Image quality assessment across viewing distances: A comparison study of CSF-based and rescaling-based metrics

***Dounia Hammou*[1], *Lukáš Krasula*[2], *Christos G. Bampis*[2], *Zhi Li*[2], *Rafał K. Mantiuk*[1]**

[1]*Department of Computer Science and Technology, University of Cambridge, UK*
[2]*Netflix Inc., Los Gatos, CA, USA*

Abstract

Assessing the quality of images often requires accounting for the viewing conditions - viewing distance, display resolution and size. For example, the visibility of compression distortions may differ substantially when a video is viewed on a smartphone from a short distance and when viewed on a TV from a large distance. Nonetheless, traditional metrics are limited when applied across a diverse range of users with a diverse range of viewing environments. Metrics that account for these viewing conditions typically rely on contrast sensitivity functions (CSFs). However, it is also possible to rescale the input images to account for the change in the viewing distance. In a recent study comparing these two types of metrics, the authors did not observe any statistical difference between the metrics. Hence, in this paper, we use Fourier analysis to study the similarities and differences between the mechanisms of CSF-based and rescaling-based metrics. We compare the behaviour of each approach and investigate the correlation between their predictions across the viewing distances. Our findings demonstrate a similarity between the two approaches for high-frequency distortions when the viewing distance is increased (the images are downscaled), but not when the viewing distance is decreased (the images are upscaled) or when accounting for low-frequency distortions.

Introduction

Perceptual image quality assessment plays an important role in assessing the degradations introduced in various stages of image processing, allowing the optimization of image delivery for each observer. However, modern-day users can access images across a spectrum of devices, including TVs, PCs, tablets, smartphones, VR headsets and more. These devices differ substantially in their screen size, resolution and viewing conditions. Consequently, identical images or videos appear differently when viewed on different devices or in different viewing environments [4, 31].

For example, the perceptual quality of a video displayed on a 6" smartphone is substantially different than the same video displayed on an 8K TV display. Such differences must be accounted for when assessing the perceptual quality of images and videos. However, only a limited number of metrics account for the display and viewing conditions and their effect on the visibility of distortions. Most metrics that do so are based on the contrast sensitivity function of the human visual system (HVS) [5]. Nonetheless, other metrics can be adapted for the different display and viewing conditions by rescaling the input images [14, 31].

In a recent study [17], we conducted a performance evaluation of both types of metrics aiming to determine which metric is the best for predicting the quality across viewing distances. However, our findings showed no statistical significant evidence of one metric performing better than the other.

In this paper, we study the behaviours of both CSF-based and rescaling-based quality metrics when predicting image quality across the viewing distance. In particular, we explore the differences and similarities between the two approaches and investigate the effect of these differences when predicting the quality across viewing distances.

Related Works

The effect of viewing distance on the perceptual quality of images and videos

Previous studies investigated the effect of viewing distance on the perceptual quality of images and videos. In particular, the visibility of distortions, such as the noise, blur and compression.

Most of the studies conducted on images [10,14,21] reported similar findings: an increase in viewing distance results in a decrease in the visibility of distortions, thus an increase in the quality of distorted images. In contrast, an increase in viewing distance results in a decrease in the quality of undistorted (reference) images. Keller et al. [18], Sugito et al. [28], Amirpour et al. [2] and Lachat et al. [20] analysed the effect of viewing distance on the perception of compressed and/or upscaled videos. Similarly, these studies reported the same findings, where the perceptual quality of distorted videos increased with the increase in viewing distance.

However, a study by Mikhailiuk et al. [25] reported a different observation. The authors measured the visually lossless thresholds at which image compression artefacts become imperceptible. The study was conducted across viewing distances and display brightness levels. The results showed that the effect of viewing distance is content-dependent. While some images' results aligned with the pattern previously observed, others reported an increase in distortions' visibility with the increase in viewing distance.

Given the limited research on the effect of viewing distance on the visibility of image and video distortions, there is not enough evidence to fully understand how this factor impacts the image and video quality.

Quality metrics sensitive to viewing distance

Most metrics that account for the effect of viewing distance do so by modelling the spatial sensitivity via contrast sensitivity function (CSF). A CSF [5] models the smallest contrast detectable by an average observer. It serves as a mathematical representation of the human visual system's sensitivity to contrast across different spatial frequencies, thereby allowing the estimation of perceived distortions at different viewing distances, display sizes, and display resolutions.

Different quality metrics employ the CSFs in various manners. Some utilize it as a linear filter on compared images [8, 32], while others assign weights to coefficients in the wavelet transform of images [6, 11, 30]. More recently, CSFs are commonly employed to weight the bands of a band-pass decomposition [23, 24, 26]. However, due to the contrast constancy of the visual system, the CSF alone cannot accurately model the discrimination of contrast differences above the detection threshold [12]. Therefore, the CSF is not commonly employed as a linear filter but is combined with models of contrast masking [23, 24] to achieve more accurate scores.

Other metrics can be adapted to the different viewing distances and display sizes by rescaling the input images [3, 14, 16, 17] or by pre-filtering them in the wavelet domain [14, 15].

Recent advancements in deep learning techniques have led to the development of models that incorporate viewing distance as a feature [7, 31]. However, due to the limited available datasets, learning-based metrics have not yet proven to perform well in this task.

In a recent study [17], we investigated the performance of CSF-based rescaling-based metrics using existing available datasets, including VDID2014 [14], CID:IQ [21], VCIP21 [2] and VLIC [25]. The results reported no statistical difference between the two types of metrics.

In contrast to these prior works, we do not collect a new dataset or test the performance of a new metric for the task of predicting quality across viewing distances. Instead, in this paper, we are interested in studying the similarities and differences between the behaviour of metrics that rely on either CSFs or rescaling when accounting for changes in viewing distances. This study aims to provide a more in-depth understanding of the underlying mechanisms of these two approaches.

Adapting image quality metrics for varying viewing distances

Image quality metrics can be adapted for varying viewing distances by pre-filtering compared images. The filtering methods can be broadly categorized into two groups: CSF-based and rescaling-based methods.

CSF-based image quality metrics

Contrast sensitivity functions (CSFs) provide the smallest detectable contrast, known as the contrast threshold. This measure allows the assessment of the visual system's sensitivity to contrast across many dimensions, including spatial frequencies, temporal frequencies, color, eccentricity, luminance and area size [22]. Because of its characteristics, CSFs are often employed in visual difference predictors and quality assessment metrics to estimate the visibility of distortions across these various dimensions. Because the focus of this study is to analyse the behaviour

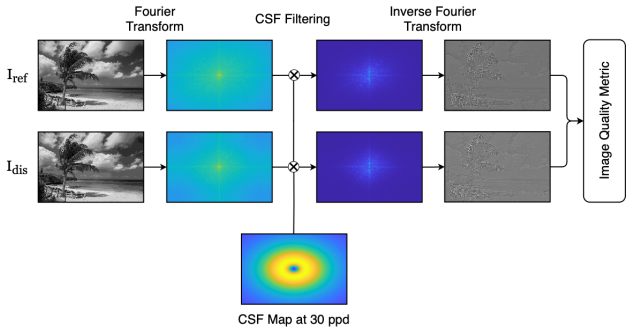

Figure 1. *Overview of the CSF-based quality assessment metric. The input is the reference and distorted grayscale images. The images are transformed to the Fourier space, and then filtered using a CSF map at a 30 ppd. The CSF-modulated images are then utilized as inputs of an existing quality metric to predict the quality score at 30 ppd.*

of the CSF filter when predicting image quality across viewing distances, we employ the stelaCSF [22] and consider the spatial frequencies dimension only. We set the temporal frequency to 0 Hz, the mean luminance to 20 cd/m^2, the area size to 1 deg^2 and the eccentricity to 0 deg.

CSFs can be employed for image quality metrics in different ways as discussed in the related work section above. But, to directly compare it to rescaling, we use CSF to pre-filter the compared images. The filtering is performed in the Fourier domain, as illustrated in Figure 1 and described in the following sections.

Input images Because the sensitivity of the visual system to colors is different than its sensitivity to luminance, CSFs are employed on separate channels when predicting the quality of color images. In particular, on the luminance, red-green, and blue-yellow channels [32].

Because our study focuses on the effect of viewing distance, we will not include the color dimension. Hence, the RGB images are converted to grayscale images.

Viewing distance To account for the viewing distance using the contrast sensitivity function, it is essential to model how the viewing distance affects visual resolution on the retina, expressed in the units of pixels per visual degree. We will refer to that measure as the effective resolution and compute it as:

$$n_{\text{ppd}} = \frac{\pi\, r_y}{360 \operatorname{atan}\left(\frac{0.5h}{d}\right)} \tag{1}$$

where r_y is the display's vertical resolution in pixels, h is the display height and d is the viewing distance, both in the same units (e.g. meters).

In the frequency domain, the corresponding Nyquist frequency is expressed as half of the effective resolution in the units of cycles per visual degree [23].

CSF filter The contrast sensitivity function is employed to filter the Fourier transform of the image. However, a CSF requires the spatial frequencies to be in the units of cycles per visual degree.

Therefore, the effective resolution is utilized to transform the frequencies from the units of cycles per pixel to the units of cycles per visual degree.

CSF-modulated image The CSF-modulated image can be expressed as follows:

$$I' = \Re\left(\mathfrak{F}^{-1}\left\{\mathfrak{F}\{I\}(\rho)\cdot\frac{S(\rho)}{\max\limits_{\rho}(S(\rho))}\right\}\right) \tag{2}$$

where I' is the filtered image, I is the input image, $\mathfrak{F}$ is the Fourier transform, $\Re$ is the real part of a complex number, ρ is the spatial frequency in cycles per visual degree, and S is the contrast sensitivity function normalized to 1 by its maximum value.

Rescaling-based image quality metrics

Reconstruction function (or filter) is a function employed to estimate a continuous image given a set of discrete samples [13], for example, when resampling an image to different resolutions. Such a function can be used for both the interpolation (to infer missing pixel values) or for low-pass filtering of an image before it is downsampled. There exist several reconstruction functions also known as rescaling functions such as the bilinear [27], bicubic [19] and Lanczos [9] functions.

Bilinear function The bilinear function is one of the fundamental rescaling functions that linearly interpolates between two adjacent samples in one dimension. The rescaling kernel is expressed as

$$K(x) = \Lambda(x) \tag{3}$$

where Λ is the triangular impulse function.

In practice, when rescaling images, the function calculates the weighted average of the four neighbouring pixels by linearly interpolating across both dimensions independently.

Bicubic function In contrast to the bilinear function, the bicubic function employs the cubic kernel to offer a smoother interpolation, while also considering sixteen neighbouring pixels. The kernel is expressed as

$$K(x) = \begin{cases} \frac{3}{2}|x|^3 - \frac{5}{2}|x|^2 + 1 & \text{for } |x| \leq 1 \\ -\frac{1}{2}|x|^3 + \frac{5}{2}|x|^2 - 4|x| + 2 & \text{for } 1 < |x| < 2 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Lanczos function In contrast to other functions, the Lanczos function is the most similar to the sinc function, which is the ideal low pass filter (but which cannot be used because of infinite support). It employs a Lanczos kernel: a sinc function windowed by the sinc window. The kernel is expressed as

$$K(x) = sinc(x)sinc\left(\frac{x}{a}\right)\Pi\left(\frac{x}{2a}\right) \tag{5}$$

where Π is the rectangular impulse and $2a$ is the window size that determines the number of neighbouring pixels to be considered. In most cases, a is set to 3.

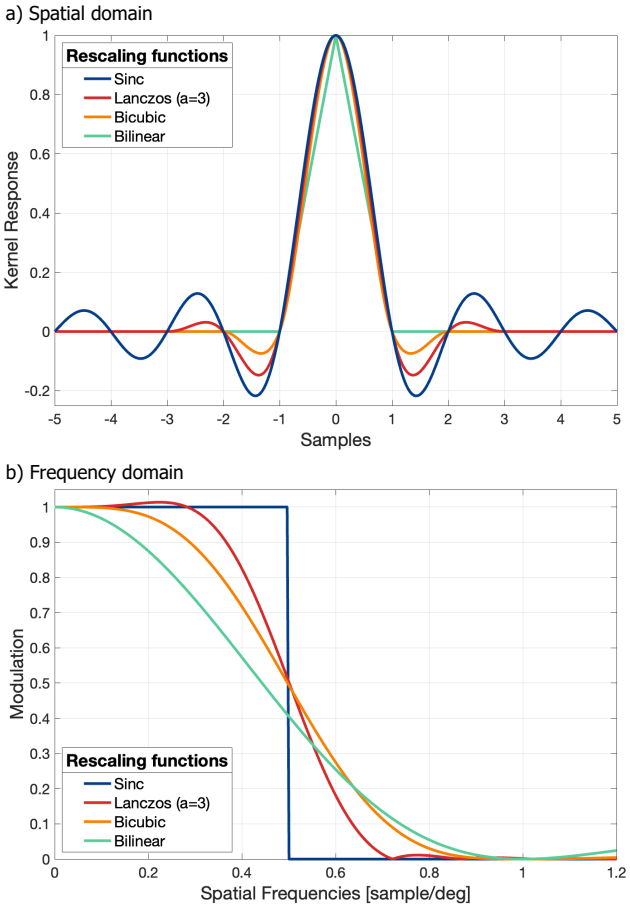

Figure 2. *Comparison between the rescaling functions in the spatial domain (top) and frequency domain (bottom). The dark blue curve represents the sinc function, the ideal low pass filter, and the other curves represent the described rescaling functions.*

An illustration of the rescaling functions in both spatial and Fourier domains is shown in Figure 2.

Image rescaling can account for the viewing distance by simulating the change of an image projected on the retina [14, 17, 31] — the retinal image will get proportionally larger as the viewing distance gets smaller, as shown in Figure 3. Here, we take a different view of the rescaling process. An image rescaling function is a low-pass filter, which can resemble the right portion (low-pass) of the CSF. If we can find the right cut-off frequency of the rescaling filter, we can make rescaling closely approximate the low-pass behaviour of the CSF.

We vary the cut-off frequency for each rescaling function to minimize the difference between the function and the right portion (from its peak) of the CSF. The selected cut-off frequencies in cycles per degree (cpd), as well as the rescaling functions, are illustrated in Figure 4.

Hence, given the cut-off frequency of the rescaling filter, the rescaling factor is given by:

$$z = \frac{r_{\text{ppd}}}{n_{\text{ppd}}} = \frac{2f_{\text{cutoff}}}{n_{\text{ppd}}} \tag{6}$$

where n_{ppd} is the image's effective resolution, f_{cutoff} is the rescal-

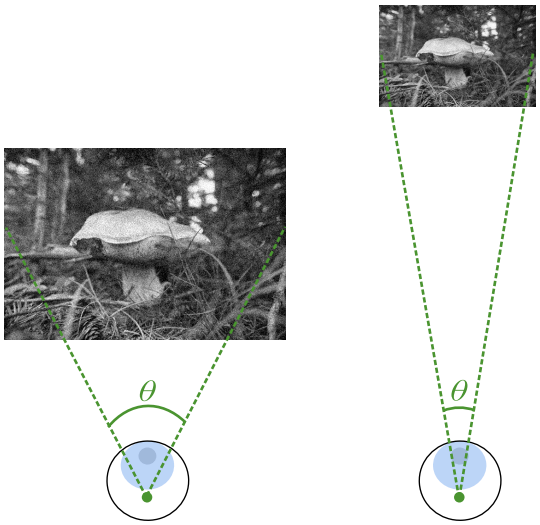

Figure 3. *Illustration of the effect of viewing distance on the visibility of distortions. As shown in the figure, the larger the viewing distance, the smaller the viewing angle, and the less visible the distortions. This behaviour is modelled by rescaling the images appropriately for each viewing distance.*

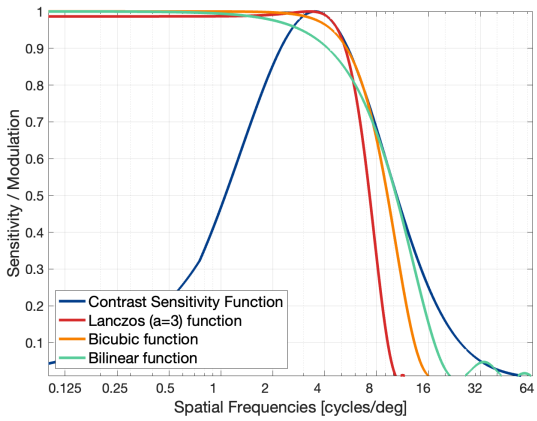

Figure 4. *Comparison between the contrast sensitivity function and the rescaling functions in the frequency domain. The cut-off frequency of each rescaling function is selected so that the rescaling curve matches the CSF at high frequencies. The selected cut-off frequency for each rescaling function is set as follows: the Lanczos3 function: $f_{cutoff} = 7.3$ cpd, the bicubic function: $f_{cutoff} = 9.3$ cpd, and the bilinear function: $f_{cutoff} = 11.7$ cpd.*

ing filter cut-off frequency, and r_{ppd} is the rescaling function optimal effective resolution. Note that this approach is different than what was done in previous work, which relied on finding the cut-off frequency in training [17] or ad-hoc assumptions [14, 31].

Comparison analysis of CSFs-based and rescaling-based quality metrics

In this section, we will analyse the similarities and differences between the rescaling-based and CSF-based quality metrics. In particular, their filter responses for different distortions across varying viewing distances. This process will allow us to comprehend the differences in their behaviours when accounting for the varying viewing distances. Furthermore, we will compare the quality predictions across the viewing distances using both types of metrics.

Frequency analysis

It is essential to understand the behaviour of both the contrast sensitivity and rescaling functions when filtering the distortions of the images at each viewing distance. Therefore, we perform a study analysis in the frequency domain by comparing the CSF, the spectral response of the rescaling functions, and the power spectrum of the distortions across the varying viewing distances.

To compute the power spectrum of the distortions, we select 40 images from the DIV2K dataset [1, 29], each degraded by seven distortions in five levels. The chosen distortions are the white Gaussian noise, Gaussian blur, JPEG, JPEG2000, resampling, display non-uniformity and adaptive scalable texture compression. We modelled the display non-uniformity by altering the brightness around the edges of the image, which was done by decreasing the luma uniformly with eccentricity. Furthermore, for each pair of reference I_{ref} and distorted I_{dis} image, we calculate the difference $D = I_{\text{ref}} - I_{\text{dis}}$. The difference is used as a representation of the distortion and consequently to calculate the power spectrum of each distortion at each level. The power spectrum of D is expressed as follows:

$$P(\rho_i) = \sum_{\rho_x,\rho_y \in \Omega} \left|F(\rho_x,\rho_y)\right|^2$$

$$\text{where} \quad \Omega = \left\{\rho_x,\rho_y : e_{i-1} \leq \sqrt{\rho_x^2+\rho_y^2} < e_i\right\} \text{ and } \rho_i = \frac{e_{i-1}+e_i}{2} \tag{7}$$

where P represents the power spectrum, ρ_i the spatial frequency in cycles per sample, ρ_x the horizontal frequency, ρ_y the vertical frequency, Ω the set of frequencies ρ_x, ρ_y in the bin with the mean frequency of ρ_i, e_{i-1} and e_i the edges of the bin, and F the Fourier transform of the distortion D, such that:

$$F(\rho_x,\rho_y) = \frac{\mathscr{F}\left(D(x,y)\right)}{w \cdot h} \tag{8}$$

where w is the width of D and h is the height of D. Note that we do not normalize the power by the number of elements in each bin. This way, the power spectrum has a direct relation to the sum of squared errors (the Plancherel theorem) and, therefore, to RMSE and PSNR metrics.

To account for the variability in the images' statistics for each distortion type and distortion level, we calculate the mean, first percentile and third percentile of the power spectrum of the differences of all 40 pairs. Then, we transform the spatial frequencies from the units of cycles per sample to the units of cycles per visual degree by including the effective resolution of the images.

We report the power spectrum of two distortions, the JPEG compression and the display non-uniformity, each at the third distortion level, for six effective resolutions. The effective resolutions are obtained by varying the viewing distance from 0.05 m to 2.5 m while assuming a 30" 1920×1080 display. The results are illustrated in Figure 5.

We observe from the figure that when an image is viewed from larger viewing distances or effective resolutions, as indicated by the green, yellow, orange and pink curves on the left,

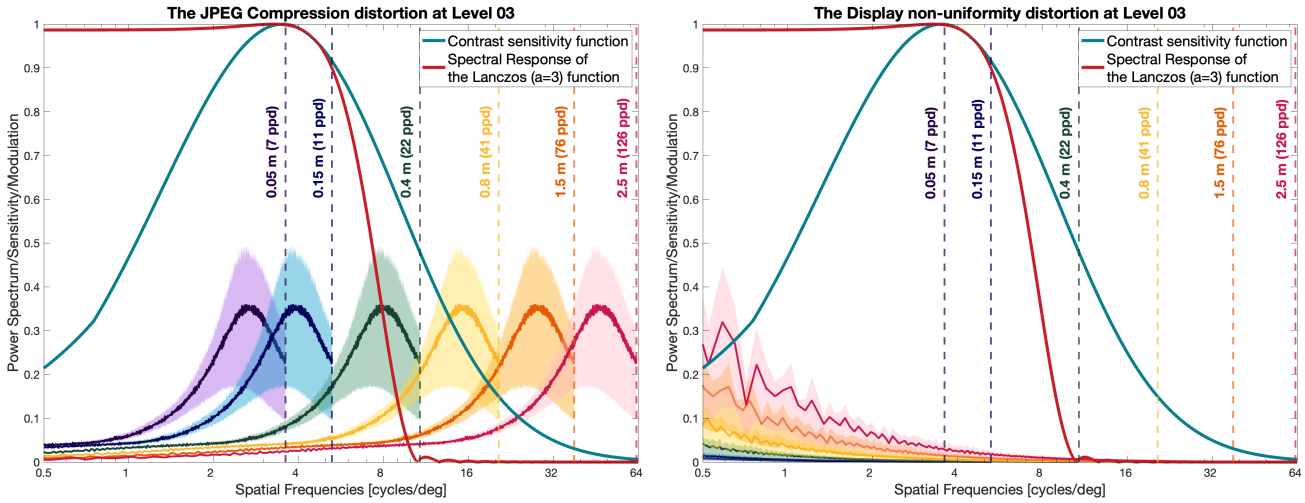

Figure 5. *Comparison of the power spectrum of the JPEG (left) and display non-uniformity (right) distortions for different viewing distances [m]/effective resolutions [ppd] with the CSF (dark cyan) and the Lanczos (a=3) filter (red) used for rescaling ($f_{cutoff} = 7.3$ cpd and $r_{ppd} = 14.6$ ppd). The CSF filter is normalized to 1, while the power spectrum of the distortion (colored curves in the bottom) is normalized to 0.5. The mean of the power spectrum is shown in a dark color, while the area between the mean and the two percentiles is filled with a transparent color. Both filters show similar behaviour at high frequencies but differ at low frequencies.*

the power spectrum of the distortion shifts toward higher frequencies (in [cycles/degree]). Because the CSF is less sensitive to high frequencies (dark cyan curve) and because of the low-pass filter nature of the rescaling function (red curve), both filters perform similarly in reducing the visibility of high-frequency distortions at larger viewing distances, such as the JPEG distortion. Nevertheless, for low-frequency distortions, such as the display non-uniformity distortion, it becomes evident that the rescaling filter and the CSF will perform differently when accounting for the viewing distance. Because in this scenario, the rescaling function does not filter the low frequencies, while the band-pass characteristic of the CSF makes the metric less sensitive to them.

Moreover, these differences between the CSF and rescaling filter at low frequencies (below 5 ppd) are also noticeable when the viewing distance is reduced and the power spectrum of the high-frequency distortions is shifted towards lower frequencies (violet and blue curves on the left). Notably, under such conditions, we can expect a difference in the metric predictions between the two approaches. Because for small viewing distances, the rescaling function will not apply any changes in the distortions; hence, it will not be able to account for the changes in small viewing distances. However, the sensitivity of the visual system decreases at these small viewing distances, which the CSF will be able to account for.

Analysis of metric' predictions

To validate our findings, we analyse the predictions of the two types of quality metrics across viewing distances/effective resolutions. We select the Peak-Signal-to-Noise Ratio (PSNR) metric, as it is the most widely used metric that is easy to adapt to both discussed approaches and because it is a pixel-based metric, it will allow us to better comprehend the effect of both filters independently from the effect of the quality metrics itself.

We compare the PSNR predictions of the JPEG compressed and images affected by screen non-uniformity using the CSF-based and rescaling-based metrics across ten viewing distances (effective resolutions from 5 to 120 ppd) while using the three described rescaling functions. The results are reported in Figure 6.

As can be observed in this figure, when accounting for the effect of larger effective resolutions than the rescaling method's effective resolution $r_{\text{ppd}} = 2f_{\text{cutoff}}$ (when downscaling the input images) on high-frequency distortions, such as the JPEG, both the CSF and the rescaling metrics behave similarly, as we can observe a very high correlation between the predictions especially using the Lanczos and bicubic functions (top row). We believe that is due to the fact that these two functions had the highest match with the CSF, as shown in Figure 4.

However, when accounting for smaller viewing distances (when upscaling the images), we can observe that the quality predictions remained constant for the rescaling-based metric. This is due to the fact that the rescaling function spectral response remains constant for low frequencies; thus, it does not include any new information to the image when upscaling it. However, the CSF was able to account for such small effective resolutions, as well as for the very low sensitivity of the visual system to very small viewing distances, where the quality increases.

Moreover, when accounting for the effect of the viewing distance on low-frequency distortions, such as display non-uniformity, we can observe that the quality remains constant for all effective resolutions. As explained beforehand, this is because the rescaling function spectral response remains constant for low frequencies; consequently, it does not account for the effect of viewing distance. In contrast, we can observe that the CSF is able to account for such distortion, and because of the low sensitivity of the visual system towards low frequencies, we can observe that, unlike the usually reported findings, the quality decreases with the increase of the viewing distance.

From these findings, we can argue that the rescaling-based metrics can mimic the human visual system sensitivity to contrast when accounting for larger viewing distances and for distortions

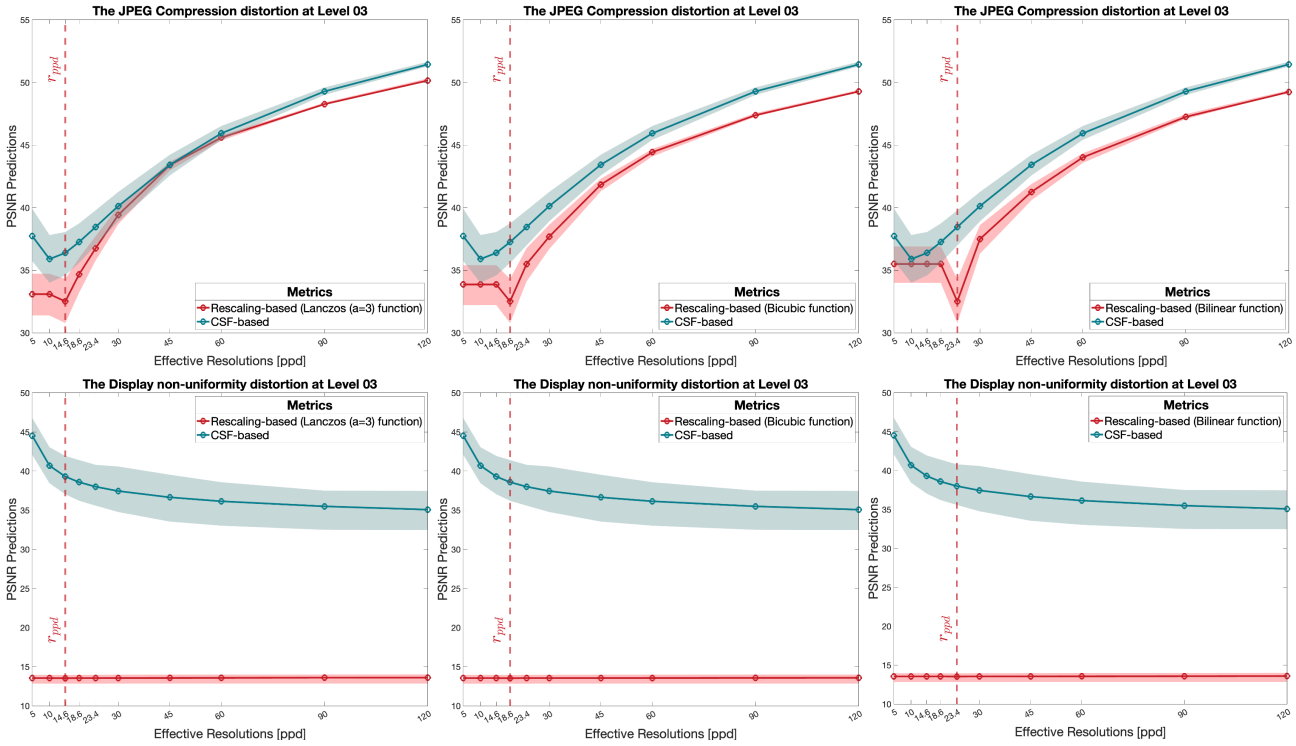

Figure 6. *PSNR quality scores comparison between the rescaling-based and CSF-based metrics for the JPEG and display non-uniformity distortions (rows) and the three rescaling functions (columns). The optimal cut-off frequency for each rescaling function is used and shown as the optimal effective resolution of the rescaling function r_{ppd} in dashed points. The graph that is on the left of the dashed lines shows the scenario of having smaller viewing distances (upscaling the images), while the graph on the right shows the scenario of larger viewing distances (downscaling the images).*

that affect mostly high frequencies; however, rescaling cannot account for smaller viewing distances or low-frequency distortions. In real-life scenarios, when watching the TV, smartphone, or a tablet, such small effective resolutions are not possible to reproduce, and low-frequency distortions are rarely produced by the standard compression methods, so we can argue that for such scenarios, the rescaling-based and CSF-based metrics behave similarly, which explains the results reported in [17].

Conclusions

In this paper, we investigated the behaviour of both CSF-based and rescaling-based filtering functions in the context of predicting image quality across varying viewing distances. Unlike previous works, our study was not centred on evaluating and comparing the performance of each metric. Instead, the focus of the paper has been on a comprehensive analysis and comparison of the underlying mechanisms of both approaches. Our results indicate a similarity between the behaviour of CSF and rescaling functions when accounting for the effect of larger viewing distances for high-frequency distortions. In such a scenario, both approaches reduce the importance of high frequencies. However, both approaches show different behaviour when accounting for smaller viewing distances or for low-frequency distortions. Here, the rescaling functions' spectral response stays constant while not accounting for the effect of the viewing distances.

Our findings offer valuable insights into the practical applications of these filtering techniques when accounting for the varying viewing distances. Specifically, they suggest that for accounting for the effect of large effective resolutions on most distortions (high-frequency distortions), as usually done when watching a TV, tablet, or smartphone, a simple rescaling function can be employed. Conversely, when accounting for smaller viewing distances or when predicting quality loss due to low-frequency distortions, CSF-based approach is more suitable. It must be noted that our analysis applies to very simple CSF-based metrics, which use CSF to prefilter compared images. The findings could be different for more advanced metrics, which account for contrast constancy and contrast masking.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017.

[2] Hadi Amirpour, Raimund Schatz, Christian Timmerer, and Mohammad Ghanbari. On the impact of viewing distance on perceived video quality. In *2021 International Conference on Visual Communications and Image Processing (VCIP)*, pages 1–5. IEEE, 2021.

[3] Christos G Bampis, Lukáš Krasula, Zhi Li, and Omair Akhtar. Measuring and predicting perceptions of video quality across screen sizes with crowdsourcing. In *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*, pages 13–18. IEEE, 2023.

[4] Nabajeet Barman, Yuriy Reznik, and Maria G Martini. A subjective dataset for multi-screen video streaming applications. *arXiv preprint arXiv:2305.03138*, 2023.

[5] Peter GJ Barten. *Contrast sensitivity of the human eye and its effects on image quality*. SPIE press, 1999.

[6] Andrew P Bradley. A wavelet visible difference predictor. *IEEE Transactions on image processing*, 8(5):717–730, 1999.

[7] Aladine Chetouani and Marius Pedersen. Image quality assessment without reference by combining deep learning-based features and viewing distance. *Applied Sciences*, 11(10):4661, 2021.

[8] Scott J Daly. Visible differences predictor: an algorithm for the assessment of image fidelity. In *Human Vision, Visual Processing, and Digital Display III*, volume 1666, pages 2–15. SPIE, 1992.

[9] Claude E Duchon. Lanczos filtering in one and two dimensions. *Journal of Applied Meteorology and Climatology*, 18(8):1016–1022, 1979.

[10] Ruigang Fang, Dapeng Wu, and Liquan Shen. Evaluation of image quality of experience in consideration of viewing distance. In *2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*, pages 653–657. IEEE, 2015.

[11] Xinbo Gao, Wen Lu, Dacheng Tao, and Xuelong Li. Image quality assessment based on multiscale geometric analysis. *IEEE Transactions on Image Processing*, 18(7):1409–1423, 2009.

[12] B Y M A Georgeson and G D Sullivan. Contrast constancy: deblurring in human vision by spatial frequency channels. *Journal of Physiology*, 252(3):627–656, 1975.

[13] Pascal Getreuer. Linear methods for image interpolation. *Image Processing On Line*, 1:238–259, 2011.

[14] Ke Gu, Min Liu, Guangtao Zhai, Xiaokang Yang, and Wenjun Zhang. Quality assessment considering viewing distance and image resolution. *IEEE Transactions on Broadcasting*, 61(3):520–531, 2015.

[15] Ke Gu, Guangtao Zhai, Min Liu, Qi Xu, Xiaokang Yang, Jun Zhou, and Wenjun Zhang. Adaptive high-frequency clipping for improved image quality assessment. In *2013 Visual Communications and Image Processing (VCIP)*, pages 1–5. IEEE, 2013.

[16] Ke Gu, Guangtao Zhai, Xiaokang Yang, and Wenjun Zhang. Self-adaptive scale transform for iqa metric. In *2013 IEEE international symposium on circuits and systems (ISCAS)*, pages 2365–2368. IEEE, 2013.

[17] Dounia Hammou, Lukáš Krasula, Christos G Bampis, Zhi Li, and Rafał K Mantiuk. Comparison of metrics for predicting image and video quality at varying viewing distances. In *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6. IEEE, 2023.

[18] Dominik Keller, Felix von Hagen, Julius Prenzel, Kay Strama, Rakesh Rao, Ramachandra Rao, and Alexander Raake. Influence of viewing distances on 8k hdr video quality perception. In *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*, pages 209–212. IEEE, 2023.

[19] Robert Keys. Cubic convolution interpolation for digital image processing. *IEEE transactions on acoustics, speech, and signal processing*, 29(6):1153–1160, 1981.

[20] Amélie Lachat, Jean-Charles Gicquel, and Jérôme Fournier. How perception of ultra-high definition is modified by viewing distance and screen size. In *Image Quality and System Performance XII*, volume 9396, pages 306–313. SPIE, 2015.

[21] Xinwei Liu, Marius Pedersen, and Jon Yngve Hardeberg. "cid: Iq–a new image quality database". In *Image and Signal Processing: 6th International Conference, ICISP 2014, Cherbourg, France, June 30–July 2, 2014. Proceedings 6*, pages 193–202. Springer, 2014.

[22] Rafał K Mantiuk, Maliha Ashraf, and Alexandre Chapiro. stelacsf: a unified model of contrast sensitivity as the function of spatio-temporal frequency, eccentricity, luminance and area. *ACM Transactions on Graphics (TOG)*, 41(4):1–16, 2022.

[23] Rafał K Mantiuk, Gyorgy Denes, Alexandre Chapiro, Anton Kaplanyan, Gizem Rufo, Romain Bachy, Trisha Lian, and Anjul Patney. Fovvideovdp: A visible difference predictor for wide field-of-view video. *ACM Transactions on Graphics (TOG)*, 40(4):1–19, 2021.

[24] Rafal K Mantiuk, Dounia Hammou, and Param Hanji. Hdr-vdp-3: A multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content. *arXiv preprint arXiv:2304.13625*, 2023.

[25] Aliaksei Mikhailiuk, Nanyang Ye, and Rafal K Mantiuk. The effect of display brightness and viewing distance: A dataset for visually lossless image compression. 2021.

[26] Abdul Rehman, Kai Zeng, and Zhou Wang. Display device-adapted video quality-of-experience assessment. In *Human Vision and Electronic Imaging XX*, volume 9394, pages 27–37. SPIE, 2015.

[27] PR Smith. Bilinear interpolation of digital images. *Ultramicroscopy*, 6(2):201–204, 1981.

[28] Yasuko Sugito, Yuichi Kondo, Daichi Arai, and Yuichi Kusakabe. Modeling perceived quality on 8k vvc video under various screen sizes and viewing distances. *IEEE Access*, 10:97237–97247, 2022.

[29] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 114–125, 2017.

[30] Abhinau K Venkataramanan, Cosmin Stejerean, and Alan C Bovik. Funque: Fusion of unified quality evaluators. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 2147–2151. IEEE, 2022.

[31] Nanyang Ye, Krzysztof Wolski, and Rafal K Mantiuk. Predicting visible image differences under varying display brightness and viewing distance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5434–5442, 2019.

[32] Xuemei Zhang, Brian A Wandell, et al. A spatial extension of cielab for digital color image reproduction. In *SID international symposium digest of technical papers*, volume 27, pages 731–734. Citeseer, 1996.