# VISIBILITY PREDICTION OF FLICKER DISTORTIONS ON NATURALISTIC VIDEOS

*Lark Kwon Choi*[1], *Lawrence K. Cormack*[2], *Alan C. Bovik*[1]

[1]Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, USA
[2]Department of Psychology, The University of Texas at Austin, Austin, TX, USA

## ABSTRACT

We conducted a series of human subjective studies where we found that the visibility of flicker distortions on naturalistic videos is strongly reduced when the speed of coherent object motion is large. Based on this, we propose a model of flicker visibility on naturalistic videos. The model predicts target-related activation levels in the excitatory layer of neurons for a video using spatiotemporal backward masking. The target-related activation level is then shifted and scaled according to video quality level changes that cause flicker distortions. Finally, flicker visibility is predicted based on neural flicker adaptation processes. Results show that the predicted flicker visibility using the model correlates well with human perception of flicker distortions on naturalistic videos.

***Index Terms***— Motion, flicker distortion, visibility of distortion, backward masking, and video quality

## 1. INTRODUCTION

As the global volume of digital videos grows exponentially, reliable and accurate video quality assessment (VQA) is essential to satisfy users' Quality of Experience [1]. Since humans are generally the ultimate recipients of videos, the perceptual quality of the received videos is significant in the design of video processing systems, from capture to display. One important component in the design of VQA models that remains poorly understood is the effect of temporal visual masking on the visibility of temporal distortions.

Digital videos not only suffer from spatial distortions such as blocking, blurring, ringing, mosaic patterns, and noise, but are also degraded by temporal artifacts including motion compensation mismatches, flicker, mosquito effects, ghosting, jerkiness, smearing, and so forth [2]. Early VQA models incorporated features for capturing spatial artifacts in each frame, then applied temporal weights to pool the spatial quality scores. However, a temporally weighted pooling of spatial quality scores does not explain perceptually relevant temporal distortions [3]. Hence, researchers have started to explore temporal variations of spatial distortions [4] and to directly measure temporal distortions [3].

The mere presence of spatial or temporal distortions does not imply quality degradation since the visibility of distortions can be strongly reduced or completely eliminated by visual masking effects [5]. Visual masking occurs when a stimulus (the mask) is superimposed on another stimulus (the target) typically of similar motion, frequency, color, or orientation [6]. For example, the detectability of a deviation $I + \Delta I$ from a patch luminance $I$ is proportional to the ratio $\Delta I / I$, so a localized distortion $\Delta I$ is more likely visible in a dark region than a bright one [7]. This is called luminance masking. In contrast masking, local high-frequency energy in an image reduces the visibility of other high-frequency features such as noise [8]. Spatial masking is well-known and plays a central role in the design of image quality models [9], video compression [10], and watermarking [11].

Regarding temporal masking, psychophysical tests have been executed using flashes, sine wave-gratings, and vernier stimuli. Recently, a striking "motion silencing" illusion [12] was presented, where the salient temporal changes of objects in luminance, color, size, and shape appear to stop in the presence of large, coherent object motions. This implies that object motion dramatically alters the visibility of temporal distortions. Although a spatiotemporal filter model precisely predicts when human perceives a motion silencing illusion as a function of object change rate and velocity [13], since the effect has been studied on highly synthetic stimuli such as moving dots, there is a significant need to understand the impact of motion on flicker visibility in naturalistic videos.

In video processing research, temporal masking was first studied in the early days of analog TV. It was found that human observers could not perceive a temporary reduction of the spatial details in TV signals after scene changes [14]. Later, temporal masking of distortions was studied in the context of video compression. Netravali *et al.* [10] examined the effects of luminance transitions on the perception of quantization noise. Puri *et al.* [15] designed an adaptive video coder using the visibility of noises on textures, edges, and flat areas. Haskell *et al*. [16] proposed that observers are more tolerant of distortions in moving images than in stationary images. However, implementations of video compression have been largely heuristic based on anecdotal evidence. More recently, the just-noticeable distortion (JND) has been applied to adaptive image coding [17] and on the visibility of noises in videos [18]. Although JND is related to the visibility of distortions, the effect of object motion has not been explicitly analyzed in this context.

Here we propose a visibility prediction model of flicker distortions on naturalistic videos. The model predicts target-related activation levels in the excitatory layer of neural networks for displayed video frames against immediately following frames via spatiotemporal backward masking [19]. The target-related activation level is then shifted and scaled based on the flicker intensity. Finally, an accumulation or adaptation process is applied. The predicted results correlate well with human perception of flicker visibility.

## 2. PREDICTION OF FLICKER VISIBILITY

### 2.1. Target-related Activation Level of Neurons

Neurons at the retina collectively receive inputs from the photoreceptors (rods and cones) and produce "center-surround" excitatory-inhibitory responses to local cone (or rod) cell signals and their surrounding neighbors, yielding a reduced-entropy residual signal. Each receptive field (RF) describing a neuron's response may be well modeled as having an excitatory layer and an inhibitory layer [7]. An excitatory layer increases firing rates, while an inhibitory layer suppresses the firing of the cells. Two layers interact in an antagonistic way, and the overlap of RFs controls firing rates to increase spatial resolution. The excitatory-inhibitory interactions can be modeled via neural field equations [20].

We predicted the firing rate of neurons, called target-related activation levels, on naturalistic videos via Hermen's spatiotemporal backward masking. Since backward masking is the most dominant temporal masking in naturalistic videos as proposed by interruption theory [21] and scene change tests [22], we adopted Hermen's model. The target-related activation level monotonically relates to the percentage of correct responses to neural stimuli. In our context, it is the visibility of a current frame masked by following frames.

Neural interactions between and within excitatory and inhibitory layers were first computed on the test videos used in human psychophysical experiments [23]. Example frames of the test videos are shown in Fig. 1. We assigned a current frame as the target and following two frames as the masks, as shown in Fig. 2. Since we were interested in the impact of motion on flicker visibility, we selected the moving object as a region of interest (ROI) and ignored or zeroed other regions. The population activities of excitatory layers ($A_e$) and inhibitory layers ($A_i$) can be expressed following [19]:

$$\tau_e \frac{\partial A_e(x,t)}{\partial t} = -A_e(x,t) + h_e\{w_{ee}(A_e * W_e)(x,t) + w_{ie}(A_i * W_i)(x,t) + I(x,t)\}. \quad (1)$$

$$\tau_i \frac{\partial A_i(x,t)}{\partial t} = -A_i(x,t) + h_i\{w_{ei}(A_e * W_e)(x,t) + w_{ii}(A_i * W_i)(x,t) + I(x,t)\}. \quad (2)$$

The parameters $\tau_e$ and $\tau_i$ are time constants; $w_{ee}$, $w_{ei}$, $w_{ie}$, and $w_{ii}$ are coupling strengths; subscripts $e$ and $i$ mean excitatory and inhibitory layers; $x$ is a two-dimensional position vector in one of the neural layers; and $t$ means time. The symbol '$*$' denotes convolution. The functions $h_{e,i}(x) = x \cdot s_{e,i}$ $(x)$ when $x > 0$, otherwise, $h_{e,i}(x) = 0$, where $s_e$ and $s_i$ are neuronal gain. Recurrent interaction $W_{e,i}$ between the layers is modeled by Gaussian kernels. The input into both layers is computed by

$$I(x,t) = (S * V)(x,t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(x',t)V(x-x')dx' \quad (3)$$

on a video frame $S$, while the input kernel $V$ is the difference of Gaussians [19]. The spatial kernels are set so that each cell receives recurring excitation from its spatial neighbors and recurring inhibition from a larger set of cells around the excitatory set of cells. Hence, these equations show a neural network that describes the recurrent center-surround RF responses of each cell. Finally, the target activation level,

$$T = \int_x A_e(x, r_0) \cdot S_T(x)dx, \quad (4)$$

where $S_T$ is the ROI of the target frame, and $r_0$ is the readout time of the excitatory activity (110ms) after the target onset. We allowed the size of $S_T$ to vary with the video content unlike Hermens's fixed constant target by dividing it by the size of the ROI. We observed only small variations of $T$ when more mask frames were added or when the readout time was varied in the range [80, 140] ms.

The target activation levels are shown in Fig. 3 for the "Bb" and "BMX" test videos. One percent of the original $T$ was shown. Although $T$ effectively captures the suppressed visibility of the target as a function of object motion, it does not discriminate different flicker intensities. We interpret this result to mean that $T$ generally captures the temporal variation of the visibility of distortions on the targets, but the visibility of flicker distortions is apparently also dependent on flicker intensity. For example, when the flicker intensity is small, the visibility of the flicker distortions may vary within a small range, while when the flicker intensity is large, flicker visibility may fluctuate over a large range.

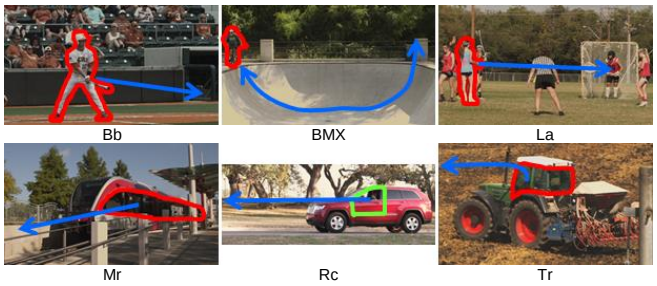


Fig. 1. Example frames of the test videos used in the experiments. The red/green marked regions indicate moving objects, while blue arrows denote the approximate paths of moving objects [23].

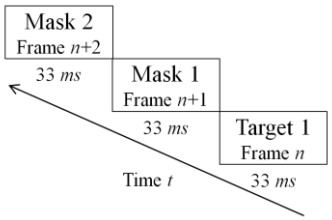


Fig. 2. The target and masks on naturalistic video stimuli.

### 2.2. Initial Flicker Visibility

Physiologically, the target (flicker) in the test videos results from the intensity changes caused by alternations of quantization parameter (QP) in H.264 video compression (e.g., between QP44 and QP26), while the mask (object motion) originates from the intensity changes created by natural object movements. We selected this form of flicker as such rate changes are caused by adaptive rate control algorithms. Thus, the appearance of the flicker distortion is more realistic than, for example, simple luminance flicker. Fig. 4 shows examples of the target and masks for different object motion in a video. Fig. 4a shows frames containing different object motions. Fig. 4b shows the target and the masks. The target is computed as the difference between the luminance frame for QP44 and QP26 at the same frame number, while the masks are the frame differences of consecutive frames. The intensities of the ROI are inversely rendered by magnifying the original intensities by a multiplicative factor of 10 for comparisons. While the target intensity is similar at different object motions, the intensities of the masks are quite different. The inset boxes in Fig. 4c and 4d show the object energies of frame differences for small object motions. QP changes between (QP32, QP26) and (QP44, QP26) caused different flicker intensities.

Since the visibility of flicker distortions also depends on flicker intensity, we shifted and scaled $T$ as a function of quality level changes corresponding to QP alternations. We estimated the perceptual quality level changes using the multi-scale structural similarity (MS-SSIM) index [24], then applied the logarithm to the quality level changes in order to compute flicker quality factor [25]. Let $msssim_{jq}$ denote the MS-SSIM index on video $j$ at QP $q$ (= 26, 32, 38, and 44), $msssim_{ijq}$ be the MS-SSIM index on video $j$ at frame number $i$ and QP $q$, and $N$ be the total number of frames. The flicker intensity, $fi_{jp}$, of video $j$ for a QP change between QP26 and QP $p$ (= 32, 38 and 44) is then given by,

$$msssim_{jq} = \frac{1}{N}\sum_{i=1}^{N} mssim_{ijq}, \tag{5-1}$$

$$fi_{jp} = \log(msssim_{j26} - msssim_{jp}). \tag{5-2}$$

In addition, since video quality remains depend on the video content even when the QP is fixed, to unbias $fi$ from content, we converted $fi$ to Z-scores, then mapped the Z-scores to human flicker visibility scores, $hfv$, using a leave-one-out cross validation protocol [26]. Each time we selected one video from the six videos shown in Fig. 1, trained the necessary parameters on the other five videos, then predicted the initial flicker visibility as follows:

$$z_{fi\_jp} = \frac{fi_{jp} - \mu(fi_j)}{\sigma(fi_j)}, \tag{6-1}$$

$$z_{hfv\_jp} = \frac{hfv_{jp} - \mu(hfv_j)}{\sigma(hfv_j)}, \tag{6-2}$$

$$\hat{z}_{hfv\_jp} = \beta_2 + \frac{\beta_1 - \beta_2}{1 + e^{-(z_{fi\_jp} - \beta_3)/|\beta_4|)}}, \tag{6-3}$$

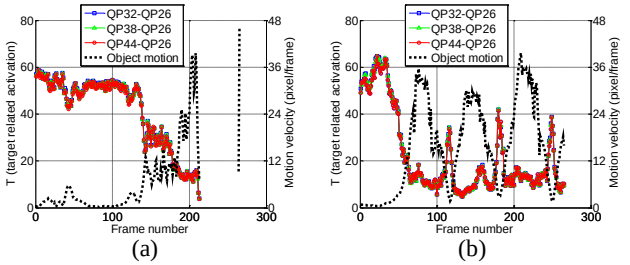


Fig. 3. The target related activation levels: (a) Bb and (b) BMX in spatiotemporal backward masking. Motion velocity is overlapped.

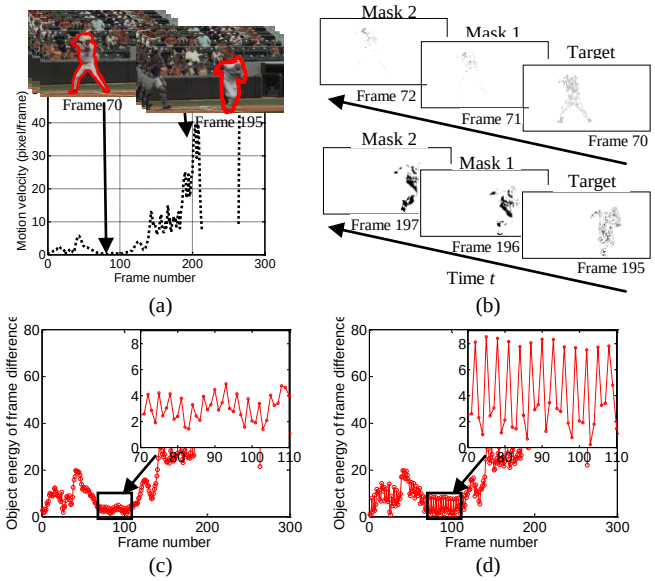


Fig. 4. Target (flicker) and masks (object motion). (a) Frames at different object motions. (b) Magnitude of target and masks. Object energy of frame differences: (c) between QP32 and QP26 and (d) between QP44 and QP26 at the same object motion.

$$\mu(hfv_j) = \alpha_2 + \alpha_1 \times \hat{z}_{hfv\_jp}, \tag{6-4}$$

$$\sigma(hfv_j) = \frac{1}{M}\sigma(hfv_{\{j\}^c}),\ j=1, 2, \ldots 6, \tag{6-5}$$

$$hfv_{jp} = \hat{z}_{hfv\_jp} \times \sigma(hfv_j) + \mu(hfv_j) \tag{6-6}$$

where $z_{fi\_jp}$ and $z_{hfv\_jp}$ are Z-scores for $fi_{jp}$ and $hfv_{jp}$; $\{\mu(fi_j), \sigma(fi_j)\}$ and $\{\mu(hfv_j), \sigma(hfv_j)\}$ are {mean, standard deviation} at video $j$ for $fi_{jp}$ and $hfv_{jp}$, respectively; and the hat symbol "^" means predicted quantities. We predicted the parameters $\alpha$ and $\beta$ following [27] using a least squares fit of $\mu(hfv)$ at $\{j\}^c$, where $j$ = 1, 2, …, 6. $\hat{\sigma}(hfv_j)$ is the mean of $\sigma(hfv)$ at $\{j\}^c$. We used $\widehat{hfv}_{jp}$ as the initial flicker visibility at video $j$ and QP $p$. Since we obtained the initial flicker visibility at the 51th frame [23], we used the value of $hfv$ at the 51th frame to predict $\widehat{hfv}_{jp}$. Next, since the range of flicker distortion visibility also depends on the initial flicker visibility, we shifted and scaled the $T$ values as follows:

$$T_{ss\_jp} = (hfv_{jp}/100) \times (T - T^{51}) + hfv_{jp}, \tag{7}$$

where $T_{ss\_jp}$ is the shifted and scaled $T$ at video $j$ and QP $p$, and $T^{51}$ is $T$ at the 51th frame. We define $T_{ss}$ as a general change pattern of the flicker visibility.

### 2.3. Prediction Algorithm of Flicker Visibility

Human's sensitivity to flicker distortions is accumulated or attenuated after prolonged exposure to flickering stimuli [28]. Due to the limited dynamic range of a neuron, visual processing in the retina efficiently adjusts or adapts visual stimuli. When a viewer is exposed to large flicker distortions for a longer period of time (e.g., 100ms), flicker visibility may be affected by "visual persistence," [29] whereby a visual stimulus is retained for a period of time beyond the termination of the stimulus. Conversely, when small flicker distortions are prolonged, the visual systems dynamically control the flicker sensitivity and allocate a finite range of neural signaling, so an observer's flicker sensitivity may be attenuated [28]. This accumulation or adaptation process may be found in viewers' responses to time-varying video quality as a "recency effect" [30] or "temporal hysteresis" [31]. We extend the flicker accumulation and adaptation by accounting for the impact of motion on flicker visibility.

We filtered the obtained $T_{ss}$ and object motion using a temporal Gaussian weighting function [32] to remove noise. The Gaussian window duration was one second. We then applied visual persistence and recency effects to adopt the influence of flicker accumulation or adaptation as a function of stimuli duration and object motion as shown in Algorithm 1. Let $FV^i$ represent the predicted flicker visibility at the $i^{\text{th}}$ frame and $mv_{th}$ be a threshold velocity that sharply decreases flicker visibility. We obtained $mv_{th}$ (=10.1945) from the human studies [23] using the average of all measured $mv_{th}$ and found the results do not significantly differ when $mv_{th}$ varies in the range [8, 12] pixel/frame. The parameters $t_1$, $t_2$, and $t_3$ are time durations where $mv^i \le mv_{th}$, $mv^i \ge mv_{th}$ and $mv^i \ge mv^{i-1}$, and $mv^i \ge mv_{th}$ and $mv^i < mv^{i-1}$, respectively.

When $mv^i \le mv_{th}$, accumulation or adaptation effects are large, while the impact of motion on flicker visibility is small. Since visual persistence almost changes linearly with log stimulus duration at a given visual intensity, $FV^i$ may be computed as a logarithmic function of time duration $t_1$ from a base-level, $T_{ss}^i$. $\gamma$ is a rate parameter of accumulation or attenuation, which was set to -0.4356, 0.7599, and 3.3629 using a least square fit of human experiment data [23] when the QP changes from 32, 38, 44, to 26 respectively.

When $mv^i \ge mv_{th}$, both motion and flicker accumulation or adaptation affect flicker visibility. We accounted for the effect of motion by measuring the difference of $T_{ss}$ (e.g., $T_{ss}^{i+1} - T_{ss}^i$) to estimate $\gamma$. To include the memory effect, we define a memory component $x(i)$ at each frame $i$ by averaging the difference of $T_{ss}$ over the time duration $t_2$ or $t_3$. We then build a flicker visibility change rate $y(i)$ at each frame $i$ using $T_{ss}^i - T_{ss}^{i-1}$. We linearly combine the past memory component $x(i)$ and the current change rate $y(i)$, thereby yielding an overall visibility accumulation or adaptation rate that seeks to explain recency effects at frame $i$. The linear factor $\lambda$ was 0.7. The results do not vary significantly when $\lambda$ is above 0.5, similar to [31]. Finally, $FV^i$ is iteratively solved to ensure that the predicted flicker visibility is smoothly varying.

**Algorithm 1** Flicker visibility prediction algorithm

**Inputs:** $N$, $mv$, $mv_{th}$, $t_1 = 0$, $t_2 = 0$, $t_3 = 0$, $j = 0$, $\gamma$, and $T_{ss}$

**for** $i = 51 : N$
  **if** $mv^i \le mv_{th}$ **and** $j = 0$
    $t_1 = t_1 + 1$, $t_2 = 0$, $t_3 = 0$
    $FV^i = T_{ss}^i + \gamma \times \log(t_1)$, where $\gamma$ = [-0.4356, 0.7599, 3.3629]
  **else**
    **if** $mv^i \ge mv_{th}$ **and** $mv^i \ge mv^{i-1}$
      $t_2 = t_2 + 1$, $t_1 = 0$, $t_3 = 0$, $j = 1$, $k = 2$
    **elseif** $mv^i \ge mv_{th}$ **and** $mv^i < mv^{i-1}$
      $t_3 = t_3 + 1$, $t_1 = 0$, $t_2 = 0$, $j = 1$, $k = 3$
    **elseif** $mv^i \le mv_{th}$ **and** $j = 1$
      $t_1 = t_1 + 1$, $t_2 = 0$, $t_3 = 0$, $j = 1$, $k = 1$
    **end**
    $x(i) = \text{mean}[(T_{ss}^{i-t_k}, \ldots, T_{ss}^{i-1}) - (T_{ss}^{i-t_{k-1}}, \ldots, T_{ss}^{i-2})]$
    $y(i) = T_{ss}^i - T_{ss}^{i-1}$
    $FV^i = FV^{i-1} + [\lambda y(i) + (1-\lambda)x(i)] \times \log(t_k)$, $\lambda = 0.7$
  **end**
**end**

## 3. PERFORMANCE EVALUATION

We evaluated the performance of the proposed model on 24 test videos, which contain six reference videos and the corresponding 18 flicker distorted videos simulated by alternating QP pairs at (QP44, QP26), (QP38, QP26), and (QP32, QP26), respectively. As mentioned earlier, this type of flicker was used as it resembles distortions that can be caused by video rate adaptation algorithms. We refer to [23] for the details of the human psychophysical studies.

The predicted flicker visibility from the model and the 95% confidence interval (CI) of the measured flicker visibility by humans were shown in Fig. 5. We compared the results in the frame interval [51, 208] for "Bb" and [51, 265] for other videos since observers required at least 1.67 seconds (50 frames) to judge the initial flicker visibility, since when moving objects disappeared, no data was obtained, and since the last 35 frames were shifted to account for a lag response. When the object motion was small, flicker visibility was largely affected by flicker intensities and persistent durations, while the object motion was large, flicker visibility was strongly affected by motion.

One interesting observation is that large jumps of flicker visibility on the videos "La" and "Mr" were observed. These may have arisen from the drastic changes of size of the moving objects. For example, the occlusion of the player by a referee in "La" and the looming train in "Mr" influenced the prediction of the $T$ and subsequently may have affected the prediction of flicker visibility. Overall, the proposed model can effectively predict variations of flicker visibility as reported by humans.

We showed model performance in Table 1 and 2 using the Pearson's linear correlation coefficient (LCC) and the Spearman's rank ordered correlation coefficient (SROCC) between the predicted and the perceived flicker visibility on the test videos after non-linear regression, as described in [33]. Although results vary depending on the contents and QP alternations, the average of LCC values, 0.8573, is high.

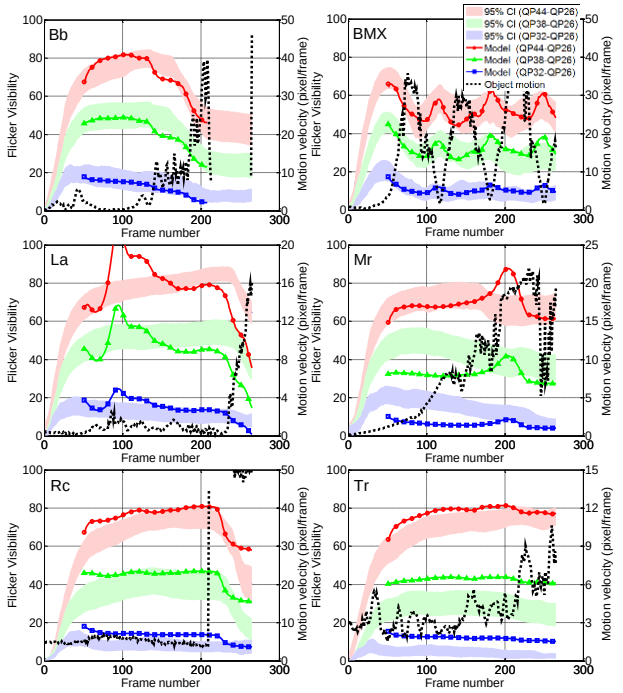


Fig. 5. Comparison between the predicted flicker visibility by the model and the 95% CI of the measured flicker visibility by humans.

Table 1. LCC of the proposed model on the test videos

| | Bb | BMX | La | Mr | Rc | Tr |
|---|---|---|---|---|---|---|
| QP44 - QP26 | 0.9534 | 0.7815 | 0.6477 | 0.8712 | 0.9183 | 0.8952 |
| QP38 - QP26 | 0.9880 | 0.7523 | 0.8721 | 0.8515 | 0.9034 | 0.9156 |
| QP32 - QP26 | 0.9796 | 0.7959 | 0.8187 | 0.7773 | 0.7882 | 0.9267 |

Table 2. SROCC of the proposed model on the test videos

| | Bb | BMX | La | Mr | Rc | Tr |
|---|---|---|---|---|---|---|
| QP44 - QP26 | 0.7069 | 0.6656 | 0.0640 | 0.0620 | 0.9284 | 0.4922 |
| QP38 - QP26 | 0.8125 | 0.3775 | 0.1050 | 0.1440 | 0.5777 | 0.6103 |
| QP32 - QP26 | 0.9880 | 0.2028 | 0.7755 | 0.5445 | 0.6975 | 0.8840 |

## 4. CONCLUSION AND FUTURE WORK

We proposed a model of flicker visibility on naturalistic videos using spatiotemporal backward masking and neural adaptation processes. Results show that the proposed model accurately predicts flicker visibility as perceived by humans. Since many factors could underlie the discrepancy between model predictions and human percepts of flicker visibility, such as the experimentally obtained model parameters and the drastic size changes of moving objects, further research is necessary. Our future work will seek to connect the visibility of flicker distortions with perceptual VQA models.

## 5. REFERENCES


[1] L. K. Choi, Y. Liao, and A. C. Bovik, "Video QoE metrics for the compute continuum," *IEEE MMTC E-Letter*, vol. 8, no. 5, pp. 26–29, Sep. 2013.

[2] M. Yuen and H. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Process.*, vol. 70, no. 3, pp. 247–278, Nov. 1998.

[3] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 335–350, Feb. 2010.

[4] A. Ninassi, O. L. Meur, P. L. Callet, and D. Barba, "Considering temporal variations of spatial visual distortions in video quality assessment," *IEEE J. Sel. Top. Signal Process.*, vol. 3, no. 2, pp. 253–265, Apr. 2009.

[5] A.C. Bovik, "Automatic prediction of perceptual image and video quality," *Proc. IEEE*, vol. 101, no. 9, pp. 2008-2024, Sep. 2013.

[6] Breitmeyer, Bruno, and H. Ogmen, *Visual masking: Time slices through conscious and unconscious vision*. vol. 41, Oxford University Press. 2006.

[7] S. E. Palmer, *Vision Science*. Cambridge, MA, USA: MIT Press, 1999.

[8] G. E. Legge and J. M. Foley, ''Contrast masking in human vision,'' *J. Opt. Soc. Amer.*, vol. 70, no. 12, pp. 1458–1470, Dec. 1980.

[9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[10] A. N. Netravali and B. Prasada, ''Adaptive quantization of picture signals using spatial masking,'' *Proc. IEEE*, vol. 65, no. 4, pp. 536–548, Apr. 1977.

[11] M. D. Swanson, B. Zhu, and A. H. Tewfik, ''Multiresolution scene-based watermarking using perceptual models,'' *J. Sel. Areas Commun.*, vol. 16, no. 4, pp. 540–550, May 1998.

[12] J. W. Suchow and G. A. Alvarez, "Motion silences awareness of visual change," *Current Biol.*, vol. 21, no. 2, pp.140-143, 2011.

[13] L. K. Choi, A. C. Bovik, and L. K. Cormack, "A flicker detector model of the motion silencing illusion," *J. Vis.*, vol. 12, no. 9, pp. 777, May 2012.

[14] A. J. Seyler and Z. Budrikis, "Detail perception after scene changes in television image presentations," *IEEE Trans. Inf. Theory*, vol.11, no.1, pp.31-43, Jan. 1965.

[15] A. Puri and R. Aravind, "Motion-compensated video with adaptive perceptual quantization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 1, pp. 351-378, Dec. 1991.

[16] B. G. Haskell, F. W. Mounts, and J. C. Candy, "Interframe coding of videotelephone pictures," *Proc. IEEE*, vol. 60, pp. 792-800, Jul. 1972.

[17] I. Hontsch and L. J. Karam, "Adaptive image coding with perceptual distortion control," *IEEE Trans. Image Process.*, vol.11, no. 3, pp. 213-222, Mar. 2002.

[18] Y. Zhao, L. Yu, Z. Chen, and C. Zhu, "Video quality assessment based on measuring perceptual noise from spatial and temporal perspectives," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 12, pp. 1890-1902, Dec. 2011.

[19] F. Hermens, G. Luksys, W. Gerstner, M. Herzog, and U. Ernst, "Modeling spatial and temporal aspects of visual backward masking," *Psychological review*, vol. 225, no. 1. pp. 83-100, 2008.

[20] H. R. Wilson and J. D. Cowan, "A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue," *Kybernetik*, vol. 13, no.2, 1973.

[21] T. J. Spencer and R. Shuntich, "Evidence for an interruption theory of backward masking," *J. Exp. Psychol.*, 85(2), pp.198-203, 1970.

[22] A. R. Reibman, and D. Poole, "Predicting packet-loss visibility using scene characteristics," *Packet Video* 2007, pp.308-317, 12-13 Nov. 2007.

[23] L. K. Choi, L. K. Cormack, and A. C. Bovik, "On the visibility of flicker distortions in naturalistic videos," in *Proc. QoMEX* 2013, pp. 164-169.

[24] Z. Wang, L. Lu, and A. Bovik, "Foveation scalable video coding with automatic fixation selection," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 243–254, Feb. 2003.

[25] J. E. Farrell, B. L. Benson, and C. R. Haynie, "Prediction flickering thresholds for video display terminals," in *Proc. SID*, vol.28, no. 4, 1987.

[26] L. Breiman, and P. Spector, "Submodel selection and evaluation in regression. the x-random case," *Int. Stat. Review*, vol. 60, no. 3, pp. 291-319, 1992.

[27] Final Report from the Video Quality Experts Group on the Validation of Objective Quality Metrics for Video Quality Assessment Phase I, VQEG, 2000, [Online]. Available: http://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseI

[28] S. Shady, D. I. A. MacLeod, H. S. Fisher, "Adaptation from invisible flicker," *Proc. PNAS*, vol. 101, no. 14, pp. 5170-5173, 2004.

[29] R. W. Bowen, J. Polar, and L. Matin, "Visual persistence: Effects of flash luminance, duration and energy," *Vision Res.*, vol. 14, no. 4, pp. 295-303, 1974.

[30] J. R. Anderson and M. Matessa, "A production system theory of serial memory," *Psychol. Rev.*, vol. 104, no. 4, pp. 728-748, 1997.

[31] K. Seshadrinathan and A. C. Bovik, "Temporal hysteresis model of time varying subjective video quality," in *Proc. ICASSP*, 2011, pp. 1153–1156.

[32] H. G. Longbothan and A. C. Bovik, "Theory of order statistic filters and their relationship to linear fir filters," *IEEE Trans. Acoust., Signal Process.*, vol. 37, no. 2, pp. 275-287, 1989.

[33] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.