VIDEO QUALITY ASSESSMENT CONSIDERING THE FEATURES OF THE HUMAN VISUAL SYSTEM

Anastasia Mozhaeva[1,2], Vladimir Mazin[2], Michael J. Cree[1][0000−0002−2973−2710], and Lee Streeter[1][0000−0002−3987−6945]

[1] The University of Waikato, Hamilton, New Zealand
[2] Moscow Technical University of Communications and Informatics, Moscow, Russia
lee.streeter@waikato.ac.nz

Abstract. Nowadays, numerous video compression quality assessment metrics are available. Some of these metrics are "objective" and only tangentially represent how a human observer rates video quality. On the other hand, models of the human visual system have been shown to be effective at describing spatial coding. In this work we propose a new quality metric which extends the peak signal to noise ratio metric with features of the human visual system measured using modern LCD screens. We also analyse the current visibility models of the early visual system and compare the commonly used quality metrics with metrics containing data modelling human perception. We examine the Pearson's linear correlation coefficient of the various video compression quality metrics with human subjective scores on videos from the publicly available Netflix data set. Of the metrics tested, our new proposed metric is found to have the most stable high performance in predicting subjective video compression quality.

Keywords: Video quality metric · the model of visibility · acceptable visual quality · artefact perception.

1 Introduction

Video quality metrics are a critical component in modern streaming video processing algorithms. Alongside compression and data transmission, accurate quality estimation is key to maximising use of bandwidth while also maximising the user experience. Video quality methods can be divided into two categories: subjective and objective quality assessment criteria [1]. The human user is typically the final recipient in typical video processing applications, so subjective quality criteria that reflects human visual perception is arguably the more important method of assessing video quality [2]. Objective quality criteria are the most common and popular method of evaluating video quality since the assessment is performed algorithmically. Objective metrics avoid expensive research with user participation. At the present stage of the development of media content transmission technologies, objective algorithms for evaluating video quality

have achieved high optimization and simplicity. Unfortunately, popular objective prediction models correlate poorly with subjective perceptions of quality by the human visual system (HVS) and depend on the systems or processes involved [2]. On the other hand, there are algorithmically complex video quality metrics (VQM) based on models of the human visual system [1], [3] , and an open question is whether complex video quality metrics based on models of the human visual system provide significantly better predictions than objective metrics. Another problem when used of visual models, is in developing video quality metrics we must represent the HVS in software, a task impeded by the limited new fundamental knowledge of the HVS perception of video content using modern equipment.

In visual perception studies, the primary focus is most often on studying the physical aspects of HSV, of which video quality is a secondary question and often not discussed [4]. In addition, in most psychophysical experiments, participants enter a controlled laboratory environment in order to stabilise tests and control any confounding factors. In other words, a reduction in the number of experiments, consequently, errors, is made [5]. The laboratory approach limits many of the problems that arise in the "real world". Therefore, developers of video quality assessments and video compression algorithms need to determine how psychophysical results relate to quality in today's video content presentation environment, without a full understanding of the processes of HVS. Modern knowledge and tests about the relationship between HVS and quality metrics are needed.

Human visual sensitivity can be characterised by a distinct region of spatio-temporal frequencies. Region boundaries determine the visibility of artefacts in the displayed information [6]. More than five years ago, a linear model of spatio-temporal contrast sensitivity [7], which determines the visibility of visual information artefacts by the human eye, was presented. However, the viewing conditions in the research differed from modern computer monitors and television screens. In particular, cathode ray tube (CRT) screens were used, which are not in common use today, and the viewing angle was controlled to be small or normal to the screen, which is not necessarily true in everyday use. Consequently, the results do not guarantee a proper description of the conditions in which media content is presently consumed. At present, the problem of imitation in the evaluations of the quality of the video of the HVS has led to the creation in 2022, of a model HVS [8] that considers stimulus parameters: spatial and temporal frequency, eccentricity, luminance, and viewing area. However, to create the model, there were data from 11 publications, where the viewing conditions were also different from modern computer monitors and television screens. At the current stage of technological development, there are no theoretical obstacles to creating a sufficiently comprehensive video quality assessment correlated with HVS. However, there exists a practical obstacle, that large-scale experiments may be extremely time-consuming and expensive. In the previous work, we presented a methodology [9] for conducting large-scale experiments and presented initial measurements of the characteristics of the HVS [1].

While many video scoring metrics are available, this paper represents the first practical consideration of a comprehensive solution to adapting a simple objective video quality metric to a typical video user experience. The objective of this paper is to demonstrate the need to research and collect new HVS data under modern circumstances of providing information. We present a comparison of commonly used video quality metrics with HVS based metrics. Our hypothesis is that video quality metrics containing HVS data perform better than other currently used metrics. Consequently, we aim to show that there is a significant opportunity to develop and extend existing video quality metrics using new HVS based data.

2 Related work

Considering all video quality assessment methods, the most popular ones are peak signal-to-noise ratio (PSNR) and structural similarity image metric (SSIM) [10]. PSNR is used more often than other methods to assess similarities between original and reconstructed images and videos. PSNR is calculated on a logarithmic scale by amplitude (in decibels). The benefit being that the HVS also perceives brightness on a logarithmic scale. In PSNR, the signal-to-noise ratio based on standard deviation never gives overestimated results [10], which is why the method is the most widely used. However, PSNR is poorly correlated with visual quality estimation and does not consider spatial and temporal psychovisual models. PSNR gives significantly underestimated results, even a slight shift of the reference and estimation frame in space or desynchronisation of video sequence in time can degrade the performance of PSNR estimation [11]. Another common metric is structural similarity of the image (SSIM) [10]. However, it has been shown analytically and experimentally that, for images with fragments of large or small mean luminance values, the local estimates of the metric are unstable [4]. SSIM does not consider different levels of absolute luminance, temporal aspects or viewing distances, and poorly correlates with human perception.

A more fundamental approach to creating video quality metrics includes low-level visual modelling based on psychophysical models, such as the contrast sensitivity function (CSF) [3]: the threshold at which a human observer can detect change in a given brightness pattern as a function of spatial and temporal frequency. The artefacts visible to the human visual system in early vision are regulated by the function of visual sensitivity. Early vision involves three processes: filtering, encoding, and interpretation [12]. Human visual sensitivity can be characterised as a reference filter in terms of spatial and temporal frequencies and there is no need for rendering beyond the limitations of the region [6].

The existing HVS models do not consider all the necessary spatio-temporal variations of the stimuli. Popular CSF models, such as Barten [13] or Daly [14], do not include the temporal frequency. Kelly's spatio-temporal CSF [15] consider spatial and temporal frequencies without luminance. In 2016, Watson and Ahumada presented a linear model of spatio-temporal contrast sensitivity, called the "visibility pyramid" [7]. The authors constructed an exhaustive description

of spatial and temporal contrast sensitivity, and its dependence on retinal illumination, and also derived a number of strong relations, all from the modest results of much older studies [16–18]. StelaCSF, unlike previous studies, tries not to get a perfect fit for individual datasets but creates a single model that can explain all datasets without overfitting [8]. To model the 5-dimensional contrast sensitivity space, the authors combined data from 11 publications on CSF. The main interest of modelling is datasets. However, the equipment and, therefore, viewing conditions in older studies differed from the conventional computer monitors and television screens used today.

In previous work, we presented a new video quality metric (PSNR-M) [1] that considered data for the first part of early vision, namely the filtering stage that determines which spatial and temporal fluctuations in stimuli the HVS responds to. PSNR-M was created using HVS data from modern screens, but peripheral vision effects were not considered. The participants controlled the spatial aspects, flicker amplitude, and the brightness of the round sinusoidal lattice pattern [9]. Contrast thresholds were measured at 8 different spatial and 15 temporal frequencies at 3 different brightness levels (L). The resultant HVS model is reproduced in Fig. 1 [1].

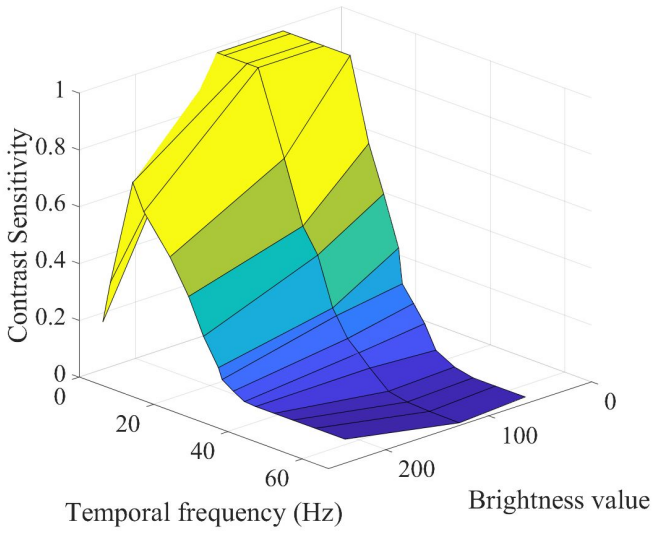

Fig. 1. The temporal- frequency response of the HVS with respect to brightness [1].

3 Video quality assessment considering the features of the human visual system: adaptation aspects, spatial and temporal frequency, eccentricity, luminance

We propose a PSNR based video quality assessment method that also incorporates measurements of the HVS, which we call PSNR-M+. The HVS factors included are spatial and temporal frequency, eccentricity (distance from the centre of fovea in visual degrees), luminance, adaptation aspects of the HVS, and stimulus size. This new metric is an extention of our earlier PSNR-M [1] which considered a small dataset on the dependence of separate spatial and temporal HVS characteristics on luminance, which was proven effective in comparison to PSNR [1].

3.1 Video Quality Assessment Metric

The proposed method centres around a weighted PSNR calculation which we proceed to elucidate,

$$\mathrm{PSNR}'(I(t), I_R(t), t) = \mathrm{PSNR}(I(t)K(t), I_R(t)K(t), t), \tag{1}$$

where I is a compressed frame, I_R is the reference uncompressed frame, and $K(t)$ is a weight coefficients matrix for t frames. The quality of the distorted video is measured incorporating both the spatio-temporal-luminance component and a peripheral component. A flow diagram framework of the methodology for calculating $K(x, y, t)$ is shown in Fig. 2.

In the spatio-temporal block of Fig. 2, we use the weight function $H_{L,f_t}(f_x, f_y)$ measured in earlier work [1], f_x, and f_y are spatial frequencies, f_t is temporal frequency, and L is luminance. This weight function models the ability of the human visual system to respond to spatio-temporal change, as measured via the CSF on a modern in-plane-switching LCD screen. The impulse response may be found via the following Fourier transform pair

$$h_{L,f_t}(x, y) \overset{F}{\Longleftrightarrow} (H_{L,f_t}(f_x, f_y)), \tag{2}$$

where $\overset{F}{\Longleftrightarrow}$ is the (invertible) Fourier transform. Filtering is performed in the spatial domain via.

$$I_R^{'}(x, y) = I_R(x, y) * h_{L,f_t}(x, y), \tag{3}$$

where $*$ is convolution. The spatio-temporal-luminance weighting factor, K_{stL}, is then computed as

$$K_{\mathrm{st}L}(x, y) = \frac{I_R^{'}(x, y)}{I_R(x, y)}. \tag{4}$$

The task of the region of interest (ROI) sampling is to identify objects that are more significant for the HVS. In this work, the ROI distinguishes individual objects using a variant of the watershed algorithm [20]. No more than five objects

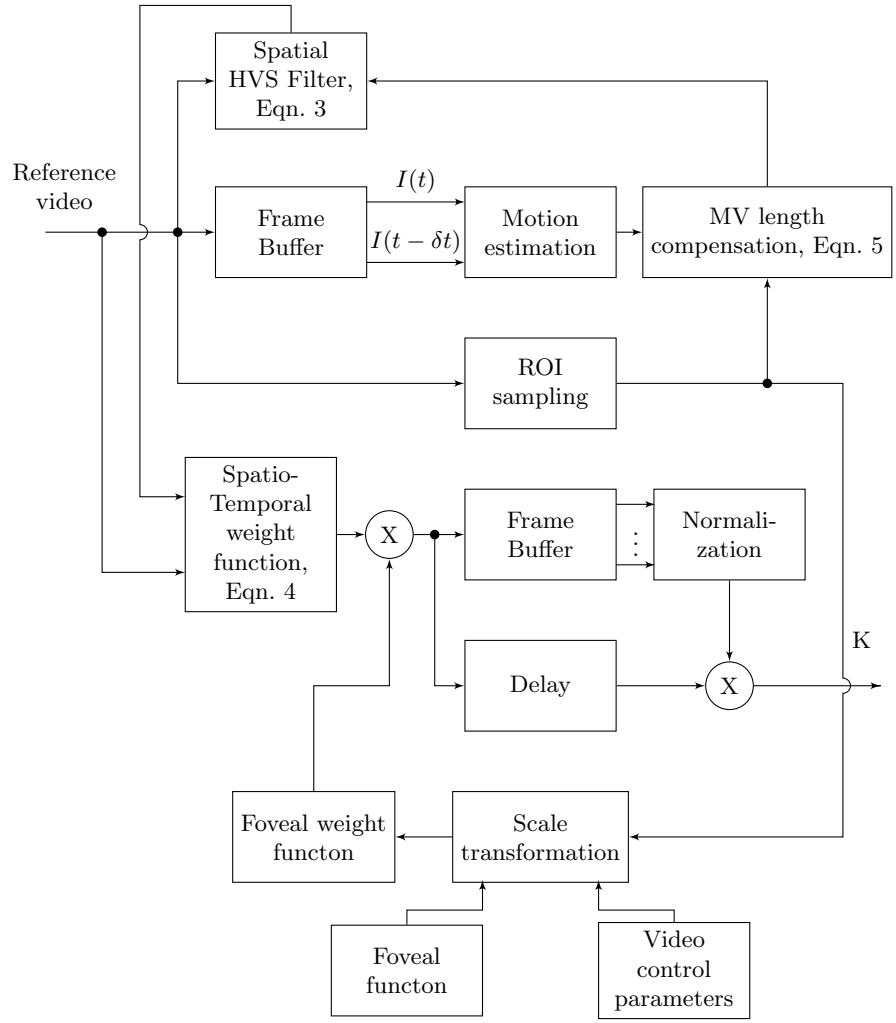

Fig. 2. The framework of the methodology for weight estimate.

are selected close to the centre, with fewer objects selected if the total area of the selected clustered object is greater than a user determined threshold [1].

The motion estimation utilises an adaptation of the MPEG block matching technique [19], which uses a 16×16 pixel block size, method with $32\times$ 32 pixel search area. The motion vectors, v(x,y,t), are "compensated" by subtracting the average within the ROI, v_{ROI}, viz.

$$v_{\mathrm{comp.}}(x, y, t) = v(x, y, t) - v_{\mathrm{ROI}}, \tag{5}$$

The peripheral coefficient is an approximate foveation function, which models the decrease of focus from the centre of the ROI outwards. (Foveation being

the blurring which increases from the centre of vision, which in our metric is accounted for by decreasing weight from the centre of the ROI.) The exact viewing angle, and hence the foveation function, depends on the screen size and resolution of the display device. Since the user can look at different places on the screen depending on the ROI, in order to find the viewing angle, we first find the centre of the ROI. Then, assuming that the user is at such a distance that there is 1 pixel in the center of the screen that corresponds to 1 minute of arc of vision [21], we find the viewing angle for the pixel as

$$\alpha(x,y) = \arctan\left(\frac{\pi\sqrt{(x-x_c)^2+(y-y_c)^2}}{1080}\right), \tag{6}$$

where, x_c, and y_c are the centre of the ROI, and 1080 is the number of pixels width of the screen.

For the foveation function, we take the characteristic of the spatial distribution of receptors as described by Gonzalez and Woods [20] and fit an approximate function to that distribution, viz.

$$A(\alpha) = \left(1.26\times10^5\right)e^{-0.71\alpha} + \left(1.5\times10^4\right) - 512\alpha + 11\alpha^2 - 0.08\alpha^3, \tag{7}$$

where we temporarily exclude the spatial dependency for brevity. (Note that the large numbers are normalised by the PSNR$'$ calculation below.) The visual resolution decreases from the region of interest to the periphery according to the above expression [22]. Then the peripheral coefficient in the weight function in the metrics is determined by

$$K_{pr}\,(x,y) = A(\alpha(x,y)). \tag{8}$$

In Fig. 3 we show an example of the $K_{\mathrm{pr}}(x,y)$ pattern.

The methodology for weight estimation developed and presented above, is introduced into the PSNR metric by weighting the function and the original video sequences via

$$\mathrm{PSNR}' = 20\log_{10}\frac{\sqrt{\sum_{t_n=-\frac{n}{2}}^{\frac{n}{2}}\sum_{x,y}K_{stL}^2\,(x,y,t_0+t_n)\,K_{pr}^2\,(x,y,t_0+t_n)}}{\sqrt{(n+1)\sum_{x,y}(I-I_R)^2K_{stL}^2\,(x,y,t_0)\,K_{pr}^2\,(x,y,t_0)}}, \tag{9}$$

where t_0 is the current frame, and t_n is the number of frames from the current frame, $n = \Delta t f_v$, Δt is the reaction time of the eye to frames (without considering the cognitive factors) [1], and f_v is the number of frames per second. In this work we took $\Delta t = 0.8$ seconds [23].

4 Methods

We compare the proposed metric to a range of metrics found in the literature and in current use. These metrics include: the most popular metrics at the current

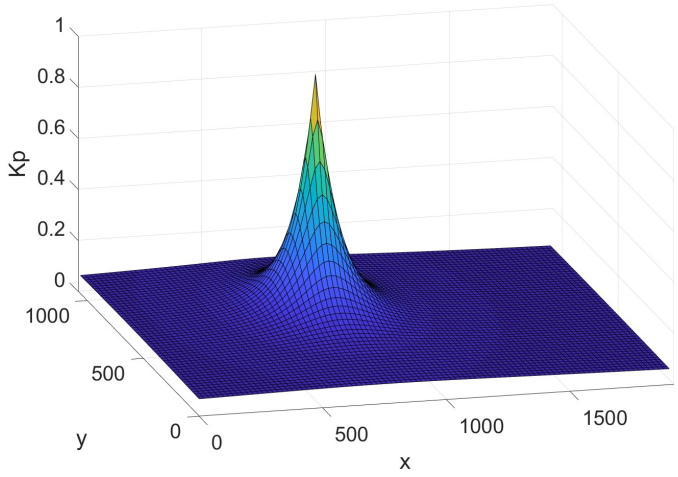

Fig. 3. An example of K_{pr}, where the frame size is 1920x1080 and the centre of the ROI is set to (760, 640).

moment (PSNR, SSIM); video multimethod assessment fusion (VMAF) [25], a method based on machine learning on databases of real users' evaluations of the quality of training videos [26]; spatio-temporal reduced reference entropy difference (STRRED) [27] and HDR-VQM [28], which are metrics that address temporal aspects; and metrics which considered the HVS visibility models such as PSNR-M and FovVideoVDP.

The LIVE-NFLX [29, 30] was used to test and compare the video quality metrics. The LIVE-NFLX data set consists of 112 compressed (hence distorted) videos. In this work we use 12 of these videos. The LIVE-NFLX database is selected because it represented highly realistic content with quality of experience responses to various design dimensions, including varying compression rates in the form of simulated varying transmission video bit rates over the course of each video.

5 Results and Discussion

The mean opinion scores (MOS), or subjective score, provided by the LIVE-NFLX database were used to compare the performance of PSNR-M+ with the presented above VQM. The scatter plots of the VQM algorithms under comparison are shown in Fig. 4. In Fig. 5, the metrics and their Pearson's linear correlation coefficient (PLCC) with the subjective quality scores are compared. As can be seen Fig. 5, the proposed metric gives the most consistent positive correlation. The gain in performance is mostly due to PSNR-M+'s ability to generalize predictions across video sequences. The strong compression distortion represented in this dataset and the difficulty of replicating subjective scores are

Table 1. Comparison of the video quality metrics correlating and not correlating with HVS, Pearson correlation coefficient (PLCC).

VQA	NFLX C4S1	NFLX C4S2	NFLX C6S1	NFLX C6S2	FCVR C1	FCVR C2
PSNR	0.577	0.117	0.648	-0.284	0.999958	0.999990
SSIM	0.623	0.075	0.715	-0.314	0.999961	0.999997
FVVDP	0.567	0.112	0.654	-0.284	0.999993	0.999999
STRRED	-0.248	-0.029	-0.451	-0.232	0.644864	0.836653
HDR-VQM	0.562	0.093	0.537	-0.178	0.999936	0.999936
PSNR-M	0.55	0.134	0.642	-0.191	0.999858	0.999891
PSNR-M+	0.613	0.803	0.607	0.572	0.999933	0.999945

apparent in the correlations herein. For example, in the scatter plots in Fig. 4, HDR-VQM predicts a comparable correlation between the prediction and real subjective score estimates for only half of the studied video sequences. HDR-VQM does a good job predicting spatial processing but does not model temporal processing.

It is appropriate to compare PSNR-M+, in which the model of HVS is derived from data based on modern LCD screens, to FovVideoVDP [3] which was based on much older data derived from experiments using cathode ray tube (CRT) screens. Both metrics consist of CSF models. In 2022, a visibility model StelaCSF (an improved FovVideoVDP model, the exact model we test herein) was presented based on 11 data sets that also use the older CRT screen data [8]. Their results show that stelaCSF can explain current data sets better than other previously existing models. FovVideoVDP, as well as PSNR-M, show variable results in low contrast objects and low bitrate. When comparing FovVideoVDP with PSNR-M+, the metrics show approximately similar results for video sequences with a normal bitrate. However, for videos containing sufficient motion, PSNR-M+ gives an average of 15% higher correlation.

The primary purpose of the analysis is to observe the behaviour of current VQM measures when deployed for predicting the perceptual quality of video content. From Fig. 5, the new HVS data based model, taken on modern equipment, gives an increase in the stability of the correlation of the obtained metric value with the real subjective score when evaluating video quality. Fig. 5 compares the correlation intervals of video quality metrics with the real subjective score on video sequences. PSNR-M+ has a better correlation interval than the VQM algorithms under comparison. Or in other words, the metric more effectively predicts the perception of videos, with different content and distortions, by the human visual system.

Despite the objectively better stability of the developed method, there is still potential room for improvement. For example, the motion compensation is coarse and could be refined to better reflect the motion of individual moving objects. The size of objects selected by the ROI stage could also be refined. The use of more videos from more disparate data sets will also enhance our confidence in our correlation scores. Moreover, our HVS model is limited and

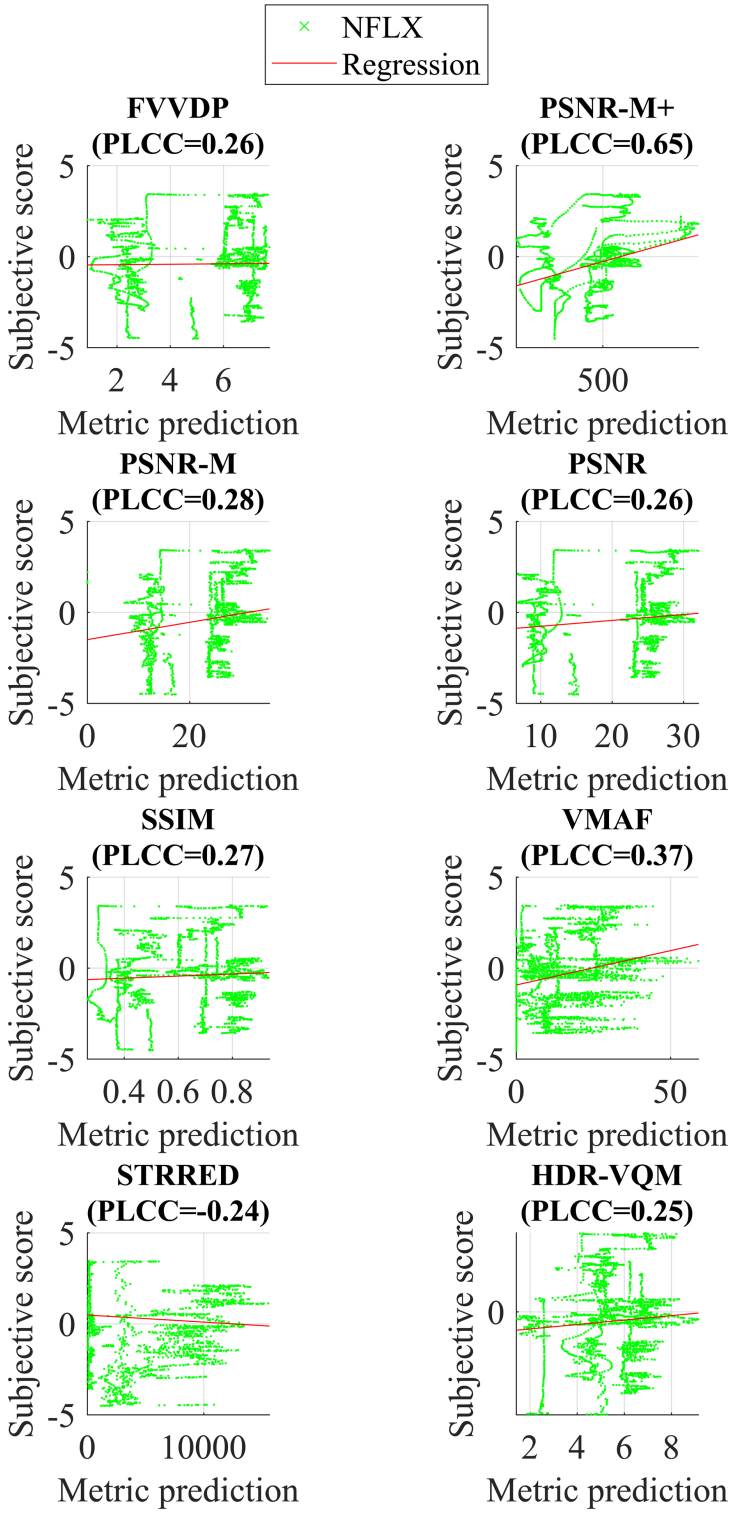

Fig. 4. Visualisation of LIVE-NFLX database results.

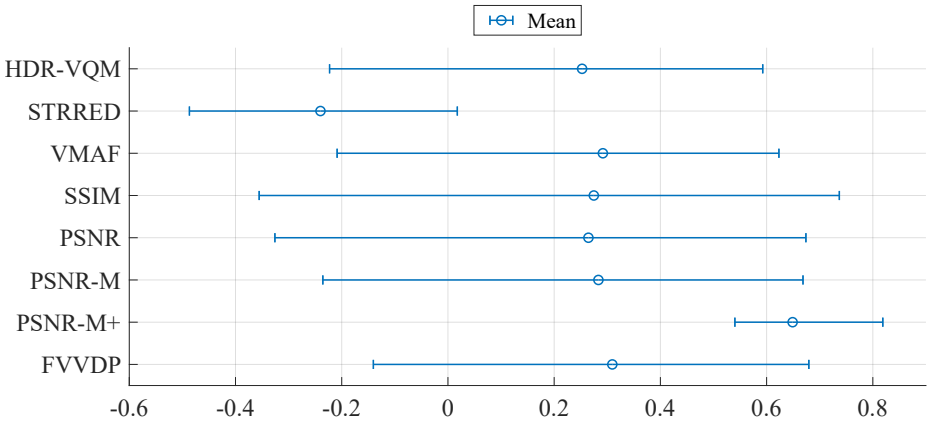

Fig. 5. Correlation interval of video quality metrics on video sequences LIVE-NFLX. The new proposed metric, PSNR-M+, has the most consistent high correlation of the metrics tested herein.

does not simulate certain aspects of vision such as inter-channel masking, eye movement, and a model of peripheral vision.

6 Conclusion

Our work demonstrates that metrics based on psychophysical HVS models explain human perception of video quality, outperforming statistically based metrics. In this test the proposed metric, which extends PSNR by incoporating recent data of the HVS taken using modern equipment, is comparable to the best complex algorithmic metrics. However, our HVS model is still somewhat limited and we plan to increase the volume of data that informs our CSF model, and to incorporate other aspects such as inter-channel masking, eye movement and peripheral vision. Also, testing the proposed metric against, at least, four independent databases with a minimum of 15 videos each and related statistics will be carried out in future work.

References

1. A. Mozhaeva, I. Vlasuyk, A. Potashnikov, L. Streeter, "Full Reference Video Quality Assessment Metric on Base Human Visual System Consistent with PSNR," *2021 28th Conference of Open Innovations Association (FRUCT)*, pp. 309-315, 2021.
2. P. Mohammadu, A. Ebrahimi-Moghadam, S. Shirani, "Subjective and Objective Quality Assessment of Image: A Survey," *Majlesi Journal of Electrical Engineering*, vol. 9, no. 1, pp. 55–83, 2015.
3. R. Mantiuk, G. Denes, A. Chapiro, A. Kaplanyan, G. Rufo, R. Bachy, T. Lian, and A. Patney, "Fovvideovdp: a visible difference predictor for wide field-of-view video," *ACM Transactions on Graph* , vol.40, pp.1-19, 2021.

4. Z. Ying, H. Niu, P. Gupta1, D. Mahajan, D. Ghadiyaram, A. Bovik, "From Patches to Pictures (PaQ-2-PiQ): Mapping the Perceptual Space of Picture Quality," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 3575–3585.
5. J. Kulikowski, "Some stimulus parameters affecting spatial and temporal resolution in human vision," *Vision Research*, vol. 11, pp. 83-93, 1971.
6. A. Watson, "High Frame Rates and Human Vision: A View Through the Window of Visibility," *SMPTE Motion Imaging Journal*, vol. 122, pp. 18-32, 2013.
7. A. Watson, A. Ahumada, "The pyramid of visibility," *Journal of Vision*, vol. 16, no. 12, pp. 567, 2016.
8. R. Mantiuk, M.Ashraf, A. Chapiro, "stelaCSF - A Unified Model of Contrast Sensitivity as the Function of Spatio-Temporal Frequency, Eccentricity, Luminance and Area," *ACM Transactions on Graphics (Proc. of SIGGRAPH 2022)*, vol. 41, no. 4, article no. 145, 2022.
9. A. Mozhaeva, I. Vlasuyk, A. Potashnikov, M. J. Cree, L. Streeter, "The Method and Devices for Research the Parameters of The Human Visual System to Video Quality Assessment," *2021 Systems of signals generating and processing in the field of onboard communications*, pp 1–5, 2021.
10. Z. Wang, A. Bovik, "Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures," *IEEE Signal Processing Magazine*, vol.26(1), pp.98-117, 2009.
11. K. Seshadrinathan, A. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE transactions on image processing* , vol.19, no. 2, pp.335-350, 2009.
12. National Research Council,*Human Performance Models for Computer Aided Engineerings*, The National Academies Press, 1989
13. P. Barten, "Formula for the contrast sensitivity of the human eye," *Proceedings Volume 5294, Image Quality and System Performance*, pp. 231–238, 2003.
14. S. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," *Proc. SPIE 1666, Human Vision, Visual Processing, and Digital Display III*, 1992.
15. D. Kelly, "Motion and vision. II. Stabilized spatio-temporal threshold surface," *Journal of Optical Society of America*, vol. 10, no. 10, pp.1340–1349, 1979.
16. H. de Lange, "Research into the dynamic nature of the human fovea-cortex systems with intermittent and modulated light. attenuation characteristics with white and colored light," *Journal of Optical Society of America*, vol. 48, no. 11, pp.777–784, 1958.
17. F. Campbell, G. Cooper, and J. Robson, "Application of Fourier analysis to the visibility of gratings," *The Journal of Physiology*, vol. 179, no. 3, pp.551–566, 1968.
18. F. van Nes and M. Bouman, "Spatial modulation transfer in the human eye," *Journal of Optical Society of America*, vol. 57, no. 3, pp.401–406, 1967.
19. A. Potashnikov, I. Vlasuyk, I. Augstkaln "Analysis of methods for detecting moving objects of different types on video image," *Fundamental problems of radio electronic instrumentation*, pp. 1201–1204, 2017.
20. R. Gonzalez, R. Woods "Digital Image Processing," *Technosphere Moscow*, 2012.
21. K. Poynton "Digital Video and HD. Algorithms and Interfaces," 2nd Edition, *Morgan Kaufmann Publishers*, MA, USA, 2012.
22. I. Vlasuyk "Development of a model of the human visual system for the method of objective image quality control in digital television systems," *Telecommunications and Transportation*, pp 189–192, 2009.
23. D.H. Hubel, "Eye, brain, vision," *Moscow: Mir*, 1990.

24. A. Mozhaeva, A. Potashnikov, I. Vlasuyk, and L. Streeter, "Constant Subjective Quality Database: The Research and Device of Generating Video Sequences of Constant Quality," *International Conference on Engineering Management of Communication and Technology (EMCTECH)*, Vienna, Austria, pp. 1–5, 2021.
25. L. Tsung-Jung, L. Yu-Chieh, L. Weisi and C.-C. Jay Kuo, "Visual quality assessment: recent developments, coding applications and future trends", *APSIPA Transactions on Signal and Information Processing* vol. 2, no. 1, 20pp, 2013.
26. Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
27. R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 4, pp. 684–694, 2012.
28. M. Narwaria, M. P. Da Silva, P. Le Callet, "HDR-VQM: An objective quality measure for high dynamic range video.," *Signal Processing: Image Communication*, vol. 35, pp 46–60, 2015.
29. C. G. Bampis, Z. Li, A. K. Moorthy, I. Katsavounidis, A. Aaron, and A. C. Bovik, "Study of Temporal Effects on Subjective Video Quality of Experience. ," *IEEE Transactions on Image Processing*, vol. 26, no. 11, pp. 5217–5231, 2017.
30. C. G. Bampis, Z. Li, A. K. Moorthy, I. Katsavounidis, A. Aaron and A. C. Bovik, *"LIVE Netflix Video Quality of Experience Database,"* [Online]. Available: //live.ece.utexas.edu/research/LIVE/NFLXStudy/index.html
31. Y. Jin, M. Chen, T. Goodall, A. Patney, and A. C. Bovik, "Subjective and objective quality assessment of 2D and 3D foveated video compression in virtual reality." *IEEE Transactions on Image Processing"*, to appear.
32. Y. Jin, M. Chen, T. Goodall, Z. Wan, and A. C. Bovik, "Study of 2D foveated video quality in virtual reality." *Proc. SPIE 11510, Applications of Digital Image Processing XLIII,"* , pp. 18–26, 2020.
33. Y. Jin, M. Chen, T. Goodall, A. Patney, and A. C. Bovik, "LIVE-Facebook Technologies-Compressed Virtual Reality (LIVE-FBT-FCVR) Databases." *Online: http://live.ece.utexas.edu/research/LIVEFBTFCVR/index.html, 2019.*