Video Rater: A Framework for Subjective Evaluation of Rendering Artifacts

Karol Cyganik[1] and Dariusz Brzezinski[2]

[1]Advanced Micro Devices, Inc. [2]Institute of Computing Science, Poznan University of Technology, Poland
karol.cyganik@amd.com



(a) System overview. (b) The Video Rater interface showing synchronized playback of ground truth and distorted sequences with interactive controls.

Figure 1: *Overview of the Video Rater system and its interface.*

Abstract

Temporal reconstruction techniques used in real-time rendering, including temporal anti-aliasing and neural upscaling methods, improve visual quality but introduce characteristic spatial and temporal artifacts. Existing image and video quality metrics are largely optimized for compression artifacts and fail to capture these degradations. To enable the development of data-driven metrics for rendering artifacts, we present a video assessment framework aligned with ITU-R BT.500 recommendations. The proposed framework is accompanied by a customizable open-source Video Rater web application (https://github.com/kordc/Video-rater) designed to gather subjective ratings. We demonstrate the framework's and tool's usability through a use-case survey of over 200 videos with various rendering artifacts (https://video-rater.cs.put.poznan.pl/).

CCS Concepts

• ***Computing methodologies*** → ***Machine learning; Computer vision; Image processing; Rendering;***

1. Introduction

In pursuit of photorealism at high frame rates, modern real-time rendering engines increasingly rely on temporal reconstruction techniques, such as AMD FidelityFX™ Super Resolution (FSR) [AMD26]. However, the improved perceptual quality achieved with temporal reconstruction techniques often comes at the cost of unique spatial and temporal artifacts, such as ghosting and flicker.

Common image and video quality metrics (PSNR, VMAF [LAK*16]) are optimized for compression artifacts and fail to adequately penalize rendering-specific video artifacts. While recent work attempts to address this, existing methods are either designed to assess the quality of a single rendered image, ignoring temporal artifacts [ANAM*20], or trained on relatively small datasets [JST*25]. Moreover, the development of dedicated per-

© 2026 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.2312/egp.20261000

delivered by EUROGRAPHICS DIGITAL LIBRARY www.eg.org diglib.eg.org

ceptual metrics for detecting rendering artifacts is hindered by a lack of standardized methodologies and tools.

In this work, we present a subjective video assessment framework based on industry standard recommendations, accompanied by a dedicated Video Rater application. The framework is designed to support the creation of large-scale perceptual datasets for creating new video quality metrics. We demonstrate the framework through a study on game-rendering artifacts, including blur, ghosting, flickering, aliasing, Moiré patterns, fireflies, and shimmering.

2. Methodology

Although demonstrated on temporal artifacts in games, the proposed methodology is applicable to any video quality assessment scenario. To ensure statistically reliable subjective scores, we designed a controlled assessment methodology based on the International Telecommunication Union BT.500 recommendations [Uni23]. Rendering artifacts are assessed by showing pairs of videos: a *distorted* video and its *ground truth* (GT). Each distorted video is played in a loop alongside its GT equivalent in a dedicated video player. Simultaneous distorted and GT playback allows precise quality assessment. Each video is categorized by its metadata, such as *source* (scene), *artifact type*, and *artifact strength*. Videos are assigned a score using a discrete five-point scale (1–5), following the standard practice for subjective television and video quality assessment [Uni23]. The definition of the quality scale may be found in Figure 1b. Each score is additionally linked to the user's *screen resolution* and their *demographic data*.

Users provide demographic data at the start of the survey. They are categorized as *Experts* (graphics R&D professionals) or *Common users* (e.g., game players) to analyze how domain knowledge and prior experience with artifact detection influence artifact perception and quality judgment. Experts are provided with an extended view offering additional interaction options, whereas other users are presented with a simplified interface that mimics typical gameplay conditions. To assess the significance of each reviewer's ratings, one of the ten videos is a ground-truth video paired with itself (video-identity check), and one of the ten videos is repeated (repeatability evaluation). If a user fails to give the GT video the highest possible score, or rates the repeated pair with a score that differs from the first rating by more than 1, all of the user's scores are treated as invalid and do not count toward the final statistics.

3. The Video Rater Tool

Following the proposed assessment methodology, we introduce the Video Rater Tool—a customizable Python web application for rating distorted videos against their ground-truth equivalents. While the framework proposes Expert and Common user categorization, the system supports flexible user grouping to accommodate diverse study designs. Similarly, the rating scale can be adjusted beyond the default five-point scale. Researchers can also add their own videos, even without corresponding GT sequences.

Upon registration, the system records the user's demographic data and calculates the effective screen resolution. An interactive tutorial guides users through the evaluation procedure. Each participant is assigned a predefined, randomized video queue that persists across sessions, allowing users to log in and resume their evaluation at any time. The main interface consists of a synchronized video player, a five-point quality rating control, and concise instructions (Figure 1b). The tool's custom video player, with a buffering strategy, ensures frame-perfect synchronization between the GT and distorted sequences. Additional controls, such as zooming, panning between distorted/GT sequences, and frame stepping, support detailed inspection of subtle spatial and temporal artifacts.

To enhance user engagement, the tool incorporates a gamification system through achievement badges. Users earn badges upon completing milestones, such as rating 10 videos, to encourage consistent participation throughout the study. Finally, a separate administrative dashboard for survey administrators supports dataset management and basic analyses of the collected ratings.

4. Use case and future work

To validate the system's capability to acquire subjective annotations at scale, we conducted an experimental study using a curated dataset of rendering artifacts. The framework was applied to assess short, high-quality sequences captured from real-time rendering scenarios. Each recorded video is 3 seconds long, consisting of exactly 90 frames at a resolution of 2160×3840. Videos are spatially cropped such that only a single dominant artifact is visible in each sequence, enabling targeted quality assessment. The study involves two participant groups (Experts and Common users) who evaluate more than 200 videos using the Video Rater tool. Videos were categorized by dominant artifacts (Fireflies, Flickering, Ghosting, Moiré patterns, Noise, Shimmering, and Overall Quality), produced using various anti-aliasing and supersampling techniques. Participants followed the proposed assessment methodology, rating each video using the five-point quality scale while comparing it against the ground truth equivalent. The deployment demonstrates the framework's readiness to consistently acquire subjective ratings that can be used to develop new perceptual metrics. Future work includes a detailed analysis of the collected ratings, the public release of the video artifact dataset, evaluation of existing objective quality metrics on the collected data, and development of new reference and no-reference metrics for neural rendering artifacts.

References

[AMD26] AMD: Amd fsr™ technologies. https://www.amd.com/en/products/graphics/technologies/fidelityfx/super-resolution.html, 2026. Accessed: 2026-01-29. 1

[ANAM*20] ANDERSSON P., NILSSON J., AKENINE-MÖLLER T., OSKARSSON M., ÅSTRÖM K., FAIRCHILD M. D.: Flip: A difference evaluator for alternating images. *Proc. ACM Comput. Graph. Interact. Tech. 3*, 2 (2020). 1

[JST*25] JINDAL A., SADAKA N., THOMAS M. M., SOCHENOV A., KAPLANYAN A.: CGVQM+D: Computer Graphics Video Quality Metric and Dataset. *Computer Graphics Forum* (2025). 1

[LAK*16] LI Z., AARON A., KATSAVOUNIDIS I., MOORTHY A., MANOHARA M.: Toward a practical perceptual video quality metric. https://medium.com/netflix-techblog/toward-a-practical-perceptual-video-quality-metric-653f208b9652, 2016. Accessed: 2026-01-29. 1

[Uni23] UNION I. T.: *Methodologies for the subjective assessment of the quality of television images*. Tech. Rep. Recommendation ITU-R BT.500-15, Radiocommunication Sector, 2023. 2

© 2026 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.