TEMPORAL MOTION SMOOTHNESS AND THE IMPACT OF FRAME RATE VARIATION ON VIDEO QUALITY

Rasoul Mohammadi Nasiri, Zhengfang Duanmu, and Zhou Wang

Dept. of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada
Emails: {r26moham, zduanmu, zhou.wang}@uwaterloo.ca

ABSTRACT

There has been a strong recent trend to improve the perceptual quality-of-experience of viewers by expanding the spatial resolution, dynamic range, color gamut, and frame rate of videos. Conceptually, increasing video frame rate should create a benefit of smoother perception of motion. However, how to measure motion smoothness is not a well resolved problem. In this study, we measure the smoothness of motion by examining the local phase correlation of complex wavelet coefcients along the temporal direction. Our experiments based on subjective-rated databases show that this novel measure provides a new means to capture the impact of frame rate on video quality, and demonstrates strong promise at improving the performance of objective video quality assessment models.

Index Terms— high frame rate, motion smoothness, natural scene statistics, local phase, video quality assessment, quality-of-experience, complex wavelet transform.

1. INTRODUCTION

With the technological advances in video acquisition and display devices, higher frame rate videos of 60 frame-per-second (fps) or higher are becoming increasingly more popular in recent years. Increase in frame rate alongside with increase in resolution or extension of color gamut create several means to provide higher quality moving pictures to end users. It has been believed that higher frame rate produces more natural and smoother motion for the viewers. However, there is not enough investigation to evaluate this improvement. On the other hand, as the video constitutes the majority of data traffic in today's communication networks [1], and frame rate, alongside resolution and quantization parameter, is one of the most effective rate control parameters, understanding the impact of frame rate changes on motion quality in video has become ever more important.

Despite the growing popularity of high frame rate videos, our understanding of human quality-of-experience (QoE) behaviors in the temporal dimension remains rather limited. Traditional services equate temporal quality to frame rate, leading to inaccurate QoE assessment. Specifically, videos at the same frame rate could have drastically different quality, which is further perplexed with bitrate and spatial resolution. To make effective and efficient use of high frame rate technology, it is important to thoroughly understand the impact of temporal smoothness in end-users' QoE.

The existing temporal video quality assessment models can be roughly categorized into data-driven and knowledge-driven approaches. Data-driven approach estimates viewers' QoE as a function of frame rate, which is calibrated with subject-rated video dataset. This class of models typically makes *a priori* assumptions about the form of the response function. For example, the video quality is assumed to have an exponential/polynomial relationship with respect to frame rate in [2] and [3], respectively. This approach suffers from several problems. First, the shape of the actual quality-frame rate function can deviate significantly from the pre-defined analytic forms. Second, the interaction between frame rate and other video parameters such as spatial resolution [4] is not accounted for. Third, content dependency is not well taken into consideration. Fourth, the content and resolution diversities of existing databases are often insufficient for properly training and validating the models.

In contrast to data-driven models, knowledge-driven approaches focus on the analysis of temporal statistical properties of videos at different frame-rates [5, 6, 7, 8]. Motion smoothness in video is one of the most important aspects of natural scene videos, which has been exploited in optical flow estimation [9], probability models of motions [10, 11], and video quality assessment [12]. It has been demonstrated that motion smoothness creates an effective tool in detecting a wide range of well-known practical distortions, including noise contamination, blurring, line or frame jittering, and frame dropping [7].

Motivated by the success of motion smoothness in video artifact detection [5, 13], we extend it to account for cross-frame rate video quality assessment. We use local phase correlation of complex wavelet coefficients in temporal direction to estimate motion smoothness and investigate the performance of this measurement on different groups of videos classified by motion type and variation.

2. METHOD

To measure motion smoothness in a video, we first assume an ideal case where there is a rigid motion in a 1-D signal signal $f(x)$ and this motion can be modeled as [12]

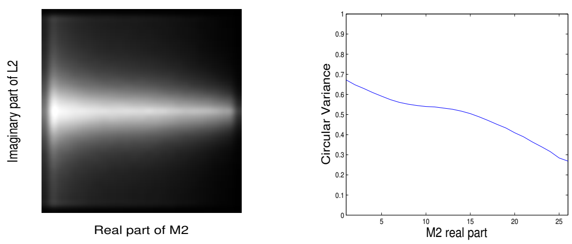
$$h(x,t) = f(x + u(t)) + b(t), \tag{1}$$

where $b(t)$ is the time varying background luminance which is approximately constant in a short period of time and $u(t)$ is the spatial movement over time. This model can be easily extended to the two dimensional space of video frames.

Consider a family of complex wavelets of the form $w(x) = g(x)e^{jw_c x}$, where $g(x)$ is varying slowly with x and w_c is the center frequency of wavelet. The variations of a mother wavelet $w(x)$ can be generated as

$$w_{s,p} = \frac{1}{\sqrt{s}}(\frac{x-p}{s}) = \frac{1}{\sqrt{s}}(\frac{x-p}{s})e^{jw_c x/s}, \tag{2}$$

where s and p are the scale and shift factors, respectively. Then the complex wavelet transform of the signal $f(x)$ can be computed as



(a) Joint histogram (b) Circular Variance

Fig. 1. Temporal motion smoothness by (a) joint histogram of (L_2, M_2); and (b) measure of circular variance on columns of joint histogram.

$$
\begin{aligned}
F(s,p) &= \int_{-\infty}^{\infty} f(x) w^*_{s,p}(x) dx \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} F(w) \sqrt{s} G(sw - w_c) e^{jwp} dw,
\end{aligned} \tag{3}
$$

where $F(w)$ and $G(w)$ are the Fourier transforms of $f(x)$ and $g(x)$, respectively. Applying this transform on the motion model of Eq. (1) leads to

$$
\begin{aligned}
H(s,p,t) &= \int_{-\infty}^{\infty} F(w) \sqrt{s} G(sw - w_c) e^{jw(p+u(t))} dw \\
&\approx F(s,p) e^{j(w_c/s)u(t)}.
\end{aligned} \tag{4}
$$

Take a logarithm on both sides, we have

$$
\log H(s,p,t) \approx \log F(s,p) + j(w_c/s)u(t). \tag{5}
$$

The imaginary part of Eq. (5) has a linear relationship with motion $u(t)$. To relate motion smoothness with this complex wavelet transform, we examine complex wavelet coefficients starting from a time instance t_0 and sample the sequence at consecutive time steps $t_0 + n\Delta t$ for $n = 0, 1, ..., N$. Using these samples in the temporal direction, an N-th order temporal correlation function is defined as

$$
L_N(s,p) \approx \sum_{n=0}^{N} (-1)^{n+N} \binom{N}{n} \log H(s,p,t_0 + n\Delta t). \tag{6}
$$

Using Eqs. (5) and (6), it can be shown that when the motion is (N-1)-th order smooth (all derivatives of $u(t)$ in degree higher than N are zero), the temporal correlation is approximately zero (i.e. $L_N(s,p) \approx 0$). Real-world videos deviate from the ideal case, and such deviation may be used as a measure of motion smoothness.

To observe how motion smoothness varies as a function of local signal energy, we define a temporal energy function

$$
M_N(s,p) \approx \sum_{n=0}^{N} \binom{N}{n} \log H(s,p,t_0 + n\Delta t), \tag{7}
$$

and examine the temporal correlation function (L_N) and temporal energy function (M_N) jointly.

An example of the joint histogram of the real part of M_N and the imaginary part of L_N for a sample video for $N = 2$ is shown as a gray-scale image in Fig. 1(a), where brighter bin indicates more frequent occurrence. In the case of perfect motion smoothness, all bright points would concentrate at the center horizontal line of 0 phase. Spread from the center indicates deviation from perfectly smooth motion. To evaluate the trend of motion smoothness with respect to the local signal energy, the circular variance (CV) [14, 15] is calculated for each column of the joint histogram by

$$
CV_q = 1 - \frac{|\sum_{p=1}^{M} h_{p,q} e^{j\theta_p}|}{\sum_{p=1}^{M} h_{p,q}}, \tag{8}
$$

where M is the number of bins in the histogram and θ_p is the center angle of bin i in a column and $h_{p,q}$ is the height of that bin. CV is bounded between 0 and 1, and the lowest value 0 is achieved when the histogram is clustered in one bin, meaning that all coefficients have the same angle, e.g. concentrated at 0 phase. The trend of CV curve is shown in Fig. 1(b). The normalized area under the curve of CV is computed to quantify motion smoothness

$$
S = \frac{\sum_{q=1}^{K} (1 - CV_q)}{K}, \tag{9}
$$

where K is the number of columns. Examples of the joint histogram for different video contents at different frame rates are given in Fig. 2. It is interesting to observe that generally motion smoothness increases as a function of video frame rate. Also note that, there is strong content-dependencies of the histogram, which in turn affect the CV curve and the subsequent smoothness measure S. In the case that a reference high frame rate (HFR) video is available in the evaluation of a test low frame rate (LFR) video, it is meaningful to assess motion smoothness relative to the reference. Therefore, we propose to normalize the smoothess measure S of a video against that of the HFR reference video

$$
\tilde{S} = \frac{\sum_{q=1}^{K} (1 - CV_q)}{\sum_{q=1}^{K} (1 - CV_q^R)}, \tag{10}
$$

where CV^R is the CV value of the reference HFR video, and $\tilde{S}$ is the normalized temporal motion smoothness (TMS) measure.

3. EXPERIMENTAL RESULTS

We evaluate the proposed TMS measure on 10 high frame rate video sequences selected from the BVI-HDR dataset [16]. Each source sequence is originally at 120 fps and is converted to lower frame rate test sequences at 60 fps, 30 fps, and 15 fps, respectively. Mean Opinion Scores (MOSs) are then collected by running a subjective test to reflect viewer's QoE on the video sequences. The details about the dataset and subjective test are reported in [16]. Using the database, we first examine how the proposed TMS measure correlates with video frame rate and human subjective QoE for individual video content. We then investigate further on motion-based content dependencies.

3.1. Validation

To better understand and to demonstrate the proposed motion smoothness measure, we examine how the joint histogram and its corresponding CV change with respect to different frame rates in Figs. 2 and 3. It can be observed from Fig. 3 that regardless of the content variation, the effect of frame rate reductions is well

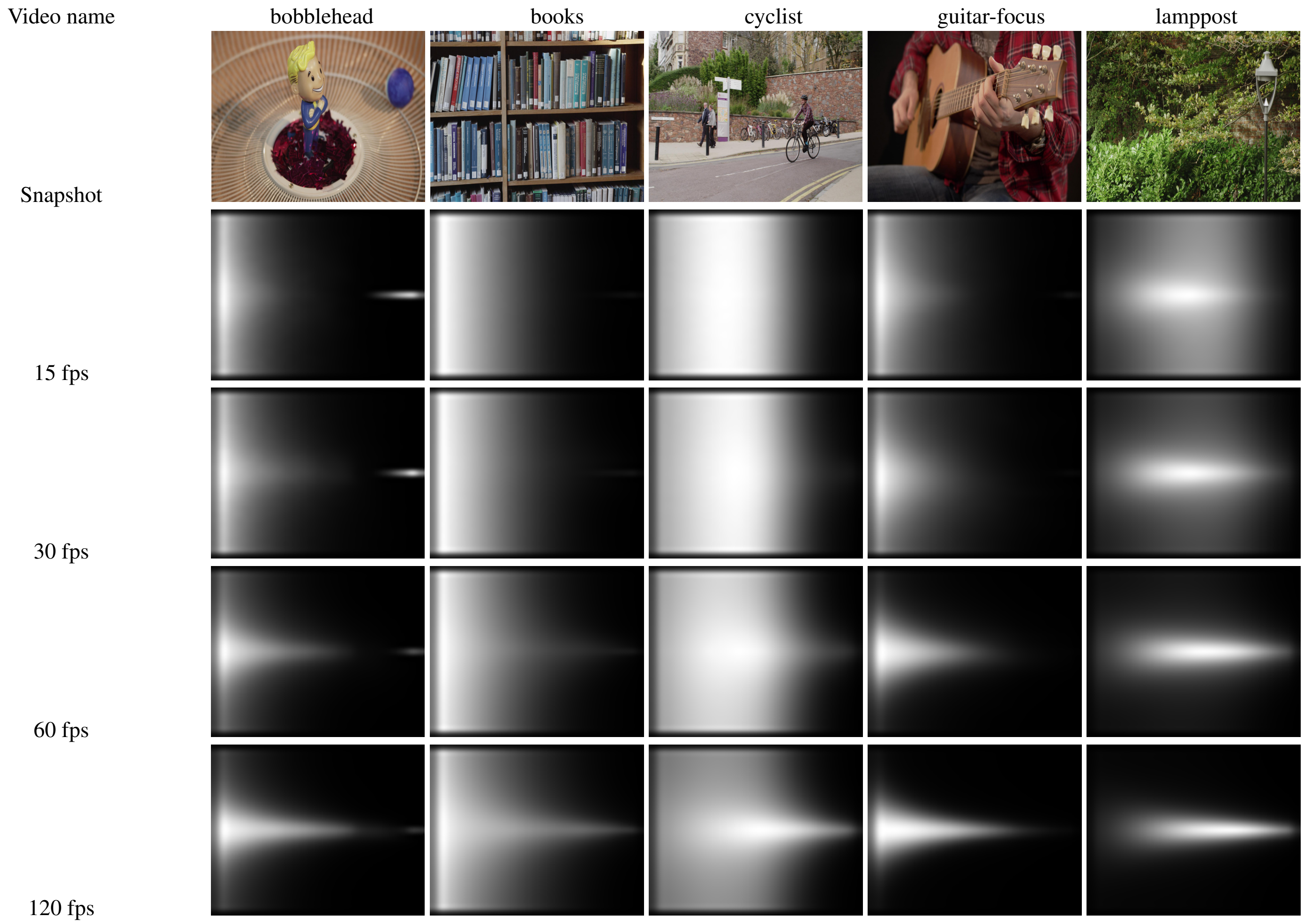


Fig. 2. Temporal motion smoothness by 2D joint histogram of (L_2, M_2) for selected videos from BVI-HFR at four different frame rates.

captured by the departure of the CV curves of the distorted videos from the reference CV curves. Specifically, the CV curve generally moves away from the reference CV curve with the decrease in frame rate. This is further confirmed by the high Spearman rank-order correlation coefficient (SRCC) between the TMS factor and MOS shown in Table 1. The only exception appears to be the "hamster" sequence, where the proposed TMS factor is unable to distinguish the reference and distorted videos. The possible reason could be that the spatial variation in motion pattern and speed are very high, or the local motion pattern in high speed refresh rate may be too complicated to be fully captured by the phase correlation between complex wavelet coefficients.

3.2. Motion Content Dependency

Although the proposed TMS factor exhibits a high correlation with perceptual quality within each content, its behavior varies significantly across different videos as is evident in Fig. 2. For example, for high motion videos such as the "cyclist", there is much larger variation from the ideal smooth motion. This motivates us to study motion-based content dependency of the proposed motion smoothness measure.

An important aspect of motion in the video is the presence of camera geometric transformation in the video captioning process. Such camera motion transformations result in global motion in video. Th global motion has important impact on the visibility of distortion in video and general perception of video quality [17]. For example, the blurring artifact is less visible in globally fast moving scenes and such effects have been considered in existing video quality models [17].

We classify videos into two groups based on the presence of global motion, and computed SRCC between the TMS factor and DMOS on each group. The experimental results are reported in Table 2. It can be seen that the proposed metric better predicts human opinions for videos containing global motion. This could be because the motion is more easily perceived in global motion videos as most of the regions of frames are moving in a consistent manner. For the local motion videos, as moving regions are part of the frames only, the global statistics based TMS factor cannot precisely reflect the impact of such local changes in the overall perceptual quality.

Motion perception provides another important perspective that is missing in the proposed motion smoothness measure to study cross-frame-rate video quality assessment. Specifically, it has been shown that the perceptual motion information content is proportional to the strength of relative motion and the inverse of global background motion [18]. A simple model to account for this relationship is given

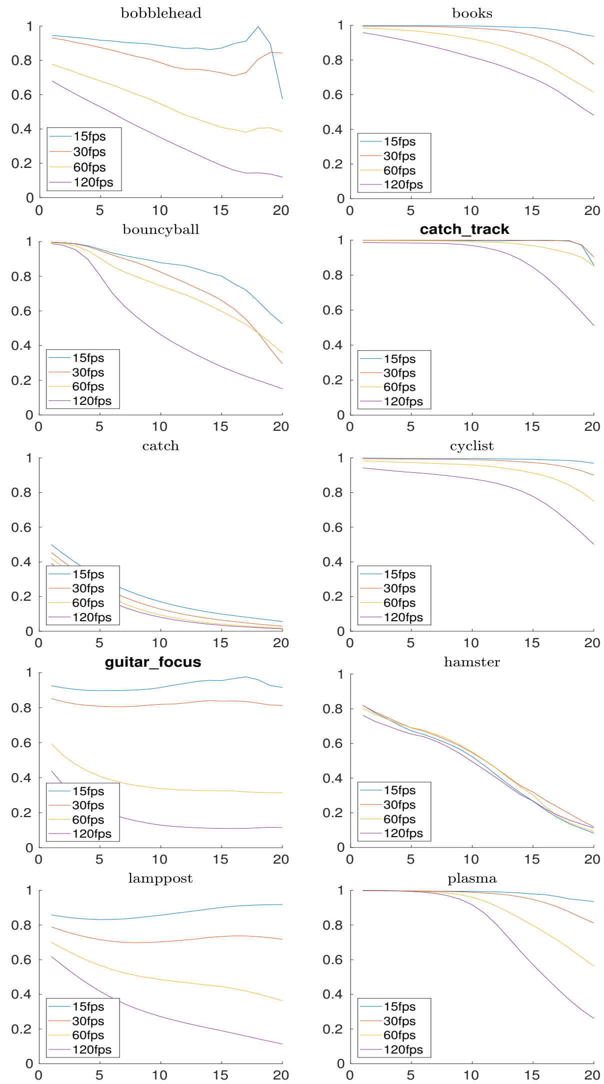


Fig. 3. Circular variance curves of 2D joint histograms of selected videos at different frame rates.

by

$$V = \frac{\sigma(\tilde{d})}{\mu(\tilde{d})} \tag{11}$$

where V, σ, μ, and $\tilde{d}$ represents the spatial motion variation, the variation of frame difference, the average of frame difference in pixels, and the temporal frame difference. Intuitively, V increases as the motion statistics becomes more complex, and decreases as the uncertainty of motion perception μ increases. It is considered a measure of spatial motion variation, or perceptual motion information content (following the principle used in [18, 6]).

We use V to classify the videos used in this study into three classes-low, medium, and high spatial motion variation, as shown in Table 1. By conducting correlation analysis as reported in Table 2, we observe that the proposed metric works better for the videos with lower variation of motion across space. For the medium variation of motion videos, the correlation is close to low V class, and the accuracy of prediction drops significantly for the class of videos with high spatial variation in motion. This could be because for these videos, the proposed method calculates the correlation of wavelet coefficients over the entire frame, while the motion is partial and the human attention could be attracted to certain moving parts of the video frames.

Table 1. Correlation of TMS factor with frame rate and DMOS for individual video sequences of different motion types (global vs local motion) and spatial motion variation levels.

video sequence	motion type	spatial motion variation	SRCC of TMS vs. DMOS
bobblehead	global	low	-1
books	global	low	-1
bouncyball	local	medium	-1
cath_track	global	low	-1
cath	local	high	-0.8
cyclist	global	low	-1
guitar_focus	local	medium	-0.8
hamster	local	high	-0.4
lamppost	local	medium	-1
plasma	local	medium	-1
mean/std	-	-	-0.90 /0.18

Table 2. SRCC between $\tilde{S}$ and DMOS for different motion-based content types.

video group	**SRCC**
all	0.78
local motion	0.62
global motion	0.93
high spatial motion variation	0.71
medium spatial motion variation	0.87
low spatial motion variation	0.93

4. CONCLUSION AND DISCUSSION

We propose a novel motion smoothness measure based on local phase correlation of complex wavelet coefficients. The proposed measure provides a novel perspective to understand the perceptual temporal quality and demonstrates strong promise in cross-frame-rate video quality assessment.

Our study on motion-based content dependencies suggests directions for future improvement. Specifically, much higher correlation between the proposed TMS factor and DMOS is achieved when the motion is global or when the motion variation is low across space, but the correlation drops significantly for the cases of local motion or high spatial variation of motion. This reveals the limitation of the global statistics based approach in the current algorithm. New measures that capture local TMS and the variation of TMS across space would be helpful. New pooling methods based on motion-based saliency detection may be employed to fuse local TMS measurement in a perceptually more meaningful way. Furthermore, how to combine the proposed TMS factor with other perceptual and statistical features to create an overall video quality measure is also worth further investigation.

5. REFERENCES

[1] Cisco Visual Networking Index Cisco, "Global mobile data traffic forecast update, 2016–2021," *Cisco White paper*, 2017.

[2] Yen-Fu Ou, Zhan Ma, Tao Liu, and Yao Wang, "Modeling the impact of frame rate on perceptual quality of video," in *Proc. IEEE Int. Conf. Image Proc.*, San Diego, CA, Oct 2008, pp. 689–692.

[3] Lucjan Janowski and Piotr Romaniak, "QoE as a function of frame rate and resolution changes," in *Future Multimedia Networking*, pp. 34–45. 2010.

[4] Rasoul Mohammadi Nasiri, Jiheng Wang, Abdul Rehman, Shiqi Wang, and Zhou Wang, "Perceptual quality assessment of high frame rate video," in *17th International Workshop on Multimedia Signal Processing (MMSP)*, Beijing, China, 2015, pp. 1–6.

[5] Kai Zeng, Tiesong Zhao, Abdul Rehman, and Zhou Wang, "Characterizing perceptual artifacts in compressed video streams," in *Proc. SPIE 9014, Human Vision and Electronic Imaging XIX*, San Francisco, CA, Feb. 2014, p. 90140Q.

[6] Zhou Wang and Qiang Li, "Video quality assessment using a statistical model of human visual speed perception," *JOSA A*, vol. 24, no. 12, pp. B61–B69, 2007.

[7] Kai Zeng and Zhou Wang, "Temporal motion smoothness measurement for reduced-reference video quality assessment," in *ICASSP*, Dallas, TX, Feb. 2010, pp. 1010–1013.

[8] Rasoul Mohammadi Nasiri and Zhou Wang, "Perceptual aliasing factors and the impact of frame rate on video quality," in *International Conference on Image Processing*, Beijing, China, 2017, pp. 3475–3479.

[9] Denis Fortun, Patrick Bouthemy, and Charles Kervrann, "Optical flow modeling and computation: A survey," *Computer Vision and Image Understanding*, vol. 134, pp. 1–21, 2015.

[10] Dawei W Dong and Joseph J Atick, "Statistics of natural time-varying images," *Network: Computation in Neural Systems*, vol. 6, no. 3, pp. 345–358, 1995.

[11] Yair Weiss, Eero P Simoncelli, and Edward H Adelson, "Motion illusions as optimal percepts," *Nature neuroscience*, vol. 5, no. 6, pp. 598–604, 2002.

[12] Zhou Wang and Qiang Li, "Statistics of natural image sequences: Temporal motion smoothness by local phase correlations," in *SPIE Human Vision and Electronic Imaging XIV*. International Society for Optics and Photonics, 2009, pp. 19–22.

[13] Yao. Wang, Zhan. Ma, and Yen-Fu Ou, "Modeling rate and perceptual quality of scalable video as functions of quantization and frame rate and its application in scalable video adaptation," in *Proc. Int. Packet Video Workshop*, Seattle, WA, May 2009, pp. 1–9.

[14] Nicholas I Fisher, *Statistical analysis of circular data*, Cambridge University Press, 1995.

[15] Kantilal Varichand Mardia, *Statistics of directional data*, Academic press, 2014.

[16] Alex Mackin, Fan Zhang, and David R Bull, "A study of subjective video quality at various frame rates," in *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2015, pp. 3407–3411.

[17] Zhou Wang, Ligang Lu, and Alan C Bovik, "Video quality assessment based on structural distortion measurement," *Signal processing: Image communication*, vol. 19, no. 2, pp. 121–132, 2004.

[18] Alan A Stocker and Eero P Simoncelli, "Noise characteristics and prior expectations in human visual speed perception," *Nature neuroscience*, vol. 9, no. 4, pp. 578–585, 2006.