Perceptual Flicker Visibility Prediction Model

Lark Kwon Choi and Alan C. Bovik; Laboratory for Image and Video Engineering (LIVE), Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, USA

Abstract

The mere presence of spatiotemporal distortions in digital videos does not have to imply quality degradation since distortion visibility can be strongly reduced by the perceptual phenomenon of visual masking. Flicker is a particularly annoying occurrence, which can arise from a variety of distortion processes. Yet flicker can also be suppressed by masking. We propose a perceptual flicker visibility prediction model which is based on a recently discovered visual change silencing phenomenon. The proposed model predicts flicker visibility on both static and moving regions without any need for content-dependent thresholds. Using a simple model of cortical responses to video flicker, an energy model of motion perception, and a divisive normalization stage, the system captures the local spectral signatures of flicker distortions and predicts perceptual flicker visibility. The model not only predicts silenced flicker distortions in the presence of motion, but also provides a pixel-wise flicker visibility index. Results show that the predicted flicker visibility model correlates well with human percepts of flicker distortions tested on the LIVE Flicker Video Database and is highly competitive with current flicker visibility prediction methods.

Introduction

Digital videos are increasingly pervasive due to the rapid proliferation of video streaming services, video sharing in social networks, and the global increase of mobile video traffic [1], [2]. The dramatic growth of digital videos and user demand for high-quality video have necessitated the development of precise automatic perceptual video quality assessment (VQA) tools to help provide satisfactory levels of Quality of Experience (QoE) to the end user [3].

To achieve optimal video quality under limited bandwidth and power consumption, video coding technologies commonly employ lossy coding schemes, which cause compression artifacts that can lead to degradation of perceptual video quality [4]. In addition, compressed videos can suffer from transmission distortions, including packet losses and playback interruptions triggered by channel throughput fluctuations. Since humans are generally the ultimate arbiter of the received videos, predicting and reducing perceptual visual distortions of compressed digital videos is of great interest [5].

Researchers have performed a large number of subjective studies to understand essential factors that influence video quality by analyzing compression artifacts or transmission distortions of the compressed videos [6], by investigating dynamic time varying distortions [7], and by probing the time varying subjective quality of rate adaptive videos [8]. Substantial progress has also been made toward understanding and modeling low-level visual processes in the vision system extending from the retina to primary visual cortex and extra-striate cortex [9]. As a result, perceptual models of disruptions to natural scene statistics [10] and of visual masking [11] have been widely applied to predict perceptual visual quality.

Spatial distortions are effectively predicted by VQA algorithms such as SSIM [12], VQM [13], MOVIE [14], STRRED [15], and Video-BLIINDS [16]. Spatial masking is well-modeled in modern perceptual image and video quality assessment tools, video compression, and watermarking. However, temporal visual masking is not well-modeled although one type of it has been observed to occur near scene changes [17], and been used in the context of early video compression methods [18-20]. Among temporal distortions, flicker distortion is particularly challenging to predict and often occurs on low bit-rate compressed videos.

Flicker distortion is (spatially local or global) temporal fluctuation of luminance or chrominance in videos. Local flicker occurs mainly due to coarse quantization, varying prediction modes, mismatching of inter-frame blocks, improper deinterlacing, and dynamic rate changes caused by adaptive rate control methods [21-25]. Mosquito noise and stationary area fluctuations are also often categorized under local flicker. Mosquito noise is a joint effect of object motion and time-varying spatial artifacts such as ringing and motion prediction errors near high-contrast sharp edges or moving objects, while stationary area fluctuations result from different types of prediction, quantization levels, or a combination of these factors on static regions [4], [21].

Current flicker visibility prediction methods that operate on a compressed video measure the Sum of Squared Differences (SSD) between the block difference of an original video and the block difference of a compressed video. The block difference is obtained between successive frames on macroblocks. When the sum of squared block differences on an original video falls below a threshold, a static region is indicated [22]. The ratio between luminance level fluctuation in the compressed video and that in the original video has also been used [23]. To improve the prediction of flicker-prone blocks, a normalized fraction model was proposed [24], where the difference of SSDs between the original and compressed block differences is divided by the sum of the SSDs. These methods have the virtue of simplicity, but the resulting flicker prediction performance is limited and content-dependent. Another method included the influence of motion on flicker prediction, where motion compensation was applied prior to SSD calculation [25]. The mean absolute discrete temporal derivatives of the average DC coefficient of DCT blocks was used to measure sudden local changes (flicker) in a VQA model [16]. Current flicker prediction methods are limited to block-wise accuracy. Further, human visual system (HVS)-based perceptual flicker visibility e.g., considering temporal visual masking, has not yet been extensively studied.

Recently, Suchow and Alvarez [26] demonstrated a striking "motion silencing" illusion, in the form of a powerful temporal visual masking phenomenon called change silencing, where the salient temporal changes of objects in luminance, color, size, and shape appear to cease in the presence of large object motions. This motion-induced failure to detect change not only suggests a tight coupling between motion and object appearance, but also reveals that commonly occurring temporal distortions such as flicker may

be dramatically suppressed by the presence of motion. To understand the mechanism of motion silencing, physiologically plausible explanations have been proposed [26-29]. However, since the effect has only been studied on highly synthetic stimuli such as moving dots, we performed a series of human subjective studies on naturalistic videos, where flicker visibility is observed to be strongly reduced by large coherent object motions [30-33]. A consistent physiological and computational model that detects motion silencing might be useful to probe perceptual flicker visibility on compressed videos.

In this paper, we propose a new perceptual flicker visibility prediction model based on motion silencing. The new perceptual flicker visibility prediction model is a significant step towards improving the performance of VQA models by making possible a model of temporal masking of temporal distortions. The new model measures the bandpass filter responses to a reference video and a corresponding flicker video using a localized multiscale 3D space time Gabor filter bank [34], [35], a spatiotemporal energy model of motion perception [36], and a divisive normalization model of nonlinear gain control in primary visual cortex [37]. We observed that flicker produces locally separated spectral signatures that almost lie along the same orientation as the motion tuned plane of the reference video but at a distance. The captured V1 responses for the flicker induced spectral signatures generally decreased when object speeds increase. Next, we measured the local difference of bandpass responses at each space-time frequency orientation and defined the sum of the magnitude responses as a perceptual flicker visibility index. The proposed model predicts temporal masking effects on flicker distortions and thereby shows highly competitive performance against previous flicker visibility prediction methods.

Background: Motion Perception

Motion perception is the process of inferring the speed and direction of moving objects. Since motion perception is important for understanding flicker distortions in videos, we model motion perception in the frequency domain. Watson and Ahumada [38] proposed a model of how humans sense the velocity of moving images, where the motion-sensing elements appear locally tuned to specific spatiotemporal frequencies.

Assuming that complex motions of video without any scene changes can be constructed by piecing together spatiotemporally localized image patches undergoing translation, we can model the local spectral signatures of videos when an image patch moves [38]. An arbitrary space-time image patch can be represented by a function $a(x, y, t)$ at each point x, y, and time t, and its Fourier transform by $A(u, v, w)$ where u, v, and w are spatial and temporal frequency variables corresponding to x, y and t, respectively. Let λ and ϕ denote the image patch horizontal and vertical velocity components. When an image patch translates at constant velocity $[\lambda, \phi]$, the moving video sequence becomes $b(x, y, t) = a(x - \lambda t, y - \phi t, t)$. The spectrum of a stationary image patch lies on the u, v plane, while the Fourier transform shears into an oblique plane through the origin when the image patch moves. The orientation of this plane indicates the speed and direction of motion.

Prediction of Perceptual Flicker Visibility

Linear Decomposition

Natural environments are inherently multi-scale and multi-orientation, and objects move multi-directionally at diverse speeds. To efficiently encode visual signals, the vision system decomposes

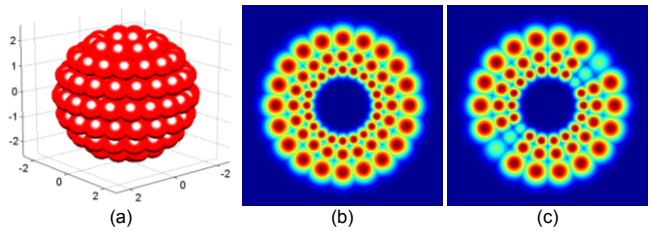

Figure 1. Gabor filter bank in the frequency domain. (a) Geometry of the Gabor filter bank. (b) A slice of the Gabor filter bank along the plane of zero temporal frequency. (c) A slice of the Gabor filter bank along the plane of zero vertical spatial frequency.

the visual world over scales, orientations, directions, and speeds. Cortical neurons in Area V1 are selective for spatiotemporal frequency and orientation, while neurons in Area MT are selective for the velocity of visual stimuli [37], [39]. Since the responses of simple cells in Area V1 are well-modeled as linear and bandpass [34], [35], linear decompositions are widely used to model the spatiotemporal responses to video signals [14], [40].

The receptive field profiles of V1 simple cells are well modeled by Gabor filters [34], [35]. Hence, we used a bank of spatiotemporally separable Gabor filters to model the responses of V1 simple cells to videos. A 3D spatiotemporal separable Gabor filter $h(\mathbf{x})$ is the product of a complex exponential with a Gaussian envelope:

$$h(\mathbf{x}) = \frac{1}{(2\pi)^{3/2}\,|\sum|^{1/2}} \exp\left(-\frac{\mathbf{x}^T \sum^{-1} \mathbf{x}}{2}\right) \exp\left(j\mathbf{U}_0\mathbf{x}\right), \qquad (1)$$

where $\mathbf{x} = (x, y, t)$ is a spatiotemporal coordinate in a video sequence, and $\mathbf{U_0} = (U_0, V_0, W_0)$ is the center frequency of the Gabor filter. $\sum$ is the covariance matrix of the Gaussian envelope.

We implemented a Gabor filter bank as in [14] and [40]. Three scales of filters were deployed, with 57 filters at each scale on the surface of a sphere centered at the space-time frequency origin, as shown in Fig. 1. The largest radial center frequency was 0.7π radians per sample, and the filters were sampled out to a width of three standard deviations. A total of 171 filters were used: 10, 18, 15, 10, and 4 filters were tuned to five different speeds, s = tan (φ), where the vertical angle φ = (0 20, 40, 60 and 80) degrees and orientations θ at every 18, 20, 24, 36, and 90 degrees, respectively. The number of oriented filters was determined such that adjoining filters intersected at one standard deviation following [40]. We also included a Gaussian filter centered at the frequency origin to capture the low frequencies in the video. The standard deviation of the Gaussian was selected so that the Gaussian would intersect the coarsest scale of bandpass filters at one standard deviation [14].

Modeling Visual Cortical Neurons

The responses of V1 neurons were modeled using the spatiotemporal energy model in [36] with divisive normalization [37]. The motion energy within a spatiotemporal frequency band was extracted by squaring the responses of quadrature (sine and cosine) Gabor filter components and summing them:

$$E(\varphi,\theta) = \left[h_{\sin}(\varphi,\theta) * I\right]^2 + \left[h_{\cos}(\varphi,\theta) * I\right]^2, \qquad (2)$$

where $h_{\sin}(\varphi, \theta)$ and $h_{\cos}(\varphi, \theta)$ are the sine and cosine Gabor filters at φ and θ, respectively, and I is the luminance level of the video.

The quantity (2) models the response of an individual neuron to a particular band of spatiotemporal frequencies. In order to

agglomerate the combined responses of all cortical neighborhoods that include cells tuned for the full range of orientations and directions, the response of each neuron is normalized to limit its dynamic range of responses without altering the relative responses of neurons in the pool [37]. The energy response of the n^{th} simple cell S_n is divided by the sum of the neighboring energy responses:

$$S_n(\varphi,\theta) = K\frac{E_n(\varphi,\theta)}{\sum_{\varphi,\theta} E_n(\varphi,\theta)+\sigma^2}, \quad (3)$$

where K determines the maximum attainable response, and σ is a semi-saturation constant. Here $K = 4$ and $\sigma = 0.2$ as was used in [37] in agreement with recorded physiological data. The model V1 complex cell responses C_n are found by averaging the responses (3) along scales on constant space-time frequency orientations:

$$C_n(\varphi,\theta) = \sum_m c_{nm} S_m(\varphi,\theta), \quad (4)$$

where c_{nm} (> 0) are weighting factors. We used constant values although they could be Gaussian in spatial distance [37].

Spectral signatures of flicker videos

To understand the distributions of the spectral signatures of flicker videos, we generated translational motion videos by moving an image patch (1280 × 720 pixels) from a large static image (4140 × 2330 pixels) at constant speeds (e.g., 0, 1, 2, and 3 pixels per frame) in horizontal, vertical, and diagonal directions. Then, we simulated quantization flicker by alternately compressing the video with different QP pairs (e.g., QP26 and QP44) every 3 frames using an H.264 codec. We measured the model V1 responses for these videos with and without flicker distortions, separately.

For the reference video, the spectral signatures formed a flat plane on the u, v axes when there was no motion. The plane tilted with respect to the u, v axes when motion was increased, as illustrated in Fig. 2(a). For the flicker videos, we observed that the spectral signatures separated along the same orientation as the motion tuned plane of the reference video but at a distance from the reference spectral plane, as shown in Fig. 2(b). Larger flicker intensities (e.g., caused by more separated QP pairs such as QP44 and QP26) produced larger responses to the flicker-induced spectral signatures. We also observed that the model V1 responses for the flicker-induced spectral signatures generally decreased when motion increased.

We executed the same spectral analysis by simulating quantization flicker on the videos in the LIVE VQA database [6] and observed similar results. Changing patterns in the spectral signatures of flicker videos (Fig. 2) also agreed with the changing patterns of the spectral signatures from the physiological experiments on motion silencing using the dot stimuli [29].

Perceptual Flicker Visibility Index

We developed a perceptual flicker visibility index based on how flicker changes the spectral signatures of a video and how motion influences the resulting V1 responses. Shifted or separated spectral signatures not present in a reference video might cause flicker distortions. Therefore, we devised an approach to capture perceptual flicker visibility by measuring locally shifted energy deviations relative to those on the reference video at each subband.

Next, we define the sum of the difference as a perceptual flicker visibility index

$$FV = \sum_{\varphi,\theta} \left| C^r(\varphi,\theta,\mathbf{x}) - C^d(\varphi,\theta,\mathbf{x}) \right|, \quad (5)$$

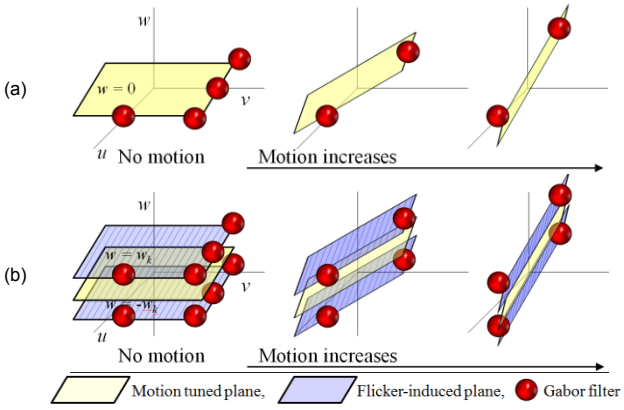

Figure 2. Schematic illustration of the spectral signatures that form motion tuned planes: (a) reference videos and (b) flicker videos.

where $C^r(\varphi,\theta,\mathbf{x})$ and $C^d(\varphi,\theta,\mathbf{x})$ are the model V1 responses on the reference and flicker videos in (4), respectively. Larger values of FV indicate larger flicker visibility, while zero implies no flicker.

Results and Performance Evaluation

We tested the proposed perceptual flicker visibility model on the LIVE Flicker Video Database [31], [32]. The database contains 6 reference videos and 6 flicker distorted versions of each reference, for a total of 36 videos in Phase I. Quantization flicker was simulated by periodic changes of H.264 compression level by varying the QP value. We enlisted 43 subjects who rated their continuous-time flicker percepts on test videos. We refer to [30-32] for the details of flicker simulations and the human psychophysical studies.

Since some previous flicker visibility prediction models [22], [23] are only able to predict flicker visibility on static segments of video sequences, it is not possible to directly compare them with the performance of the proposed model. For the other models, we first compared the predicted flicker visibility maps obtained by each model. Next, we evaluated the performance of the proposed model against the results of the human subjective study using a correlation analysis.

Figure 3 shows flicker visibility maps obtained by the method in [22], [25], [16], and by the proposed model on one of the videos in the LIVE Flicker Video Database. Brighter regions indicate larger flicker, while the gray slashed areas denote regions that cannot be predicted for flicker visibility. In Fig. 3, the left column shows results on static or small motion scenes, while the right column presents the results on large motion scenes. As shown in Fig. 3(b), Fan's method [22] predicts flicker visibility only on static regions, while the other methods predict flicker on all regions. The flicker maps in [22], [25], and [16] are limited to block-wise accuracy, but the proposed model can predict flicker visibility with pixel precision, as shown in Figs. 3(b)-(e). Furthermore, the proposed model predicts silenced flicker distortions in the presence of large motion in agreement with the human responses. Although the methods in [25] and [16] consider moving objects using motion compensation, the estimated results were less predictive of human percepts of flicker. For example, the predicted results in [25] generally show small or almost zero flicker visibility on both small and large motion (e.g. the baseball player in Fig. 3(c)). In [25], the SSDs between the original and compressed block differences after motion compensation were divided by the SSD of the original block differences. This division

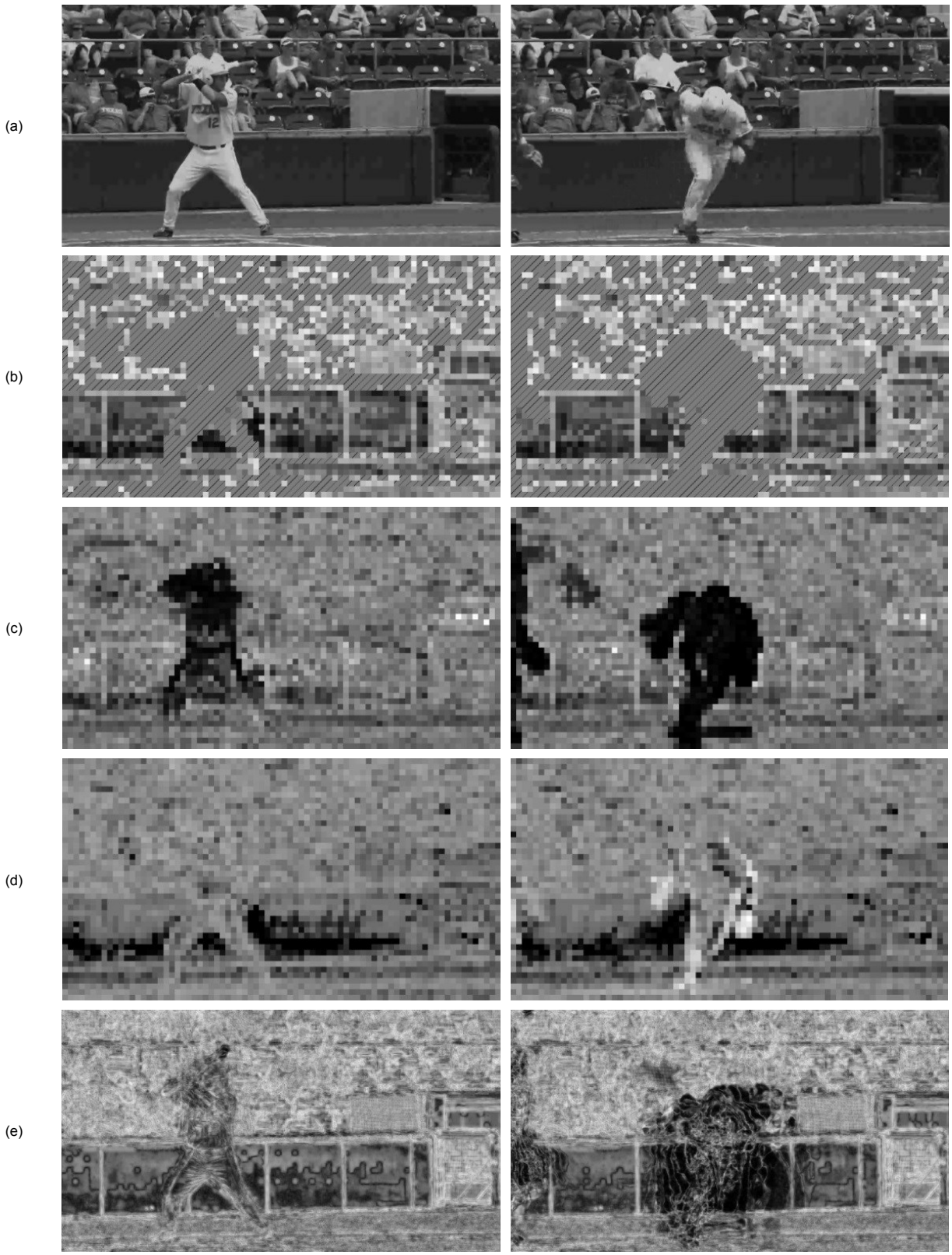

Figure 3. Comparison of predicted flicker visibility maps in the presence of static or small motion (left) and large motion (right). (a) Frames from a flicker distorted video. Flicker visibility maps: (b) Fan et al. [22], (c) Vo et al. [25], (d) Saad et al. [16], and (e) the proposed model. The gray slashed areas in (b) denote regions where the model cannot predict flicker visibility. Brighter regions indicate larger flicker visibility. The flicker maps were logarithmically compressed for rendering.

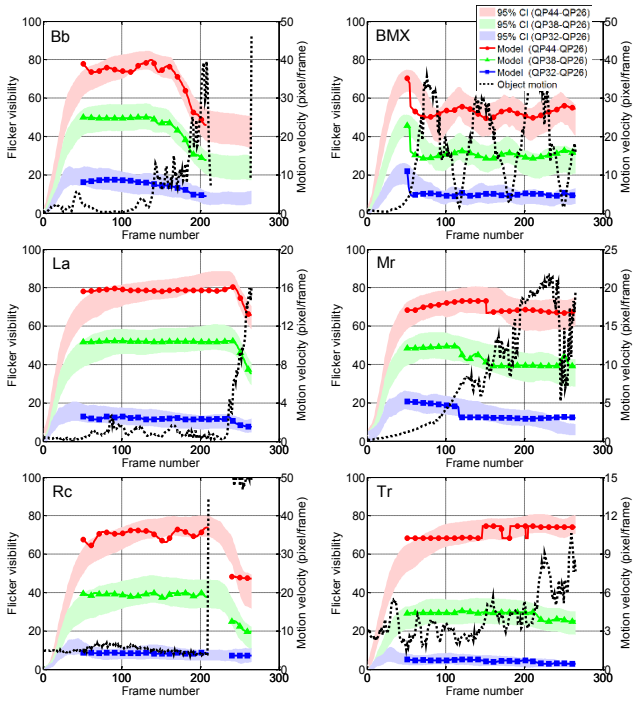

Figure 4. Predicted flicker visibility from the proposed model. Each colored band indicates the 95% confidence interval (CI) of perceived flicker from the human subjects at different QP alternations – (QP44, QP26) ( ), (QP38, QP26) ( ), and (QP32, QP26) ( ). Each solid line with different markers represents the predicted flicker visibility from the model – (QP44, QP26) (●), (QP38, QP26) (▲), and (QP32, QP26) (■), respectively.

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

Method	QP alternation	Bb	BMX	La	Mr	Rc	Tr
Choi *et al*. [31]	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
Proposed Model	QP44 - QP26	0.9614	0.5532	0.5257	0.7202	0.8603	0.6690
	QP38 - QP26	0.9918	0.4977	0.8751	0.8630	0.9206	0.8871
	QP32 - QP26	0.9326	0.4416	0.6999	0.7701	0.5833	0.8216

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

Method	QP alternation	Bb	BMX	La	Mr	Rc	Tr
Choi *et al*. [31]	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
Proposed Model	QP44 - QP26	0.7311	0.6033	0.1919	0.5929	0.5381	0.6718
	QP38 - QP26	0.8462	0.3439	0.1771	0.7795	0.3301	0.5443
	QP32 - QP26	0.8833	0.1237	0.6002	0.7215	0.5053	0.7926

may yield a small degree of flicker sensitivity on moving objects. By contrast, the method in [16] predicts heightened flicker visibility in the presence of large motion, as shown in Fig. 3(d) on the baseball player. This might result from mismatches between inter-frame blocks or large residuals between consecutive frames. On other tested flicker videos, we observed similar results.

Predicted flicker visibility values delivered by our proposed model were also compared against the human flicker visibility scores. As shown in Fig. 4, the trajectory of flicker visibility predicted over time by the model correlates well with the flicker visibility perceived by human subjects. The frame intervals [51, 208] for "Bb" and [51, 265] for the other videos were compared, since subjects needed at least 50 frames to begin to rate initial flicker visibility after each test video began. No data was obtained when objects disappeared in a scene. The frames were shifted to account for a lag response. We averaged the predicted flicker visibility over the moving object (e.g., the baseball player) and filtered the average by a temporal Gaussian weighting function, where the Gaussian window duration was one second. Then, the filtered averages were non-linear regressed [41] to match the perceived flicker visibility range in the human subjective study. Although the proposed model is able to predict small differences in flicker visibility over short periods (and long), flicker visibility rated by the subjects is very smooth owing to the limited speed of mouse movement using a hand. Hence, we applied temporal filtering before non-linearity regression. In agreement with the human percepts of flicker, the model can effectively predict temporal variations of flicker visibility, in which flicker visibility is strongly reduced when motion increases.

Tables 1 and 2 show the performance of the proposed model using the Pearson's Linear Correlation Coefficient (PLCC) and the Spearman Rank Order Correlation Coefficient (SROCC) after logistic nonlinear regression [41], respectively. Performance was compared with the neural network model (NNM) [31] on the LIVE Flicker Video Database. Although NNM predicts flicker visibility better in terms of correlations, the proposed model is highly competitive. Further, the proposed model has several advantages: the NNM method cannot produce a predicted flicker visibility map indicating details of spatiotemporal variations of flicker visibility, and it requires human results to train neural network parameters. By contrast, the proposed model predicts a detailed perceptual flicker visibility map without any prior human results.

Conclusion and Future Work

We have presented a perceptual flicker visibility prediction model based on a recently-discovered motion silencing phenomenon. The proposed model successfully predicts flicker visibility on both static and moving regions. The model not only provides a pixel-wise flicker visibility index, but also predicts silenced flicker in the presence of motion. The results show that flicker visibility predicted by the proposed model correlates well with human percepts of flicker distortions, and its performance is highly competitive with or outperforms current methods.

We believe that the proposed flicker visibility model will be useful for augmenting automatic objective VQA methods by predicting suppressed flicker distortions. We have explored the tight coupling of motion and flicker visibility via a detailed model of spectral signatures. Understanding changes in the spectral signatures arising from multiple distortions would also be helpful when predicting distortion specific or generalized spatiotemporal distortion visibility. Future work could combine this model into modern VQA models. Also of interest are time-varying models of flicker density, which will require databases of time-varying data, similar to [7].

References

[1] Cisco Corporation, Cisco Visual Networking index: Global mobile data traffic forecast update, 2014-2019.

[2] Interactive Advertising Bureau, IAB Mobile Video 2015: A global perspective.

[3] L. K. Choi, Y. Liao, and A. C. Bovik, "Video QoE metrics for the compute continuum," IEEE Commun. Soc. Multimed. Tech. Comm. (MMTC) E-Lett., vol. 8, no. 5, pp. 26-29, 2013.

[4] K. Zeng, T. Zhao, A. Rehman, and Z. Wang, "Characterizing perceptual artifacts in compressed video streams," in Proc. SPIE HVEI, 2014.

[5] A. C. Bovik, "Automatic prediction of perceptual image and video quality," Proc. IEEE, vol. 101, no. 9, pp. 2008-2024, 2013.
[6] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," IEEE Trans. Image Process., vol.19, no.6, pp.1427-1441, 2010.
[7] K. Moorthy, L. K. Choi, A. C. Bovik, and G. de Veciana, "Video quality assessment on mobile devices: Subjective, behavioral and objective studies," IEEE J. Sel. Topics Signal Process., vol. 6, no. 6, pp. 652-671, 2012.
[8] C. Chen, L. K. Choi, G. de Veciana, C. Caramanis, R. W. Heath, Jr., and A. C. Bovik, "Modeling the time-varying subjective quality of http video streams with rate adaptations," IEEE Trans. Image Process., vol. 23, no. 5, pp. 2206-2221, 2014.
[9] M. Carandini, J. B. Demb, V. Mante, D. J. Tolhurst, Y. Dan, B. A. Olshausen, J. L. Gallant, and N. C. Rust, "Do we know what the early visual system does?" J. Neurosci., vol. 25, no. 46, pp. 10577-10597, 2005.
[10] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," Annu. Rev. Neurosci., vol. 24, pp. 1193-1216, 2001.
[11] B. Breitmeyer and H. Ogmen, Visual Masking: Time Slices Through Conscious and Unconscious Vision, New York, NY, USA: Oxford University Press, 2006.
[12] 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, 2004.
[13] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," IEEE Trans. Broadcast., vol. 10, no. 3, pp. 312-322, 2004.
[14] 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, 2010.
[15] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatiotemporal entropic differencing," IEEE Trans. Circuits Syst. Video Technol., vol. 23, no. 4, pp. 684-694, 2013.
[16] M. A. Saad and A. C. Bovik, "Blind prediction of natural video quality," IEEE Trans. Image Process., vol. 23, no. 3, pp. 1352-1365, 2014.
[17] 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, 1965.
[18] B. Girod, "The information theoretical significance of spatial and temporal masking in video signals," in Proc. SPIE Human Vis. Visual Process. and Digital Display, pp.178-187, 1989.
[19] A. Puri and R. Aravind, "Motion-compensated video with adaptive perceptual quantization," IEEE Trans. Circuits Syst. Video Technol., vol. 1, pp. 351-378, 1991.
[20] B. G. Haskell, F. W. Mounts, and J. C. Candy, "Interframe coding of videotelephone pictures," Proc. IEEE, vol. 60, pp. 792-800, 1972.
[21] M. Yuen and H. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," Signal Process., vol. 70, no. 3, pp. 247-278, 1998.
[22] X. Fan, W. Gao, Y. Lu, D. Zhao, Flickering reduction in all intra frame coding, JVT-E070, in Proc. the Joint Video Team of ISO/IEC MPEG & ITU-T VCEG Meeting, 2002.
[23] J. Yang, J. B. Park, and B. Jeon, "Flickering effect reduction for H.264/AVC intra frames," in Proc. SPIE vol. 6391, 2006.
[24] E. Gelasca and T. Ebrahimi, "On evaluating video object segmentation quality: A perceptually driven objective metric," IEEE J. Sel. Topics Signal Process., vol. 3, no. 2. pp. 319-335. 2009.
[25] D. T. Vo, T. Q. Nguyen, S. Yea, and A. Vetro, "Adaptive fuzzy filtering for artifact reduction in compressed images and videos," IEEE Trans. Image Process., vol. 18, no. 6, pp. 1166-1178, 2009.
[26] J. W. Suchow and G. A. Alvarez, "Motion silences awareness of visual change," Curr. Biol., vol. 21, no. 2, pp.140-143, 2011.
[27] 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, 2012.
[28] M. Turi and D. Burr, "The motion silencing illusion results from global motion and crowding," J. Vis., vol. 13, no. 5, 2013.
[29] L. K. Choi, A. C. Bovik, and L. K. Cormack, "Spatiotemporal flicker detector model of motion silencing," Perception, vol. 43, no. 12, pp. 1286-1302, 2014.
[30] L. K. Choi, L. K. Cormack, and A. C. Bovik, "On the visibility of flicker distortions in naturalistic videos," in Proc. IEEE QoMEX, 2013.
[31] L. K. Choi, L. K. Cormack, and A. C. Bovik, "Motion silencing of flicker distortions on naturalistic videos," Signal Process. Image Commun., vol. 39, pp. 328-341, 2015.
[32] L. K. Choi, L. K. Cormack, and A. C. Bovik, "LIVE Flicker Video Database," Online: http://live.ece.utexas.edu/research/quality/live_flicker_video.html, 2015.
[33] L. K. Choi, L. K. Cormack, and A. C. Bovik, "Eccentricity effect of motion silencing on naturalistic videos," in Proc. IEEE GlobalSIP, 2015.
[34] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two dimensional visual cortical filters," J. Opt. Soc. Amer. A, vol. 2, no. 7, pp. 1160-1169, 1985.
[35] A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," IEEE Trans. Pattern Anal. Mach. Intell., vol. 12, no. 1, pp. 55-73, 1990.
[36] E. H. Adelson and J. R. Bergen, "Spatiotemporal energy models for the perception of motion," J. Opt. Soc. Amer. A, vol. 2, no. 2, pp. 284-299, 1985.
[37] D. J. Heeger, "Normalization of cell responses in cat striate cortex," Vis. Neurosci., vol. 9, no. 2, pp. 181-197, 1992.
[38] A. B. Watson and A. J. Ahumada, "Model of human visual-motion sensing," J. Opt. So. Amer. A, vol. 2, no. 2, pp. 322-342, 1985.
[39] E. P. Simoncelli and D. J. Heeger, "A model of neuronal responses in visual area MT," Vis. Res., vol. 38, no. 5, pp. 743-761, 1998.
[40] D. Fleet and A. Jepson, "Computation of component image velocity from local phase information," Int. J. Comput. Vis., vol. 5, no. 1, pp. 77-104, 1990.
[41] 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, 2006.

Author Biography

Lark Kwon Choi received his BS in electrical engineering at Korea University (2002), the MS in electrical engineering and computer science at Seoul National University (2004), and his PhD in electrical and computer engineering at The University of Texas at Austin (2015) under the supervision of Dr. Alan. C. Bovik. His research interests include image and video quality assessment, spatial and temporal visual masking, motion perception, and perceptual image and video quality enhancement.

Alan C. Bovik is the Cockrell Family Regents Endowed Chair Professor at The University of Texas at Austin, where is Director of the Laboratory for Image and Video Engineering (LIVE). He received his PhD from the University of Illinois, Urbana. He is an Honorary Member of IS&T. He received a Primetime Emmy Award in 2015 for his work on the development of video quality prediction models which have become standard tools in broadcast and post-production houses throughout the television industry.