cefet-rj.br, ORCID: 0000-0003-2698-4346) are with the Department of Telecommunications Engineering (CCGTEL), Federal Center for Technological Education Celso Suckow da Fonseca (CEFET/RJ), Rio de Janeiro, RJ, Brazil. M. F. Pinto (e-mail: milena.pinto@cefet-rj.br, ORCID: 0000-0001-6916-700X) is with the Department of Electronic Engineering (CCGELT), Federal Center for Technological Education Celso Suckow da Fonseca (CEFET/RJ), Rio de Janeiro, RJ, Brazil.

This study was financed in part by the Coordination for the Improvement of Higher Education Personnel (CAPES), Finance Code 001; the National Council for Scientific and Technological Development (CNPq), Finance Code 304590/2023-1; and the Carlos Chagas Filho Research Support Foundation of the State of Rio de Janeiro (FAPERJ), Finance Code E-26/201.298/2022.

Digital Object Identifier: 10.14209/jcis.2025.11

Submission: 2025-03-20, First decision: 2025-06-06, Acceptance: 2025-11-12, Publication: 2025-12-05.

[1]https://www.domo.com/learn/infographic/data-never-sleeps-12

[2]https://www.github.com/eduardoscalzer/trick

The remainder of this paper is organized as follows. Section II discusses the related work. Section III presents the system architecture. The results are shown in Section IV. Finally, Section V concludes this paper and points out future work.

II. Related work

There have been several studies regarding video streaming over the years, each employing different approaches [2], [5], [16], [17]. Schwarzman *et al.* [18] propose a model for rate-based and buffer-based Adaptive Bit Rate (ABR), a technology that adjusts video transmission based on such metrics. Ge *et al.* [19] implement caching of video content at the network edge to enhance video delivery and improve users' QoE. Dao *et al.* [20] investigate which factors impact live streaming quality, such as architectures, and protocols.

In addition to works focused on video streaming, there are those addressing inattention and distraction [13], [21]. Duffy and Thain [22] point out that people tend to underestimate the frequency of their smartphone-checking habits while doing tasks. Indeed, people may involuntarily glance away while browsing the Web [23] or consciously do so to read video subtitles [24]. In both cases, viewers can become distracted and miss details on the screen.

Considering the state of the art, this is the first paper to jointly address resolution transcoding and the presence of specific characteristics in video frames that may trigger inattention or a lack of detail perception. Thus, these moments in the video can become suitable for reducing resolution, saving bandwidth without compromising user experience. There is a scarcity of works investigating this intersection between transcoding and users' detail perception in video streaming. Hence, this paper stands out for its originality, presenting an innovative approach in an often neglected area.

III. System modeling

The modeling adopted here involves a video-on-demand streaming application, as shown in Fig. 1. Video files are stored on a server and delivered through HTTP Adaptive Streaming (HAS) upon a client's request. Videos are divided into segments, which are sequences of frames, encoded at multiple resolution levels [2]. Then, during playback, the client requests segments, and each one may have a different resolution.

Traditional ABR schemes, such as HAS, typically rely on bandwidth and buffer levels to decide the resolution (transcoding) of a whole segment [17]. However, this results in prolonged low-resolution viewing, which may reduce QoE, as the users are more likely to notice quality degradation (resolution drop). Hence, this paper proposes the TRICK strategy. TRICK transcodes each frame solely based on its visual content. For example, resolution can be reduced in frames with high motion (e.g., races). Thus, only certain frames of a segment are in low resolution, which can preserve QoE. Furthermore, low-resolution frames decrease video size in bytes, potentially reducing transmission time and network load.

TRICK selects resolution based on the visual characteristics of the video scenes. This is done to analyze how each can impact QoE, due to the user's propensity for inattention during these scenes. Within the scope of this paper, the following characteristics are chosen:

- Close-up: videos with many facial details, colors, and little camera changes.
- Darkness: videos with low lighting and lack of details.
- Motion: videos with intense camera movement.
- Brightness: videos with glaring lighting.

The selection of characteristics, as illustrated in Fig. 2, is primarily based on the authors' intuition, considering common video content types. However, this is also supported by previous work. Mkwawa *et al.* [11] examine how network load affects streaming quality, noting that slow- and fast-motion videos require different bit rates. Rossi *et al.* [13] discuss how contrast (i.e., differences between bright and dark areas) and lighting, which is directly related to color perception, influence visual quality. Here, close-up and motion scenes are expected to easily distract users. Dark scenes are inherently harder to see clearly, possibly disturbing the user's vision, while bright scenes are the opposite. Therefore, it is essential to study how transcoding, associated with these characteristics, affects QoE.

TRICK operates as follows. First, the video is preprocessed to identify keyframes for transcoding. These frames contain scenes of close-up, darkness, motion, and brightness. This information can be extracted from threshold conditions (e.g., luminance levels, pixel motion intensity). Then, for each keyframe, a low resolution (in L) is assigned at random. In this paper, these four characteristics are associated with lower resolutions to assess the impact on QoE, as surveyed in Section IV. Keyframe and resolution information can be stored in a simple data structure, such as a list or array, with a maximum size of f, where f is the number of video frames. This preprocessing can be done in advance, prior to user requests. Thus, the server only verifies this structure and sends the transcoded video segments to the user, avoiding additional transmission delays imposed by TRICK. The user must inform the server if they choose to use TRICK, akin to specifying a video resolution of 720p or 1080p, for example. Therefore, the transmission overhead added by TRICK is not significant.

Pseudocode for TRICK is given in Algorithm 1. TRICK has a time complexity of $O(N)$, as detailed in the Appendix. This low complexity is important because the preprocessing occurs on the server, which handles requests from multiple clients in a real scenario; hence, it should not be overloaded. Additionally, TRICK needs to run only once per video to set a resolution for each frame, making $O(N)$ an acceptable complexity.

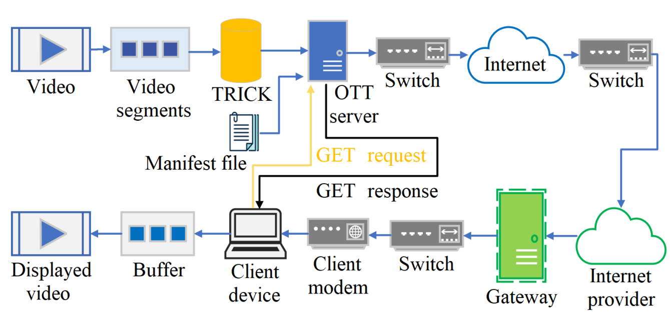

Fig. 1. Network architecture including TRICK.

(a) (b) (c) (d)

Fig. 2. Examples of frames in 1080p for the videos (a) Close-up, (b) Darkness, (c) Motion, and (d) Brightness.

Algorithm 1: TRICK.

input: original video V, set of low resolutions L, set of high resolutions H, maximum number of consecutive frames in low resolutions c, number of frames to be encoded in each resolution n_r

output: transcoded video V_T

$V_T \leftarrow \emptyset$

let F be the set of frames in V that include at least one of the following characteristics: close-up, darkness, motion, and brightness

while there are fewer frames in V_T than in V

  let v be the first frame in V not yet added to V_T

  if $v \in F$ **then**

    add frame v and its subsequent frames to V_T, using one of the resolutions in L, until c frames are added or there are n_r frames of the selected resolution r in V_T

    let v be the first frame in V not yet added to V_T

  add frame v to V_T, using one of the resolutions in H for which n_r has not been reached yet

return V_T

IV. Results and discussion

The videos used in this paper are one minute long, and are generated by concatenating short clips (10 to 20 s) retrieved from Pexels[3], a website that provides copyright-free videos. These clips are obtained through a keyword search for each of the four characteristics: Close-up, Darkness, Motion, and Brightness. As a result, each video consists almost entirely of frames exhibiting the target characteristic (i.e., motion).

Each video is available in 480p, 720p, and 1080p, with a frame rate of 24 fps, commonly used in the film industry[4]. The H.264 standard is utilized for video compression due to its efficiency and widespread support even in legacy devices [25]. As TRICK is agnostic in terms of video encoding and frame rate, more recent techniques [26], such as AV1 and H.266, can also be employed. Additionally, unlike traditional ABR schemes, which adapt quality based on bandwidth and buffer levels, TRICK operates independently of these parameters.

From each video, four versions are created, referred to as exhibitions. These exhibitions include: the video entirely in 480p, entirely in 1080p, and two implementations using TRICK, which employ transcoding. As the videos are composed almost entirely of keyframes, the frames to be transcoded were selected by visual inspection, following Algorithm 1.

The two implementations of TRICK are TRICK-3 and TRICK-10, which operate by limiting the value of c, as per Algorithm 1. This is the maximum duration a video can stay at low resolutions, up to 3 s or 10 s, respectively. Upon reaching this limit, it must switch to a higher resolution. Low-resolution frames are distributed across half of the video's duration (totaling 20 s in 480p and 10 s in 720p), which implies a 50% quality degradation. Here, degradation is the proportion of frames in low resolution over the total video duration. This is done by setting the number of frames in each resolution (n_r in Algorithm 1). This assumption is intended to simulate a challenging scenario, giving users more opportunities to notice the quality variation. The 50% value was chosen to ensure participants were exposed to a balanced distribution of frames, as a smaller proportion could reduce the number of resolution changes. If users are not dissatisfied with the degradation, then TRICK performs its role well.

The opinion survey involves presenting the four exhibitions of each video to 90 participants, with question sessions conducted between the exhibitions of different videos. This is done in person and offline to prevent connection issues from affecting the participants' QoE. Although participants did not use smartphones during the exhibitions, they may still become inattentive or distracted, as they may unconsciously look away, blink, or glance at a wall. They are asked about their perception of resolution changes and their satisfaction regarding the exhibition. Satisfaction reflects the participants' subjective opinion regarding the viewing experience [10], in terms of being positive and pleasant [12]. Additionally, participants rate visual quality from 1 (poor) to 5 (excellent) based on personal criteria [9] to compose the MOS [27].

A. Opinion analysis

Tab. I presents average MOS values and percentages for both satisfaction and resolution change perception. These results are for all exhibitions of the four videos, with confidence intervals at the 95% level, and assuming a normal distribution. Additionally, MOS values are also estimated using the UVQ model [28], an industry-grade model from YouTube that resembles a subjective quality assessment.

One of the objectives of TRICK is to ensure that resolution changes do not affect the participants' experience. Although more than half of the participants claim to notice resolution changes in TRICK exhibitions, even the exhibitions fully in 480p and 1080p also have high values of resolution change perception. This is even more evident in the results for the video Close-up; TRICK-3 and TRICK-10 are perceived as a smoother experience here. This is possibly due to them not remembering the variation (low attention span), misinterpreting the blur as resolution, or eyestrain [13]. The participants are ordinary individuals, most of whom did not claim to work with image processing or photography during the evaluations. Furthermore, this also suggests that, in all cases, at least 36.7% of people cannot correctly recognize a resolution change. However, even if some people are inclined to perceive changes where they do not occur, TRICK exhibitions are rated as satisfactory by up to 84.4% of participants, which indicates that TRICK can provide a good QoE.

In terms of MOS, transcoding is favorable for both Darkness and Close-up, as TRICK-3 and TRICK-10 are significantly better than 480p, and even equivalent to 1080p in the case of Close-up. Darkness has the lowest overall MOS values, indicating that participants tend to dislike very dark scenes, regardless of resolution. This is evident as the MOS for Darkness in 1080p is similar to the MOS for Close-up in 480p. Regarding the video Motion, TRICK's MOS is only about 0.45

[3]https://www.pexels.com/

[4]https://www.adobe.com/creativecloud/video/discover/frame-rate.html

TABLE I
OPINION SCORE RESULTS FOR EVERY EXHIBITION.

Video	Exhibition	MOS (Survey)	MOS (UVQ)	Perception of change	Satisfaction
Close-up	1080p	3.98 ± 0.14	4.14	68.9%	85.6%
	480p	3.48 ± 0.16	3.93	70.0%	78.9%
	TRICK-3	3.82 ± 0.17	4.08	67.8%	81.1%
	TRICK-10	3.92 ± 0.19	4.07	54.5%	84.4%
Darkness	1080p	3.57 ± 0.20	3.09	53.3%	73.3%
	480p	2.50 ± 0.23	2.91	44.4%	34.4%
	TRICK-3	3.02 ± 0.20	3.04	78.9%	52.2%
	TRICK-10	3.18 ± 0.19	3.05	76.7%	61.1%
Motion	1080p	3.91 ± 0.18	3.76	41.1%	77.8%
	480p	3.17 ± 0.22	3.62	45.6%	55.6%
	TRICK-3	3.45 ± 0.20	3.71	62.2%	70.0%
	TRICK-10	3.46 ± 0.21	3.68	58.9%	72.2%
Brightness	1080p	4.20 ± 0.17	3.65	36.7%	86.7%
	480p	3.31 ± 0.20	3.61	60.0%	64.4%
	TRICK-3	3.23 ± 0.21	3.60	82.2%	56.7%
	TRICK-10	3.44 ± 0.21	3.61	67.8%	65.6%

lower than that of 1080p. Transcoding is barely perceivable compared to non-transcoded exhibitions, the resolution change perception for TRICK-10 is only 13.3% higher than that for 480p. This suggests that fast-moving frames favor transcoding, as participants are distracted by trying to follow the scene.

Among the four characteristics, brightness is the most challenging for TRICK. Brightly lit scenes hinder transcoding, as TRICK-3 and TRICK-10 show MOS values similar to 480p's in both the survey and the UVQ model, with resolution changes noticed by up to 82.2% of participants. Unlike the other videos, there are no visual distractions here; the scenes are sharp and clear. Brightness is linked to better contrast and detail perception, as more light is captured by the photoreceptors in the human eye [13], [21]. Light is also directly related to visual comfort and well-being. This suggests that bright scenes should be kept at a single resolution, either 1080p (for better MOS) or 480p (for bandwidth savings, despite QoE).

In summary, TRICK performs well and surpasses the exhibitions in 480p for Close-up, Darkness, and Motion. This stands for the MOS values of both opinion survey and UVQ model. However, this does not apply for the video Brightness, as this characteristic shows itself to be unfavorable for transcoding.

B. Data size analysis

In addition to preserving QoE, TRICK can reduce transmitted data and save bandwidth, as low-resolution frames have smaller data sizes. To evaluate this reduction, all previously discussed exhibitions are considered here, alongside a full 720p exhibition and two TRICK versions with 25% degradation. As discussed, degradation refers to the number of low-resolution frames inserted, adjusted by n_r in Algorithm 1. It is worth mentioning that the opinion survey only includes the TRICK versions with 50% quality degradation because it represents a severe degradation case, with multiple resolution changes and low-resolution frames. This decision is due to the importance of testing TRICK in a challenging scenario where participants can more noticeably perceive these changes.

Fig. 3 shows values for each exhibition, normalized against 1080p's file size. TRICK versions at 50% degradation produce slightly smaller files than at 25%, due to having more low-resolution frames. Despite having different frame-by-frame transcoding decisions (due to different c values), TRICK-3 and TRICK-10 achieve a similar video size reduction. On average, video size decreases by about 10%, reaching up to 18% for the video Darkness. For longer content, such as movies on an OTT platform, this can save several GBs [6], reduce user download time, and ease bandwidth consumption.

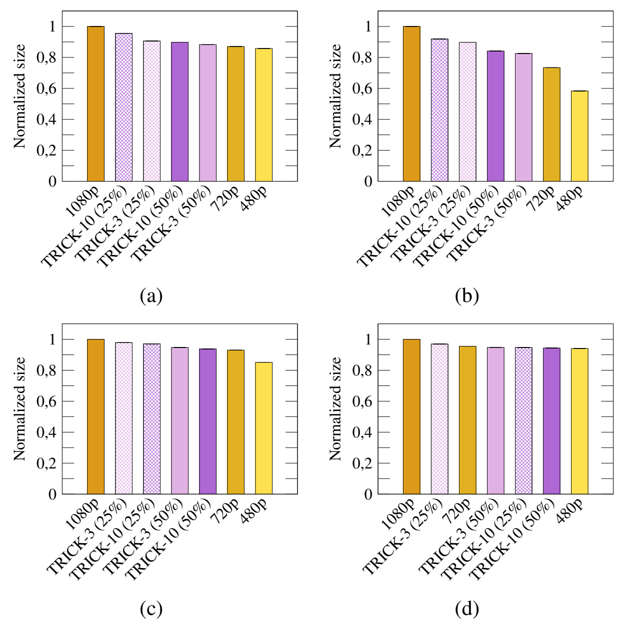

Fig. 3. File sizes for the videos (a) Close-up, (b) Darkness, (c) Motion, and (d) Brightness.

V. CONCLUSION AND FUTURE WORK

This paper proposes the TRICK strategy, which aims to mask resolution drops in videos so that they are not perceived by users and do not hinder their QoE. This is done by selecting keyframes for resolution transcoding. These frames contain scenes with significant darkness or camera movement, for example. Two implementations of TRICK, TRICK-3 and TRICK-10, are proposed. Each limits the time a video can be continuously displayed in low resolution to 3 s and 10 s, respectively. Subsequently, an opinion survey with 90 participants is conducted. Exhibitions for four different videos are presented: Close-up, Darkness, Motion, and Brightness. Results include MOS from the opinion survey and UVQ (YouTube's model for MOS estimation), resolution change perception, and participants' satisfaction.

TRICK has positive impacts on QoE for the videos Close-up, Darkness, and Motion; even achieving a MOS equivalent to that of a 1080p exhibition in the case of Close-up. However, scenes with intense brightness are not suitable for transcoding and should remain at a constant resolution. Additionally, there is an average reduction of 10% in file size for all TRICK exhibitions, reaching up to 18% for the video Darkness.

Future work includes additional MOS collections, and simulating a communication network to compare TRICK with traditional ABR approaches, as well as the automatic detection of keyframes using machine learning. Another potential direction is adapting TRICK for live streaming scenarios, in which keyframes cannot be identified beforehand because the video content is not known in advance.

REFERENCES

[1] S. C. Madanapalli, A. Mathai, H. H. Gharakheili, and V. Sivaraman, "ReCLive: Real-Time Classification and QoE Inference of Live Video Streaming Services," in *2021 IEEE/ACM 29th Int. Symp. on Qual. of Service (IWQOS)*. Piscataway, NJ, United States: IEEE, 2021, pp. 695–702. doi: 10.1109/iwqos52 092.2021.9 521 288.

[2] A. Borges, B. Zatt, M. Porto, and G. Correa, "A systematic literature review on video transcoding acceleration: challenges, solutions, and trends," *Multimedia Tools and Appl.*, vol. 83, p. 64079–64108, 2024. doi: 10.1007/s11042-023-17862-w.

[3] T. M. Santamaria-López, S. H. J. Encalada, K. I. C. Avila, and W. G. T. Samaniego, "Analysis Of The Transmission Models Used In Video Streaming And The Video Encoding Systems Most Commonly Used In The Transmission Of Content," *Migration Lett.*, vol. 21, p. 597–608, 2024. Available at: https://migrationletters.com/index.php/ml/article/view/7975.

[4] G. Esakki, A. S. Panayides, V. Jalta, and M. S. Pattichis, "Adaptive Video Encoding for Different Video Codecs," *IEEE Access*, vol. 9, pp. 68 720–68 736, 2021. doi: 10.1109/ACCESS.2021.3077313.

[5] A. A. Barakabitze, N. Barman, A. Ahmad, S. Zadtootaghaj, L. Sun, M. G. Martini, and L. Atzori, "QoE Management of Multimedia Streaming Services in Future Networks: A Tutorial and Survey," *IEEE Commun. Surv. & Tut.*, vol. 22, pp. 526–565, 2020. doi: 10.1109/comst.2019.2958784.

[6] A. S. Andrae, "New perspectives on internet electricity use in 2030," *Eng. and Appl. Sci. Lett.*, vol. 3, pp. 19–31, 2020. doi: 10.30538/psrp-easl2020.0038.

[7] S. Petrangeli, T. Wu, T. Wauters, R. Huysegems, T. Bostoen, and F. De Turck, "A Machine Learning-Based Framework for Preventing Video Freezes in HTTP Adaptive Streaming," *J. of Netw. and Comput. Appl.*, vol. 94, p. 78–92, 2017. doi: 10.1016/j.jnca.2017.07.009.

[8] B. Chen, Z. Yan, Y. Zhang, Z. Yang, and K. Nahrstedt, "LiFteR: Unleash Learned Codecs in Video Streaming with Loose Frame Referencing," in *21st USENIX Symp. on Netw. Syst. Des. and Implementation (NSDI 24)*. Santa Clara, CA, United States: USENIX Association, 2024, pp. 533–548. doi: 10.5555/3 691 825.

[9] M. Farias, "Quality of Experience of Immersive Media – New Challenges," in *29th Brazilian Symp. on Multimedia and the Web*. Porto Alegre, RS, Brazil: SBC, 2023, pp. 3–4. doi: 10.1145/3 617 023. 3 617 025.

[10] C. Quadri, S. Gaito, R. Bruschi, F. Davoli, and G. P. Rossi, "A MEC Approach to Improve QoE of Video Delivery Service in Urban Spaces," in *2018 IEEE Int. Conf. on Smart Comput. (SMARTCOMP)*. Piscataway, NJ, United States: IEEE, 2018, pp. 25–32. doi: 10.1109/smartcomp.2018.00 095.

[11] I. Mkwawa, A. A. Barakabitze, and L. Sun, "Video Quality Management over the Software Defined Networking," in *2016 IEEE Int. Symp. on Multimedia (ISM)*. Piscataway, NJ, United States: IEEE, 2016, pp. 559–564. doi: 10.1109/ism.2016.0122.

[12] M. Rodrigues, K. Moura, K. Branco, V. Lelli, W. Viana, R. Andrade, and I. Santos, "Medication time? a user experience evaluation of mobile applications targeting people with diabetes," in *28th Brazilian Symp. on Multimedia and the Web*. Porto Alegre, RS, Brazil: SBC, 2022, pp. 275–283. doi: 10.1145/3 539 637.3 558 045.

[13] S. Rossi, N. Kara-José, E. M. Rocha, and N. Kara-Júnior, "Influence of lighting on visual performance," *Arquivos Brasileiros de Oftalmologia*, vol. 87, pp. 1–7, 2024. doi: 10.5935/0004-2749.2023-0257.

[14] J. Ducharme, "Why everyone's worried about their attention span: and how to improve yours [Online]," 2023. Available at: https://time.com/6302294/why-you-cant-focus-anymore-and-what-to-do-about-it.

[15] B. Duffy and M. Thain, "Do we have your attention? How people focus and live in the modern information environment [Online]," 2022. Available at: https://www.kcl.ac.uk/policy-institute/assets/how-people-focus-and-live-in-the-modern-information-environment.pdf.

[16] S. Afzal, V. Testoni, C. E. Rothenberg, P. Kolan, and I. Bouazizi, "A holistic survey of multipath wireless video streaming," *J. of Netw. and Comput. Appl.*, vol. 212, pp. 1–41, 2023. doi: 10.1016/j.jnca.2022.103581.

[17] C. Timmerer, H. Amirpour, F. Tashtarian, S. Afzal, A. Rizk, M. Zink, and H. Hellwagner, "HTTP Adaptive Streaming: A Review on Current Advances and Future Challenges," *ACM Trans. on Multimedia Comput., Commun. and Appl.*, vol. 21, pp. 1 – 27, 2025. doi: 10.1145/3736306.

[18] S. Schwarzmann, P. Breitbach, T. Zinner, and M. Rost, "Modeling Adaptive Video Streaming Using Discrete-Time Analysis," in *31st Int. Teletraffic Congr. (ITC 31)*. Piscataway, NJ, United States: IEEE, 2019, pp. 121–129. doi: 10.1109/itc31.2019.00 028.

[19] C. Ge, N. Wang, W. K. Chai, and H. Hellwagner, "QoE-assured 4K HTTP Live Streaming via Transient Segment Holding at Mobile Edge," *IEEE J. on Sel. Areas in Commun.*, vol. 36, pp. 1816–1830, 2018. doi: 10.1109/jsac.2018.2845000.

[20] N. Dao, A. Tran, N. H. Tu, T. T. Thanh, V. N. Q. Bao, and S. Cho, "A Contemporary Survey on Live Video Streaming from a Computation-Driven Perspective," *ACM Comput. Surv.*, vol. 54, pp. 1–38, 2022. doi: 10.1145/3519552.

[21] C. Tzouvaras, A. Dimara, A. Papaioannou, K. Karatzia, C.-N. Anagnostopoulos, S. Krinidis, K. I. Arvanitis, D. Ioannidis, and D. Tzovaras, "A Guide to Visual Comfort: An Overview of Indices and Its Applications," in *Artif. Intell. Appl. and Innov. (AIAI 2023). IFIP WG 12.5 Int. Workshops*. Cham, ZG, Switzerland: Springer Nature, 2023, pp. 183–194. doi: 10.1007/978–3–031–34 171–7_14.

[22] B. Duffy and M. Thain, "Are attention spans really collapsing? Data shows UK public are worried – but also see benefits from technology [Online]," 2022. Available at: https://www.kcl.ac.uk/news/are-attention-spans-really-collapsing-data-shows-uk-public-are-worried-but-also-see-benefits-from-technology.

[23] A. Carneiro, C. Gonzales, and C. Morimoto, "EyePursuitLinks - an Eye-pursuit Based Interface for Web Browsing Using Smart Targets," in *29th Brazilian Symp. on Multimedia and the Web*. Porto Alegre, RS, Brazil: SBC, 2023, pp. 16–24. doi: 10.1145/3 617 023.3 617 058.

[24] F. de Sá, D. Cavalcanti, and V. Becker, "Testes com usuários para análise de emoções em conteúdos audiovisuais utilizando EEG e eye tracking," in *29th Brazilian Symp. on Multimedia and the Web*. Porto Alegre, RS, Brazil: SBC, 2023, pp. 63–66. doi: 10.5753/webmedia_estendido.2023.235 663.

[25] W. Moina-Rivera, M. Garcia-Pineda, J. Gutiérrez-Aguado, and J. M. Alcaraz-Calero, "Cloud media video encoding: review and challenges," *Multimedia Tools and Appl.*, vol. 83, pp. 81 231–81 278, 2024. doi: 10.1007/s11042-024-18763-2.

[26] M. Uhrina, L. Sevcik, J. Bienik, and L. Smatanova, "Performance Comparison of VVC, AV1, HEVC, and AVC for High Resolutions," *Electron.*, vol. 13, pp. 1–21, 2024. doi: 10.3390/electronics13050953.

[27] ITU-T Recommendations, "ITU-T P.800.1 [Online]," 2016. Available at: https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=12972.

[28] Y. Wang and F. Yang, "UVQ: Measuring YouTube's Perceptual Video Quality [Online]," 2022. Available at: https://research.google/blog/uvq-measuring-youtubes-perceptual-video-quality/.

APPENDIX

COMPLEXITY ANALYSIS

The time complexity of Algorithm 1 can be broken up as follows. V_T is defined in $O(1)$ time. Line 2 involves traversing V to generate F, resulting in $O(N)$ complexity, given that V has size N. As the number of frames is presumably much greater than the number of characteristics observed (i.e., $N \gg 4$), this verification takes $O(N)$ time. In the While loop, lines 4 and 7 perform an assignment operation, which is $O(1)$. Line 5 verifies if $v \in F$ in $O(1)$ time, as the elements in F and V are ordered similarly. Each time line 6 adds $\min\{c, n_r\}$ frames to V_T, L is traversed; similarly, H is traversed for adding a single frame in line 8. Thus, during each iteration of the loop, line 6 performs $\min\{c, n_r\} \times L$ operations, while line 8 performs H operations. Since the number of low and high resolution types is much smaller than N, the complexities for lines 6 and 8 are simplified to $O(\min\{c, n_r\})$ and $O(1)$, respectively. Considering that each iteration adds $\min\{c, n_r\} + 1$ frames to V_T, the While loop is executed $\frac{N}{\min\{c,n_r\}+1}$ times. This means that, in all cases, the loop concludes with complexity limited by $O(N)$. Finally, returning V_T is $O(1)$. By combining all these parts, the total complexity of TRICK is equivalent to $O(N)$.