VIDEO QUALITY TEMPORAL POOLING USING A VISIBILITY MEASURE

Chen Bai and Amy R. Reibman

School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, USA

ABSTRACT

Both mobile and egocentric videos contain much larger motion than broadcast videos. To estimate the quality of videos with huge motion, the temporal pooling strategy should be adaptive to the specific content. Existing methods focus more on the values and variations of frame quality scores and ignore the masking effect of motion. In this paper, a temporal pooling strategy using a visibility measure is proposed to estimate the quality of videos containing large motion, where the imperceivable details during motion are not considered. We then introduce a strategy to measure the influence of measured visibility on pooling and design a subjective test to gather data for the strategy by synthetically creating shaky videos. Our pooling method is demonstrated to be more effective than existing strategies at pooling frame scores estimated by different image quality metrics.

***Index Terms*—** visibility, temporal pooling, video quality

1. INTRODUCTION

Currently, multimedia such as mobile videos [1], egocentric videos [2], drone videos [3] and crowd-sourced interactive live streaming [4], contain much larger motion than broadcast video streaming, which strongly affects the perception of distortions in the videos. We categorize all these types of videos with heavy motion to be large motion videos (LMVs). For example, broadcast videos are often recorded by stably-mounted cameras that are static or contain low-speed motion. In contrast, egocentric videos are recorded by wearable cameras that follow human body movements with high and random speed. Outdoor live streaming is often quite shaky when it is captured using hand-held cameras.

A common strategy to estimate video quality has two steps. First, an objective image quality metric is used to measure the spatial quality score of individual video frames. Then, a temporal pooling method is applied to combine all frame quality scores over time to get a single video-level score. This strategy has been demonstrated to be an effective way to estimate video quality [5, 6]. In this paper, we focus on the temporal pooling step.

The mapping from frame-level scores to a video-level score should be adaptive to the type of videos whose quality is to be evaluated. If the video has little motion, an emphasis on motion is unnecessary and a typical pooling strategy [7] can be applied. If the video is a LMV, the imperceivable details during motion should not be considered when we interpret frame-level quality scores.

The visibility of quality degradations during motion has been studied from two perspectives. The first is the window of visibility proposed in [8], that represents human visual spatio-temporal contrast sensitivity function (STCSF). The spatial details of the image are invisible outside of the window. The boundary of the window is decided by the perceivable contrast threshold of STCSF. Another perspective is motion sharpening [9–11] in which the blurred images looks sharper when they moves fast. However, the effect has not been thoroughly described using a theoretical model.

Most current temporal pooling metrics do not consider the influence of motion so they are not suitable to measure the quality of LMVs. For example, the Minkowski pooling [7] emphasizes the influence from low quality frames. A hysteresis model [12] emphasizes the memory effects for human observer. In [13], the temporal pooling strategy considers the influence of the temporal gradients of quality scores. One model that considers the influence of motion is the human visual speed model proposed in [14]. It considers global and local speed of the frame in pooling, but it is only evaluated on television video sequences.

Our contribution in this paper is to provide an adaptive temporal pooling mechanism to estimate the quality of LMVs. We use weighted average pooling strategy that uses the function of visibility as the pooling weight to combine frame quality scores. The method can be expressed as

$$Q = \frac{\sum_i \lambda(V_i) \cdot q_i}{\sum_i \lambda(V_i)}, \tag{1}$$

where q_i and V_i are the spatial quality and visibility of frame i, respectively, and Q is the video quality score. $\lambda(\cdot)$ is the function that models how visibility influences the pooling of Q_i. We propose a visibility measure which computes the proportion of frame details that are visible under a given motion based on the window of visibility [8, 15], and systematically measure the function $\lambda(\cdot)$ based on subjective data.

There are four potential scenarios in which our temporal pooling strategy can be applied. First, we can compare the overall blurriness of videos that are captured by multiple different shaky cameras at the same time. Second, videos with motion editing can be compared relative to the original version. One example is video stabilization in which blur is more visible after stabilization. Third, videos with post-processing can be compared relative to the original. One example is video illumination enhancement that may add newly generated artifacts into frames. The artifacts may be imperceivable due to the masking effect of motion.

In this paper, a temporal pooling strategy using a visibility measure is proposed to estimate the quality of LMVs. The video quality is computed as the weighted average of frame quality scores, where the weights are the function $\lambda(\cdot)$ of visibility. We propose and illustrate our visibility measure for individual frames under a given motion in Section 2. Then we introduce the strategy to measure the function $\lambda(\cdot)$, and describe the method to gather video quality data for the strategy in Section 3. In Section 4, we implement a subjective test to gather subjective video quality scores to measure the function $\lambda(\cdot)$. The subjective data is also used to validate our visibility pooling strategy and to compare our method with other existing temporal pooling methods. The results show that our method provides the best performance.

2. VISIBILITY MEASUREMENT

In this section, we propose a visibility measurement developed based on the window of visibility [8, 15]. We overview the theory of window of visibility and define a measure of visibility to be the proportion of visible frame details.

2.1. Overview of the Window of Visibility

The basic idea of the window of visibility is that there exists a spatio-temporal window outside which the contrast is invisible [8, 15]. Let the x-coordinate be spatial frequency (cycles/degree) and y-coordinate be temporal frequency (Hz). The positive frequency part of the window is the triangle with three vertices, $(0, 0)$, $(u_0, 0)$ and $(0, w_0)$, where u_0 is spatial frequency limit and w_0 is temporal frequency limit.

According to the relationship between window limit u_0, w_0 and display luminance I in [15], u_0 is saturated at around 50 $cycles/deg$ at $I = 7cd/m^2$, and w_0 has a linear relationship with display luminance $log_{10}(I)$. We approximate as $w_0 = 15 \cdot log_{10}(I) + 35$ Hz.

Consider the motion function of a line: $m(x, t) = \delta(x - rt)$, where x is the position, t is the time, and r is the speed. The transformed moving line in the spatio-temporal domain is determined by $f(u, w) = \delta(w + ru)$, where u and w are spatial and temporal frequency, respectively. Figure 1 shows the window of visibility (u_0, w_0). The dashed line is the part of $f(u, w)$ outside the window of visibility that cannot be perceived. See more details in [15].

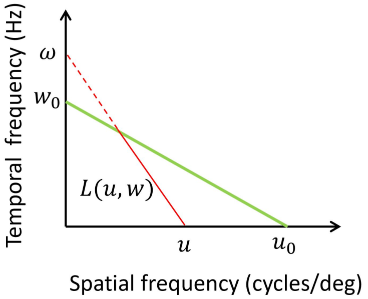

Fig. 1: Green: the window of visibility (u_0, w_0) boundary. Red: spatio-temporal content of $\boldsymbol{u}$ in which the solid line is visible, and the dashed line is invisible

2.2. Proposed Method

Based on the idea of the window of visibility, we measure the visibility to be the proportion of the overall power spectrum that is inside the window of visibility. The visibility value in an image patch is calculated as the summation of the fraction of energy of all spatial frequencies weighted by their visible proportion inside the window of visibility. The visibility for an image is then a spatially pooled average from all image patches.

Given an image patch with speed $\boldsymbol{v}$ (where all bold font parameters indicate a vector variable), we have a fixed window of visibility represented as (u_0, w_0). Let $\boldsymbol{u}$ be one spatial frequency in the image patch. We consider only the part of $\boldsymbol{u}$ parallel to $\boldsymbol{v}$ that influences the visibility. Then the temporal frequency w for $\boldsymbol{u}$ is calculated as $w = \boldsymbol{u} \cdot \boldsymbol{v}$. The u in Figure 1 is $\|\boldsymbol{u}\| \cos\theta$, where $\|\boldsymbol{u}\|$ is the length of $\boldsymbol{u}$ and θ is the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$.

We compute the fraction of energy, $P(\boldsymbol{u})$ for spatial frequency component $\boldsymbol{u}$, in the image patch to be

$$P(\boldsymbol{u}) = \frac{M(\boldsymbol{u})}{\int_{\boldsymbol{u}} M(\boldsymbol{u})}, \tag{2}$$

where $M(\boldsymbol{u})$ is the magnitude of the spatial power spectrum at $\boldsymbol{u}$. Not all spatial frequencies $\boldsymbol{u}$ will be visible because some lie out of the window of visibility.

The proportion of spatio-temporal content at $\boldsymbol{u}$ inside the window of visibility is the weight $\omega(\boldsymbol{u})$ for energy fraction $P(\boldsymbol{u})$, calculated as

$$\omega(\boldsymbol{u}) = \frac{L(u, w)}{\sqrt{u^2 + w^2}}, \tag{3}$$

where $L(u, w)$ is the length of the visible part shown in Figure 1, and $\sqrt{u^2 + w^2}$ is the total length. The visibility of image patch q in frame i is then calculated as

$$V_{iq} = \int_{\boldsymbol{u}} \omega(\boldsymbol{u}) P(\boldsymbol{u}). \tag{4}$$

The visibility of frame i is spatially pooled from 31×31

patches overlapped by 15 pixels:

$$V_i = \frac{1}{N_q} \sum_q V_{iq}, \tag{5}$$

where q is the patch index, N_q is the total number of patches. The measured V_i is not very sensitive to the chosen patch size and the spatial pooling method. Note in our actual implementation, speed refers to the viewing angular velocity, which is dependent on viewing distance.

3. STRATEGY TO MEASURE $\lambda(\cdot)$

In this section, we introduce a strategy of measuring the function $\lambda(\cdot)$ in Equation 1. This function describes the influence of visibility on the pooling of spatial quality scores. To obtain the objective and subjective quality scores to estimate $\lambda(\cdot)$, we introduce our data gathering strategy along with its motivation.

3.1. Estimate Function $\lambda(\cdot)$

The function $\lambda(\cdot)$ can be measured using D ($D > 1$) test video sequences that share the same visibility but have a different spatial quality in the temporal domain.

To measure the $\lambda(\cdot)$, we can rewrite Equation 1 as

$$Q' = \boldsymbol{q}^T \lambda(\boldsymbol{V}), \tag{6}$$

where the scaled video quality $Q' = \sum_i \lambda V_i \cdot Q$. Let K be the number of frames; the $\lambda(\boldsymbol{V})$ and $\boldsymbol{q}$ are both $K \times 1$ vectors that represent spatial quality and the function $\lambda(\cdot)$ of visibility $\boldsymbol{V}$, respectively. To get the solution of $\lambda(\boldsymbol{V})$, we construct the case that D different videos have the same visibility with different quality $Q'_1, Q'_2, \ldots, Q'_D$. The least square solution $\lambda(\hat{\boldsymbol{V}})$ for $\lambda(\boldsymbol{V})$ in Equation 6 is

$$\lambda(\hat{\boldsymbol{V}})^T = \begin{bmatrix} \boldsymbol{q_1}^T \\ \boldsymbol{q_2}^T \\ \vdots \\ \boldsymbol{q_D}^T \end{bmatrix}^{\dagger} \begin{bmatrix} Q'_1 \\ Q'_2 \\ \vdots \\ Q'_D \end{bmatrix}, \tag{7}$$

where † is the Moore-Penrose pseudo-inverse. If the D sequences have known subjective quality scores, then we can estimate their frame quality scores and compute $\lambda(\hat{\boldsymbol{V}})$.

3.2. Strategy Motivation

To measure the $\lambda(\cdot)$, D video sequences with same visibility but different spatial quality in the temporal domain need to be synthetically created, and then their subjective and objective quality need to be measured.

To ensure the D video sequences have the same visibility, we need to control the motion in each video. To create such videos, we can move a cropping window in a high-quality image to create the desired motion, and add synthetic motion blur to create frames with different spatial quality.

To gather the subjective quality, we need human observers to be able to compare the D synthetic videos. The typical strategy is to apply different amounts of one synthetic distortion into different videos, and determine the minimum differences between these synthetic videos that can be perceived by human observers. One method to apply this strategy into our case is to add the same amount of blur into all frames in one video, and add different amounts for different videos. However, the motion differences between videos would have to be quite large to maintain the same visibility, so that motion would have significant influence on the subjective evaluation. Another method is to add temporally sinusoidal blur into one video, and temporally shift the blur curve to create the other $D-1$ videos. This method does not require large variations in the amount of motion for different videos. In our implementation, we use the second method.

To gather the objective quality scores of video frames, we need to use one image quality metric. $\lambda(\cdot)$ could be measured differently when using different image quality metrics, since they may have inconsistent scales for measuring the same quality degradation. In this paper, we only use frame scores estimated by one image quality metric, LVI [2], to measure $\lambda(\cdot)$, and we test the result using other image quality metrics. The goal is to demonstrate a consistent design that can be effective for temporal pooling of any quality metric. LVI measures the relative quality between images with overlapping but not necessarily pixel-aligned content, and it has been demonstrated to be effective at providing a consistent measure for blur [2, 16].

3.3. Data Gathering Strategy

To measure $\lambda(\cdot)$, we want to create a set of videos, Γ, that has different spatial quality but the same visibility in the temporal domain.

Before introducing the strategy to create Γ, we first introduce the notations for motion, blur and visibility. Let A_j, B_j, P_j be the motion profile, blur profile and visibility profile, respectively, where j is the profile index. The profile describes the pixel shifts (profile A_j), the blur kernel size (profile B_j) and the visibility (profile P_j) temporally for each frame in a video.

Assume we have a blur profile B_0 and a visibility profile P_0 where $P_0 \propto -B_0$. If B_0 is temporally shifted to B'_0 while P_0 is maintained, there would be less masking effect for B'_0. Figure 2 shows the comparison between B_0, B'_0 and P_0 in which $1 - P_0 = c \cdot B_0$ with c to be a constant parameter. As a temporal shift is introduced into B_0 to B'_0, the overlap between B'_0 and $1 - P_0$ becomes smaller so that more blurry frames would have higher visibility.

To create videos with temporally-shifted quality, we shift the phase of the temporal quality curve in the frequency do-

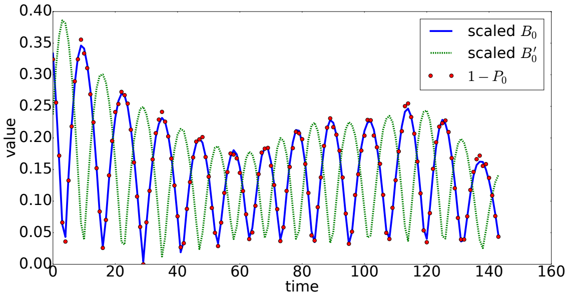

Fig. 2: Comparison between Blur profile B_0, B_0' and visibility profile $1 - P_0$

main. Assume we have a motion profile A_0 with blur profile B_0 and visibility profile P_0. By shifting B_0 in the frequency domain with phase 0.125π, 0.25π, 0.375π, 0.5π, we get blur profiles B_1, B_2, B_3 and B_4. The motion profile A_0 is then edited to become A_j', for $j = 0, 1, 2, 3, 4$, so that the video visibility profile is constant, P_0. Γ is formed with videos created by (B_j, A_j'), for $j = 0, 1, 2, 3, 4$. The goal is to demonstrate the decrease of visibility has a masking effect on frame blurriness so that the perceived video quality increases.

4. SUBJECTIVE TEST

In this section, we describe our subjective test using synthetic shaky videos. The test results are then used to estimate the function $\lambda(\cdot)$ using method described in Section 3.1. Finally, our pooling strategy is demonstrated to perform the best when compared to different temporal pooling methods across a range of existing image quality metrics.

4.1. Test Video Sets

To create synthetic videos, we start with 4 high-resolution images corresponding to test sets Γ_j, where $j = 0, 1, 2, 3$. Videos are created by moving the cropping window in the original image using the strategy described in Section 3.3.

To synthetically create Γ_j, we first create a motion profile A_j that extracts the motion from an actual captured shaky video α. Assume we want A_j to be in the frequency range from a Hz to b Hz. We first find the peak frequency F_{peak} from a Hz to b Hz in the motion frequency spectrum of α and apply a Gaussian window centered around F_{peak} to get the motion information to create A_j. Second, we compute blur B_j based on the motion A_j in which the size of the average blur filter is proportional to the pixel displacement. The visibility is then computed as $P_j(t) = \max(0, 1 - c \cdot B_j(t))$, where t is the time instance, c is a constant parameter, $P_j(t)$ and $B_j(t)$ are the visibility and the blur kernel size at time t.

Γ_1 and Γ_2 are created with frequency range between 1 and 2 Hz, and the frequency range for Γ_3 and Γ_4 are between 2 and 3 Hz. Each test set has five videos with blur phase shift $0, 0.125\pi$, 0.25π, 0.375π, 0.5π. All test videos with their corresponding reference videos and the video that is used to compute motion profiles are available at [17].

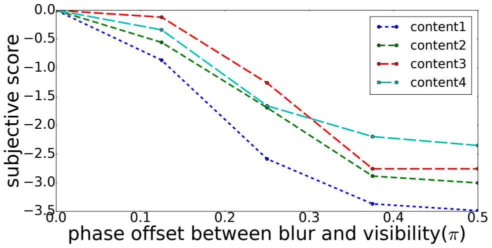

Fig. 3: Subjective scores (0: best quality in each test set)

4.2. Test Setup

Our subjective test method is paired comparison. All pairs of comparisons are videos in the same set Γ_j. A pair of test videos is presented one after another on a monitor (DELL U2718Q) that has resolution 3840×2160. The video is presented at the center of the screen with resolution 1920×1080. The background is gray at 128. Each test video is 5 seconds with frame rate 30 frames/second. Since the calculation of the visibility relies on the viewing distance, it is fixed to be 3.2 times the height of the display. Each of the 20 test participants are asked to choose *in which video can you perceive more spatial details*.

4.3. Subjective Test Results

The relative subjective qualities are estimated using the Bradley-Terry Model [18]. The test results are shown in Figure 3 where the best quality is 0 for each test content. The subjective results indicates that a larger phase difference between visibility and blur introduces more perceived quality degradations for a human observer in all four test contents. This demonstrates that our measure for visibility does have a masking effect on the perception of blurriness; low quality frames have little influence when they have low visibility.

One additional comment about content differences is that content 1 and 2 show greater quality differences between videos with phase shift 0 and 0.5π than content 3 and 4. One reason is that content 1 and 2 have lower-frequency motion than do content 3 and 4. Content 1 has much greater quality difference between videos with phase shift 0 and 0.5π than other contents, because it contains a higher proportion of regions with high spatial frequencies that enable the differences to be more perceivable.

4.4. Estimating $\lambda(\cdot)$

We estimate the function $\lambda(\cdot)$ using the method illustrated in Section 3.1. We apply the subjective results from the 4 contents to estimate $\lambda(\cdot)$, and choose the estimated model using content 1 because it achieves the highest PLCC between $\lambda(\hat{\boldsymbol{V}})$ and $\lambda(\boldsymbol{V})$ among the 4 contents.

The temporal weighting vector $\lambda(\hat{\boldsymbol{V}})$ is calculated by Equation 7, where $\boldsymbol{Q}$ is the subjective quality scores of the 5 test videos of content 1. Vector $\boldsymbol{q}$ is estimated by LVI [2].

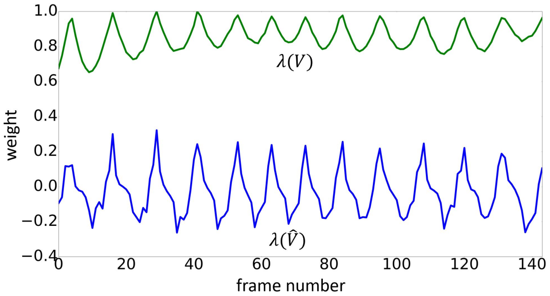

Fig. 4: Comparison between $\lambda(\boldsymbol{V})$ and estimated $\lambda(\hat{\boldsymbol{V}})$

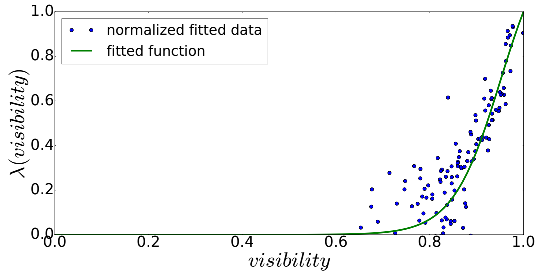

Fig. 5: Function $\lambda(\cdot)$ in Equation 1: x-axis is measured visibility V_i, y-axis is $\lambda(V_i)$.

Figure 4 shows the comparison between $\lambda(\boldsymbol{V})$ and the estimated weighting vector $\lambda(\hat{\boldsymbol{V}})$. We fit function $\lambda(\cdot)$ using the logistic function.

$$f(x) = (t_0 - t_1)/(1 + \exp(-(x - t_2)/|t_3|)) + t_1 \qquad (8)$$

Then we normalize the values after mapping, where the maximum value and minimum value for normalization is $f(1)$ and $f(0)$. The estimated $\lambda(\cdot)$ shown in Figure 5 maps measured visibility to pooling weight with fitted parameters $t_0 = 0.26, t_1 = -1.25, t_2 = 0.95, t_3 = -0.05$. Our measure for visibility is shown to have an nonlinear relationship with the pooling weight in Figure 5.

4.5. Evaluating Overall Method

The other three test video contents are used as validation for our pooling strategy. We compare our method with existing pooling strategies: average pooling, percentile pooling (70th), Minkowski pooling (p=2), speed pooling [14], temporal variation pooling [13], and hysteresis pooling [12]. In our implementation, the relative speed is zero in speed pooling, since all our test videos only contain global motion. In addition, we only consider the global temporal pooling method in [13] and set the distortion value to be the negative quality value plus the maximum quality value of the quality metric. To test the generality for different image quality metrics, we estimate the frame quality using two full-reference (FR) methods (SSIM [19], [20]), two no-reference (NR) methods (BRISQUE [21], NIQE [22]) and one mutual reference method (LVI [2]). All quality scores are normalized to be between 0 to 1 using the minimum and maximum values in [23].

Table 1 shows the Pearson linear correlation coefficient (PLCC) and Spearman rank-order correlation coefficient (SROCC) between the subjective video quality scores and the objective temporal pooling scores using different image quality metrics. For all three test contents, our pooling method shows the best overall performance.

Our method can achieve high PLCC and SROCC for two reasons. First, because of the limited number of test samples, PLCC and SROCC mainly measure if the method correctly ranks the video quality. Second, the subjective test and our proposed method are both specifically designed for the masking effect on perceived blurriness due to motion.

The results also show that our method can generalize across different contents. Our method incorporates the influence of content since our estimation of visibility computes spatio-temporal information in a single frame. In addition, we model the relationship $\lambda(\cdot)$ between visibility and pooling weight based on gathered subjective data that has better cross-content performance than considering $\lambda(\cdot)$ to be linear.

Our method is not successful when pooling BRISQUE and NIQE in content 2. BRISQUE and NIQE do not provide a consistent measure when the same amount of blur is added into pixel-shifted content. The test videos in content 2 are produced with greater frame-to-frame pixel shifts than content 3 and 4, so the BRISQUE and NIQE scores of content 2 are not as robust as in other contents.

Speed pooling has the second performance among all. It computes temporal weights based on motion, but their model parameters are only evaluated on videos with low-speed motion. The other 5 methods are not suitable for our situation. They pool the video quality using only frame scores. However, our videos are created to have similar frames scores with different visual qualities, so these methods are not capturing all the relevant information.

5. CONCLUSIONS

In this paper, we propose a temporal pooling strategy built on a measurement of visibility that is more effective at estimating the perceived blurriness of LMVs than existing pooling strategies. The visibility measure is proposed based on the window of visibility theory to compute the fraction of visible details within a single frame under a given motion. A systematic subjective test is implemented to demonstrate the masking effect on motion blur using our visibility measure. The subjective video scores are also used to estimate the influence of visibility on the pooling of frame quality scores. The test results indicate that our pooling strategy is more suitable for LMVs and can be effectively applied to pool quality scores estimated by different types of image quality metrics. The future work is to investigate the video quality assessment at the application level, for example, egocentric video quality comparison and motion-edited video quality evaluation.

Content 2

Pooling method	SSIM	VSNR	LVI	BRISQUE	NIQE
average	0.75(0.7)	0.86(0.7)	0.74(0.7)	0.49(0.3)	0.84(0.6)
percentile	0.83(0.6)	0.99(0.9)	0.86(0.6)	-0.91(-0.9)	0.71(0.4)
Minkowski	0.7(0.6)	0.84(0.7)	0.78(0.7)	0.66(0.7)	0.82(0.5)
speed [14]	0.94(0.9)	0.97(0.9)	0.94(0.9)	0.76(0.8)	0.95(0.9)
variation [13]	-0.59(-0.4)	0.94(1.0)	0.41(0.5)	0.68(0.4)	0.76(0.9)
hysteresis [12]	0.57(0.6)	0.87(0.9)	0.62(0.7)	0.86(0.9)	0.59(0.7)
visibility	0.99(1.0)	0.98(1.0)	0.99(1.0)	0.64(0.6)	0.81(0.7)

Content 3

Pooling method	SSIM	VSNR	LVI	BRISQUE	NIQE
average	-0.46(0.1)	-0.74(-0.7)	0.04(0.1)	-0.58(-0.3)	-0.2(0.0)
percentile	-0.09(0.1)	0.72(0.7)	-0.3(0.0)	-0.77(-0.9)	0.05(0.3)
Minkowski	-0.51(-0.4)	-0.96(-0.9)	0.1(0.1)	-0.55(-0.3)	-0.2(0.0)
speed [14]	0.80(1.0)	0.51(0.7)	0.65(0.6)	-0.27(-0.1)	0.41(0.2)
variation [13]	-0.54(-0.6)	-0.74(-0.7)	-0.0(0.1)	-0.57(-0.3)	-0.23(0.0)
hysteresis [12]	0.82(0.7)	-0.38(-0.3)	0.37(0.3)	0.36(0.5)	0.61(0.5)
visibility	0.98(1.0)	0.96(1.0)	0.97(1.0)	0.99(1.0)	0.98(1.0)

Content 4

Pooling method	SSIM	VSNR	LVI	BRISQUE	NIQE
average	0.22(0.0)	0.29(0.0)	0.69(0.5)	-0.03(-0.3)	-0.51(-0.4)
percentile	0.55(0.3)	0.87(0.8)	0.78(0.7)	-0.84(-0.9)	-0.72(-0.9)
Minkowski	0.04(0.0)	0.07(0.0)	0.46(0.1)	0.09(-0.3)	-0.46(-0.3)
speed [14]	0.85(0.9)	0.72(0.6)	0.89(0.9)	0.44(0.3)	0.49(0.3)
variation [13]	0.17(0.0)	0.29(0.0)	0.71(0.7)	-0.09(-0.4)	-0.64(-0.4)
hysteresis [12]	0.79(0.6)	0.5(0.1)	0.59(0.3)	0.39(0.1)	0.7(0.4)
visibility	0.99(1.0)	0.98(1.0)	0.98(1.0)	0.97(0.9)	0.98(1.0)

Table 1: PLCC (SROCC) between objective pooling scores and subjective scores.

6. REFERENCES

[1] Deepti Ghadiyaram and Janice Pan, "In-capture mobile video distortions: A study of subjective behavior and objective algorithms," *IEEE Trans. Ckts. Syst. for Video Tech.*, vol. 28, no. 9, pp. 2061–2077, 2018.

[2] Chen Bai and Amy R. Reibman, "Image quality assessment in first-person videos," *J. Vis. Commun. Image Represent*, vol. 54, pp. 123–132, 2018.

[3] Josh Harguess and Michael Reese, "Aggregating motion cues and image quality metrics for video quality estimation," in *Geospatial Informatics, Motion Imagery, and Network Analytics VIII*, 2018, vol. 10645, p. 106450A.

[4] Cong Zhang and Jiangchuan Liu, "On crowdsourced interactive live streaming: a twitch.TV-based measurement study," in *ACM Workshop on Network and Operating Systems Support for Digital Audio and Video*, 2015, pp. 55–60.

[5] Michael Seufert, Martin Slanina, Sebastian Egger, and Meik Kottkamp, "To pool or not to pool: A comparison of temporal pooling methods for HTTP adaptive video streaming," in *Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, 2013, pp. 52–57.

[6] M. Venkata Phani Kumar and Sudipta Mahapatra, "A multi-stage temporal pooling mechanism for video quality assessment," in *50th Asilomar Conference on Signals, Systems and Computers*, 2016, pp. 1853–1857.

[7] Snjezana Rimac-Drlje, Mario Vranjes, and Drago Zagar, "Influence of temporal pooling method on the objective video quality evaluation," in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, 2009, pp. 1–5.

[8] Andrew B. Watson, Albert J. Ahumada, and Joyce E. Farrell, "Window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays," *Journal of the Optical Society of America A*, vol. 3, no. 3, pp. 300–307, 1986.

[9] Stephen T. Hammett, Mark A. Georgeson, and Andrei Gorea, "Motion blur and motion sharpening: temporal smear and local contrast non-linearity," *Vision Research*, vol. 38, no. 14, pp. 2099–2108, 1998.

[10] Mark A. Georgeson and Stephen T. Hammett, "Seeing blur: motion sharpening without motion," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 269, no. 1499, pp. 1429–1434, 2002.

[11] Karen K De Valois, Tatsuto Takeuchi, and Thomas D Wickens, "Appearance of images," in *Human Vision and Electronic Imaging*, 2008, vol. 6806, p. 680605.

[12] Kalpana Seshadrinathan and Alan C. Bovik, "Temporal hysteresis model of time varying subjective video quality," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 1153–1156.

[13] Alexandre Ninassi, Olivier Le Meur, Patrick Le Callet, and Dominique Barba, "Considering temporal variations of spatial visual distortions in video quality assessment," *IEEE Journal of Selected Topics in Signal Processing: Special Issue on Visual Media Quality Assessment*, vol. 3, no. 2, pp. 253–265, 2009.

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

[15] Andrew B. Watson, "High frame rates and human vision: A view through the window of visibility," *Motion Imaging Journal*, vol. 122, no. 2, pp. 18–32, 2013.

[16] Chen Bai and Amy R. Reibman, "Subjective evaluation of distortions in first-person videos," in *Human Vision and Electronic Imaging*, 2017, number 14, pp. 110–117.

[17] "Visibility test video dataset," https://engineering.purdue.edu/VADL/resources/visibility_test/visibility_test.zip.

[18] John C. Handley, "Comparative analysis of Bradley-Terry and Thurstone-Mosteller paired comparison models for image quality assessment," in *Proc. IS&Ts Image Processing, Image Quality, Image Capture, Systems Conference*, 2001, pp. 108–112.

[19] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[20] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2284–2298, 2007.

[21] Anish Mittal, Anush K. Moorthy, and Alan C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[22] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik, "Making a completely blind image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.

[23] He Liu and Amy R. Reibman, "Software to stress test image quality estimators," in *Quality of Multimedia Experience (QoMEX)*, 2016.