Signal Processing: *Image Communication* 39 (2015) 328–341

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Motion silencing of flicker distortions on naturalistic videos

Lark Kwon Choi [a,*], Lawrence K. Cormack [b], Alan C. Bovik [a]

[a] *Laboratory for Image and Video Engineering (LIVE), Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78701, USA*

[b] *Center for Perceptual Systems (CPS), Department of Psychology, The University of Texas at Austin, Austin, TX 78712, USA*

ARTICLE INFO

Available online 27 March 2015

Keywords:
Motion silencing
Visual masking
Flicker distortion
Visibility of distortion
Video quality

ABSTRACT

We study the influence of motion on the visibility of flicker distortions in naturalistic videos. A series of human subjective studies were executed to understand how motion silences the visibility of flicker distortions as a function of object motion, flicker frequency, and video quality. We found that flicker visibility is strongly reduced when the speed of coherent motion is large, and the effect is pronounced when video quality is poor. Based on this finding, we propose a model of flicker visibility on naturalistic videos. The target-related activation levels in the excitatory layer of neurons were estimated for a displayed video using a spatiotemporal backward masking model, and then the flicker visibility is predicted based on a learned model of neural flicker adaptation processes. Experimental results show that the prediction of flicker visibility using the proposed model correlates well with human perception of flicker distortions. We believe that sufficiently fast and coherent motion silences the perception of flicker distortions on naturalistic videos in agreement with a recently observed "motion silencing" effect on synthetic stimuli. We envision that the proposed model could be applied to develop perceptual video quality assessment algorithms that can predict "silenced" temporal distortions and account for them when computing quality judgments.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The global volume of digital videos is exponentially gro wing and taking up an increasingly dominant percentage of Internet traffic. The world's mobile video traffic was 53 percent of Internet traffic in 2013 and is expected to increase to more than two-thirds by 2018 [1]. Video streaming services such as Netflix, Hulu, and YouTube proliferate and live online video services including Skype and Google+ Hangouts are expanding rapidly as well [2]. As consumer demand increases, fast, reliable, and accurate methods for evaluating video quality are also important for satisfying users' Quality of Experience [3]. Another driving force behind the proliferation of user demand is the advancement of technologies to produce better quality, higher resolution videos such as High Definition (HD) and emerging 4K videos.

Generally, digital videos suffer not only from spatial artifacts such as blocking, blurring, ringing, mosaic patterns, and noise, but are also degraded by temporal distortions including motion compensation mismatches, flicker, mosquito effects, ghosting, jerkiness, smearing, and so forth [4]. Since humans are the ultimate recipients of videos, technologies to monitor the quality of the received videos have been focused on the design of objective video quality assessment (VQA) methods. 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. Interestingly, the mere presence of spatial or temporal distortions does not have to imply quality degradation

* Corresponding author. Tel.: +1 512 917 9026.
E-mail addresses: larkkwonchoi@utexas.edu (L.K. Choi), cormack@utexas.edu (L.K. Cormack), bovik@ece.utexas.edu (A.C. Bovik).

http://dx.doi.org/10.1016/j.image.2015.03.006
0923-5965/© 2015 Elsevier B.V. All rights reserved.

since the visibility of distortions can be strongly reduced or completely eliminated by visual masking [2].

Visual masking occurs when a stimulus, called the mask, is superimposed on another stimulus, called the target. The mask is typically of similar orientation, spatiotemporal frequency, motion, or color to the target, causing the target to be less visible [5]. 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 ΔI is more likely visible in a dark region than a bright one [6]. 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 [7]. Spatial visual masking is well-known and plays a central role in the design of perceptual image quality models [8,9], video compression [10], and watermarking [11].

Regarding temporal visual masking, psychophysical experiments of temporal masking have been performed using light flashes [12], sine wave-gratings [13], and vernier stimuli [14]. 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 [15]. Later, temporal masking of distortions was studied in the context of video compression. Netravali et al. [10] investigated the effects of luminance transitions on the perception of quantization noise. Girod [16] emphasized the theoretical significance of spatial and temporal masking effects, while Johnston et al. [17] created a non-linear quantizer using simple models of spatial and temporal masking. Puri et al. [18] designed an adaptive video coder using the predicted visibility of noise on flat areas, textures, and edges. Haskell et al. [19] suggested that observers are more tolerant of distortions in moving images than in stationary images due to a presumed difficulty of focusing on details on moving objects. More recently, the just-noticeable distortion (JND) has been applied to adaptive image coding [20] and on the visibility of noise in videos [21]. While a variety of ideas and algorithms have been proposed to account for temporal masking, scene changes are sparse, and implementations to account for temporal masking in video compression have been largely heuristic based on anecdotal evidence. Although JND is related to the visibility of distortions, the influence of object motion on the visibility of distortions has not been explicitly analyzed in this context.

Flicker distortions, which are not well accounted for in current VQA models, frequently appear near moving edges and textures in interlaced videos and/or compressed videos, producing annoying visual artifacts such as edge flicker, interline flicker, and line crawling [22]. To mitigate flicker defects, variations on the Kell factor [23], which perceptually limits the bandwidth of a temporally sampled image to avoid the visibility of irritating beat frequencies, have been devised, while a variety of deinterlacing techniques [24] were proposed to reduce flicker including temporal anti-alias filters, interpolation, and motion-compensated up-conversion. Flicker also can be observed in quantized or compressed videos due to spatially localized temporal fluctuations of luminance: for example, caused by blockiness over textures or blockiness at the junction of blocks, called corner outliers. Fan et al. [25] proposed a modified encoder with improved intra-prediction of encoded sequences, while others devised post-processing filters [26,27] to reduce flicker; however, in these models masking effects were not accounted for. Recently, Ni et al. [28] performed a subjective study to understand the visibility of flicker as a function of flicker amplitude and frequency, but not how it is affected by motion.

Very recently, Suchow and Alvarez [29] demonstrated a striking "motion silencing" illusion (the illusion may be viewed at http://visionlab.harvard.edu/silencing/), in which the salient temporal changes of objects in luminance, color, size, and shape appear to stop when they move rapidly in collective motion. Many types of illusions exist [30], including some which are perceived phenomena that do not actually exist, and others where a phenomena that might ordinarily be seen are not because of modified conditions. Motion silencing illusion is an example of the latter, i.e., suppressed or eliminated perception of an existing flicker signal due to motion. This motion-induced failure to detect change not only suggests the tight coupling of motion and object appearance, but also reveals that motion can dramatically reduce the perception of salient changes in visual objects. Hence, understanding the motion silencing phenomenon as a form of temporal visual masking is important forwards deepening our understanding of the visibility of commonly-occurring temporal flicker distortions. Our previous studies [31,32] showed that motion silencing can be explained as a function of stimulus velocity, flicker frequency, and spacing between objects by using a spatiotemporal model of cortical simple cell responses. A consistent physiological and computational model that detects motion silencing might be useful to probe related motion perception effects, such as distortion visibility in compressed videos. However, since the effect has thus far only been studied on highly synthetic stimuli such as moving dots, there is a significant need to understand the impact of object motions on the visibility of temporal distortions in real-world, naturalistic videos. Furthermore, developing a visibility prediction model of temporal distortions is of great interest as a possible path towards improving the performance of VQA models.

Here, we study a new observed motion silencing phenomenon wherein flicker visibility is affected by the speed of coherent object motion on naturalistic videos. We conducted a series of human subjective studies to examine the influence of object motion on flicker visibility. By presenting compressed videos undergoing regular, periodic changes in quality levels to 43 naïve subjects engaged in two tasks ("follow the moving object" and "view freely"), we found the empirical distributions of flicker visibility on tested videos and the effect of motion on flicker visibility using a correlation analysis. Based on this finding, we propose a model of flicker visibility on naturalistic videos. The proposed model estimates target-related activation levels in the excitatory layer of neurons for displayed video frames (the target) against immediately following frames (the mask) via a spatiotemporal backward masking model [14,33], then applies flicker adaptation [34] and accumulation [35,36] processes to predict flicker visibility. The predicted results correlate well with human perception of the visibility of flicker distortions. This paper builds upon our earlier work describing on the visibility of flicker distortions [37] and prediction of flicker visibility on naturalistic videos [38].

The remainder of this paper is organized as follows. In Section 2, we describe details of a human subjective study

investigating motion silencing of flicker distortions on naturalistic videos and perform an analysis of the subjective opinion scores obtained from the study. Section 3 proposes an algorithm of flicker visibility and evaluates the performance of the algorithm on the executed human subjective study results. Finally, general conclusions and discussions are presented in Section 4.

2. Human subjective study of flicker distortions

2.1. Source videos

Source videos were obtained using a RED ONE digital cinematographic camera. The 12-bit REDCODE (.r3d) RAW data was captured from the MYSTERIUM sensor at a resolution of 3K (3072×1728) at the frame rate of 30 fps using the 42 MB/s option to ensure that the best possible acquisition quality was recorded. The RED 50–150 mm and 18–50 mm T3 zoom lens were used. The source videos were truncated to 10 s lengths and then downsampled to resolution 720p (1280×720) or cropped to 1920×720 at 30 fps. The .r3d videos were converted into uncompressed .yuv files by a combination of the "imresize" (option: bicubic) function in MATLAB and Adobe Premiere CS5. One video, "Tr," was added from a database available from the Technical University of Munich and followed the same processing. A total of eight reference videos were selected. Two videos were used for a training session, while six videos were used for the actual test. The list below describes each of the videos used in the human subjective study.

1) Baseball Batter (Bb): Shot at Disch–Falk Field, Austin, Texas on a sunny morning. A baseball player stands, then runs fast after hitting a ball. The camera was fixed.
2) BMX bike rider (BMX): Shot at the Austin BMX and skate park on a sunny afternoon. A BMX bike rider continuously moves in a roughly semi-circular arc like a pendulum. The object speed changes up and down fast. The camera was fixed.
3) Lacrosse Players (La): Shot at The University of Texas at Austin intramural fields on a sunny afternoon. Lacrosse players stand, then run towards a goal after a referee blows a whistle. This scene includes both static and rapid motions. The camera was fixed.
4) Metro Rail (Mr): Shot at a MetroRail station, Howard, Austin, Texas on a sunny morning. MetroRail starts leaving the station as increasing speed gradually. The camera was fixed.
5) Red Car (Rc): Shot in a road on a sunny afternoon. A red car moves from right to left twice. The car moves slowly at first but then faster, so this video includes abrupt object motion with a scene change. The camera was fixed.
6) Tractor (Tr): Shot at a farm field on a sunny day. A tractor turns left across the field. The object speed increases steadily. The camera follows the tractor and zooms in.

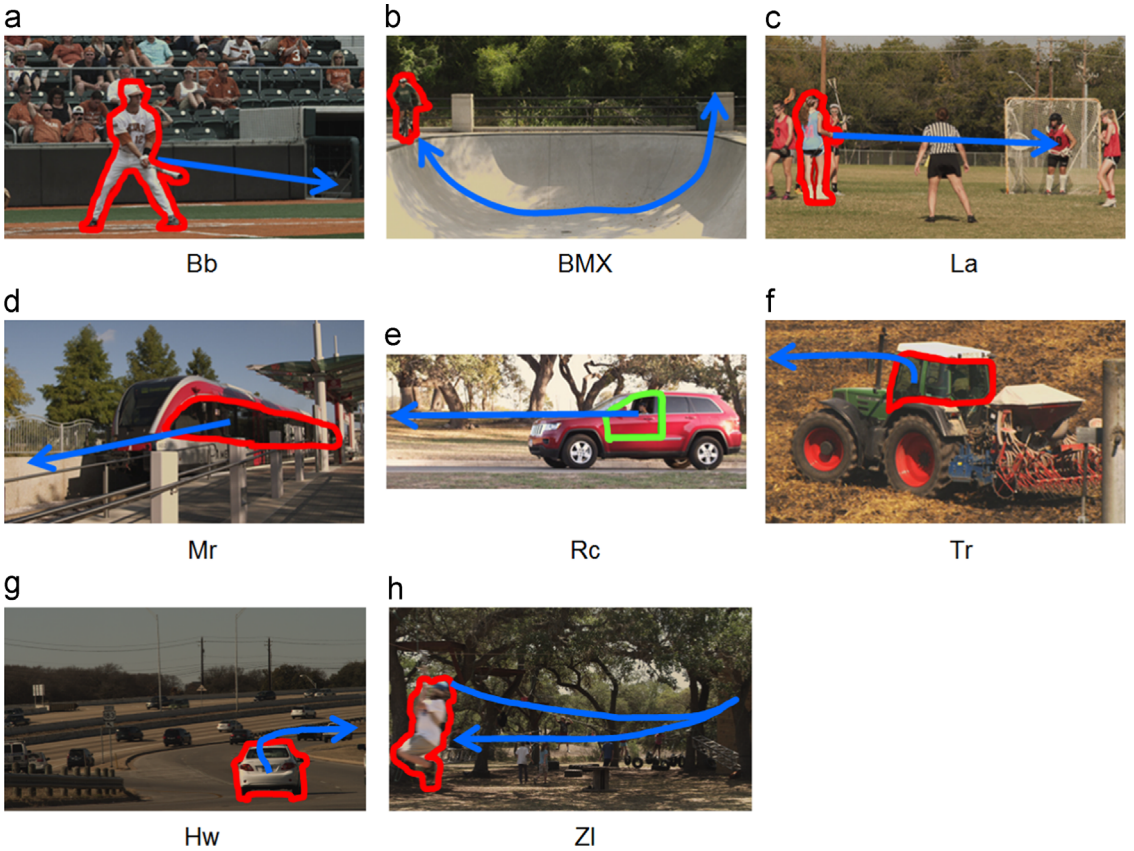

Fig. 1. Example frames of the videos used in the study. The marked areas indicate moving objects, while arrows denote the paths of object movement.

7) Highway (Hw): Shot at Highway 183, Austin, Texas on a sunny afternoon. Cars move at different speeds and directions. Some cars turn at a highway entrance. The camera was fixed.
8) Zip Line (Zl): Shot at a Stunt Ranch, Austin, Texas on a sunny morning. Children play on the ground, and a girl is running across the playground. The camera was fixed.

Fig. 1 shows sample frames from the various video sequences including diverse speeds of object motions. The marked areas show the gaze target and moving objects, while the blue arrows indicate the paths followed by the moving objects.

2.2. Distortion simulation

We simulated flicker distortions by periodically alternating subsequences of videos at different quality levels (amplitude) at three flicker frequencies (duration) as illustrated in Fig. 2. Flicker frequency means the number of quality level alternations per second. Specifically, the flicker is caused by changing frames from a high to a low quantization parameter (QP) setting and then back from low to high QP in an H.264 compression codec. Thus, the artifact is a "distortion flicker," rather than a brightness flicker. Each reference video was encoded using the JM reference implementation of H.264 Advanced Video Codec (AVC) [39] at one of four fixed QP values: QP26, QP32, QP38, and QP44. The perceptual video quality was found to be well separated using these QP values, corresponding to roughly excellent, good, poor, and bad, respectively, from QP26 to QP44. Over each flicker period, the QP alternated either from QP44 to QP26, from QP38 to QP26, or from QP32 to QP26 by inserting appropriate segments from the four already compressed videos. 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.

Further, to investigate the effect of flicker frequency on the visibility of flicker as a function of motion, videos with variable-length QP durations (periods) were also constructed, corresponding to time-varying flicker frequencies of 7.5, 5, and 3 Hz, all using a fixed alternating QP pair (QP38, QP26). Each case is illustrated in Fig. 2(a) and (b). A transition from low to high quality decides flicker magnitude. When the quality difference is large (e.g., QP44, QP26), the flicker magnitude is large, so flicker can be easily perceived. However, when the quality difference is small (e.g., QP32, QP26), the flicker magnitude is small, so flicker can be perceived less. We explore how flicker visibility at each quality difference is influenced by object motion speeds.

2.3. Test methodology

We used a single-stimulus continuous quality evaluation (SSCQE) [40] procedure with hidden reference to obtain subjective percepts of flicker visibility for the different video sequences. All subjects taking part in the study were recruited from The University of Texas at Austin. A total of 43 students served as naïve subjects. All subjects were between the ages of 20 and 35. Each was found to have normal or corrected-to-normal vision in visual acuity (Snellen Test) and color perception (Ishihara Test). We developed the user interface for the study using MATLAB and the XGL toolbox [41]. The XGL toolbox interfaced with an ATI Radeon X300E graphics card in an Intel Xeon® 2.93 Hz Dual CPU, 24 GB RAM Windows computer. All test sequences were processed and stored as raw YUV 4:2:0 files. To avoid any latency due to slow hard disk access of a large video file, each entire video was loaded into memory before its presentation. Subjects viewed videos on a 24″ screen Dell U2410 LCD monitor (Dell, Round Rock, TX, USA) with a resolution of 1920×1080 at a 60 Hz refresh rate. The entire study was conducted using the same monitor, and the video was played at the center of the display. The remaining areas of the display were black.

Before presenting a test video, an instruction frame was displayed to visualize the pre-defined moving object contained in each test video. The study interface is shown in Fig. 3. A continuously calibrated rating slider bar with Likert-like markings was shown at the right side of the screen. The rating scale ranged from 0 to 100, where the increments 0, 25, 50, 75, and 100 were marked as "Hardly," "Little," "Medium," "Highly," and "Extremely" to indicate the degree of perceived flicker visibility. The initial score displayed on the rating bar was "Medium" at the beginning of each video. During playback, the rating bar disappeared except for a white score gauge along the bar not to disturb video viewing.

An eye and head tracker (faceLAB 5, Seeing Machines) was used to monitor each subject's gaze direction. The subjects' heads were unrestrained. Gaze was calibrated by using a 9 point calibration sequence before each experiment, and was recorded at every 1/60 s into calibrated display coordinates. The viewing distance was approximately 87 cm (three times of display height).

Each subject participated in two separate tasks; Task 1, "follow the moving object" and Task 2, "view freely." Subjects performed Task 1 first and executed Task 2 after sufficient rest (e.g., one day). In Task 1, subjects were requested to fixate their eyes on the pre-defined moving object throughout the duration of the video and to rate the visibility of flicker on the object by moving the mouse up

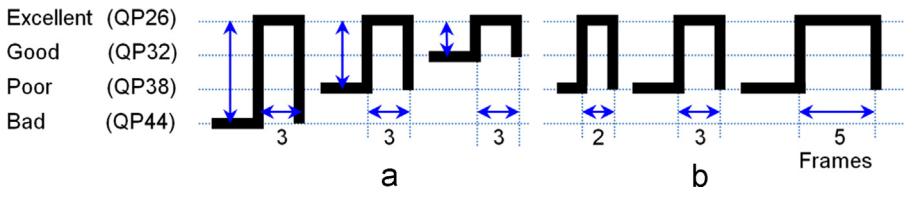

Fig. 2. Schematic illustration of the flicker distortion simulation: (a) video quality level (amplitude) changes and (b) video frame duration (frequency) changes. Every 2, 3, and 5 frames were periodically alternated, corresponding to flicker frequencies of 7.5, 5, and 3 Hz, respectively.

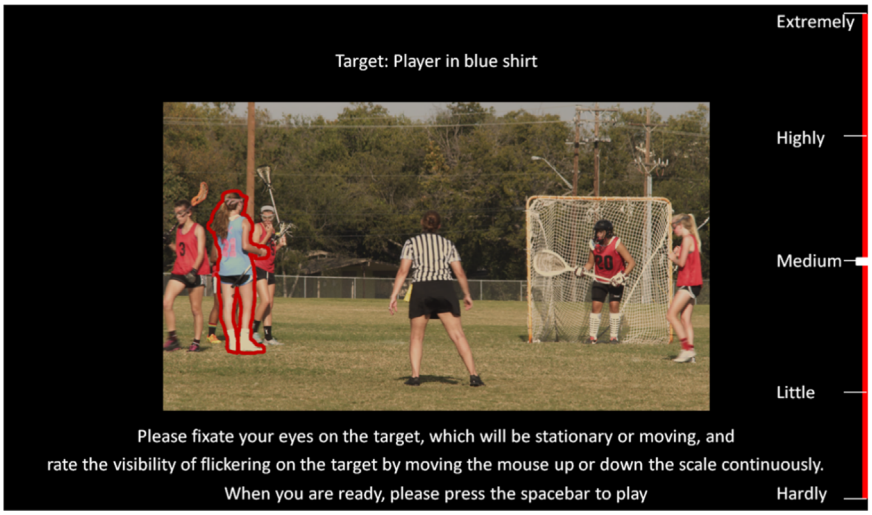

Fig. 3. The subjective study interface displaying the instruction and the video to the subject.

or down the scale continuously. In Task 2, subjects were requested to view videos naturally. Each task lasted less than 20 min, each of which consisted of 36 test videos (6 hidden reference videos and 30 flicker distorted videos) in randomized order. Prior to data collection, a short training session preceded the actual study to familiarize the subjects with the experimental procedure.

In addition, a lag response (the time difference between the perception of flicker and the movement of a mouse to rate the visibility) was measured for each session and for each subject. Subjects were asked to move a mouse up when a black dot on a white background flickered and to move a mouse down when the dot did not flicker. The black dot distinctively flickered for 2 or 3 s. Time duration from the start of the dot flicker (or no-flicker) on the screen to the mouse movement by the subject was measured five times, and then those values were averaged.

2.4. Processing of the scores

The subject screening procedure was executed in Task 1 by monitoring whether the subject followed the pre-defined, designated moving objects based on the recorded gaze positions. Since no subjects were rejected from the study, the scores from all subjects were used for the analysis.

To ensure appropriate time synchronization between the frame at which the subject visually perceived the flicker and the mechanical scoring of flicker visibility that was rated by hand, the evaluated score was matched after shifting the score signals by the average lag response time for each subject. Let s_{ijf} denote the score assigned by subject i to video j at frame number f, ms_{ijf} be the visibility score of subject i to video j at frame number f, and lag_i be the average lag response of subject i. Then the visibility score is given by

$$ms_{ijf} = s_{ij(f+lag\_i)}. \tag{1}$$

The range of lag response times was 0.47–1.17 s (14–35 frames).

In order to unbias measured flicker visibility of each flicker distorted video from its content, we calculated the difference flicker visibility scores between the score that the subject assigned to each reference video and the scores assigned to the corresponding flicker distorted videos. Specifically, let ms_{ij_reff} denote the flicker visibility score assigned by subject i to the reference no-flicker video associated with the distorted video j after lag response matching and, M_j be the total number of ratings received for video j. The difference scores ds_{ijf} are computed as

$$ds_{ijf} = ms_{ijf} - ms_{ij\_reff}. \tag{2}$$

The final flicker visibility score is

$$fvs_{jf} = \frac{1}{M_j}\sum_i ds_{ijf}. \tag{3}$$

The flicker visibility scores range continuously from 0 to 100, where 0 means that the subject failed to, or hardly perceived flicker on the moving object, while 100 means that the subject perceived flicker clearly.

2.5. Analysis of subjective flicker visibility

2.5.1. Distributions of flicker visibility against object speed

The measured flicker visibility for each video was analyzed against the speed of the pre-defined moving object. Distribution of flicker visibility against object speed is shown in Fig. 4 for the test videos in Task 1. Each column represents different content from “Bb” to “Tr.” In each subplot in Fig. 4 (a), solid red, green, and blue lines indicate the average distribution of flicker visibility for different QP alternation pairs (QP44, QP26), (QP38, QP26), and (QP32, QP26), respectively, at 5 Hz flicker frequency. The left vertical axis is the flicker visibility score, and the horizontal axis is the frame number. The dotted black line shows the average speed of the moving object, which was computed on successive frames using the optical flow algorithm in [42]. The magnitude of flow on the area of the pre-defined moving object was then averaged. The magnitude of motion flow is displayed on the right vertical axis (unit: pixel/frame). The area of the moving object associated with each video frame was segmented manually, then used to calculate object

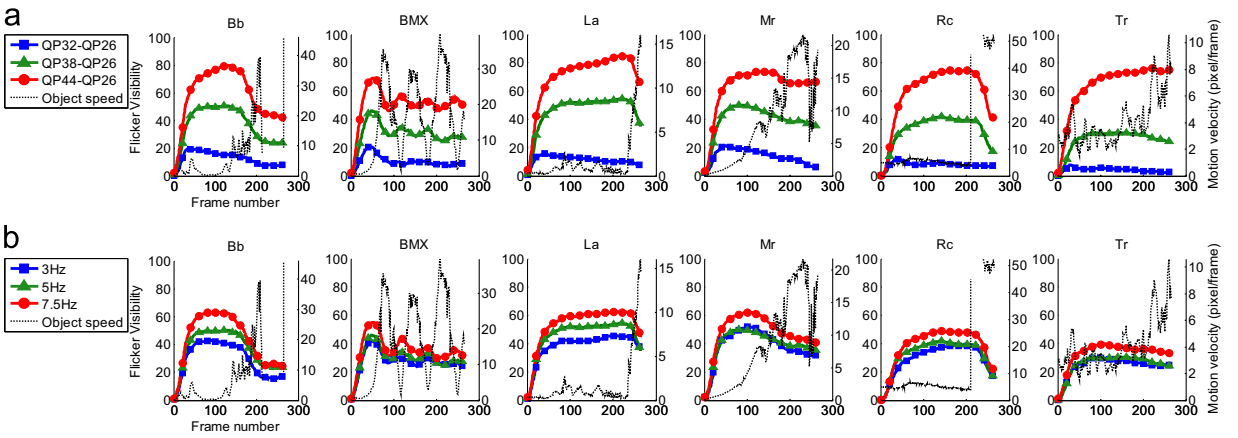

Fig. 4. Distributions of flicker visibility against object speed: (a) quality level changes at 5 Hz flicker frequency and (b) flicker frequencies at QP alternations (QP38, QP26). Solid lines indicate flicker visibility, while dotted lines show motion velocity (pixel/frame). It can be observed that when motion is large, the visibility of flicker distortion decreases, and this effect is more pronounced when the video quality is poor.

motion as well as to monitor subjects' gaze whether they followed the pre-defined object or not. We superposed the object speed on flicker visibility scores for easy comparison in Fig. 4.

Each subject spent at least 1.67 s (50 frames, which include a judgment time and a lag response) up to 3 s in order to rate the first flicker visibility after a test video began, which explains why the flicker visibility scores begin at 0 in Fig. 4. Since flicker visibility scores were shifted to compensate for the manual response latency, the last 35 frames of visibility scores for each video were not displayed. When moving objects disappeared from a scene, no data for object speed existed, so we omitted those intervals. Frame intervals [51–208], [51–265], [51–265], [51–265], [51–209, 240–265], and [51–265] were used for the analysis from "Bb" to "Tr."

The experimental results show that the visibility of flicker distortions was strongly reduced when object motion was large or increased rapidly for all test video sequences. It happened even when video quality was maintained at a given QP level, or when the apparent quality worsened due to blur as the speed of motion increased. The reduction of flicker visibility on naturalistic videos was dependent on both overall video quality (range of QP values) and on the object speed. When video quality was high, flicker visibility was low and less sensitive to motion, whereas when video quality was poor, the motion silencing effect was large. In addition, an abrupt increase of object speed (e.g., "Rc") led to a significant reduction of flicker visibility. On the other hand, slow and gradual movement (e.g., "Tr") influenced flicker visibility very little. Although the subjects were able to hold their gaze on the moving objects, less flicker was seen on fast-moving objects.

The distribution of flicker visibility at different flicker frequencies for a fixed QP alternation pair (QP38, QP26) is shown in Fig. 4(b) against object speed. The solid red, green, and blue lines show 7.5, 5, and 3 Hz flicker frequency, respectively. The dotted black line indicates the object speed. Similar to the result on quality level changes, the visibility of flicker distortion decreased when object motion was large or increased abruptly, while flicker visibility gradually reduced when motion was small or increased steadily. Although subjects generally perceived more flicker at high flicker frequencies, the influence of flicker frequency was not as large as was the effect of increased object motions.

2.5.2. Correlation analysis

We analyzed the relationship between flicker visibility and object speed using Pearson's linear correlation coefficient (LCC) for the QP level changes and flicker frequencies. Fig. 5 represents the result. Each column indicates different content. In each subplot, the vertical axis is flicker visibility, and the horizontal axis is the object speed. To indicate object motion, frame numbers are rendered using a standard color map from cool to hot. For example, for "BMX," frame 51 is blue, and frame 265 is red. LCC between flicker visibility and object motion is displayed above each subplot.

The magnitude of LCC on "Bb" is greater than 0.83, which implies that object motion has a significant influence on flicker visibility. Despite distinct negative correlation in each sweep of the rider on "BMX" (e.g., QP alternation (QP38, QP26)), the single correlation coefficient does not capture the interaction between time (motion trajectory) and flicker visibility. When we segmented each sweep with a peak-to-peak speed of motion, the frame intervals were [51–76], [77–150], [151–209], and [210–265], and LCC were −0.9727, −0.6588, −0.9062, and −0.9305, respectively. Since the object speed changed rapidly and repeatedly over a short time span, the subjects could not rate scores using a mouse as quickly as they perceived flicker. Results on "La" and "Mr" clearly show that flicker visibility decreased as object speeds increased, albeit the single LCC score does not represent silencing. For "Rc" containing the QP alternation (QP32, QP26), the LCC was −0.53 and the range of flicker visibility is small, while for other cases, the LCC magnitude exceeded 0.8 over a wide range of flicker visibility. This suggests that the reduction of flicker visibility is less significant when the overall video quality is good. One interesting observation is the positive correlation coefficient on "Tr" at the QP alternation (QP44, QP26). We speculate that when the object speed is slow and changes gradually in a poor quality video, the motion influence on flicker visibility is weaker. Camera motion

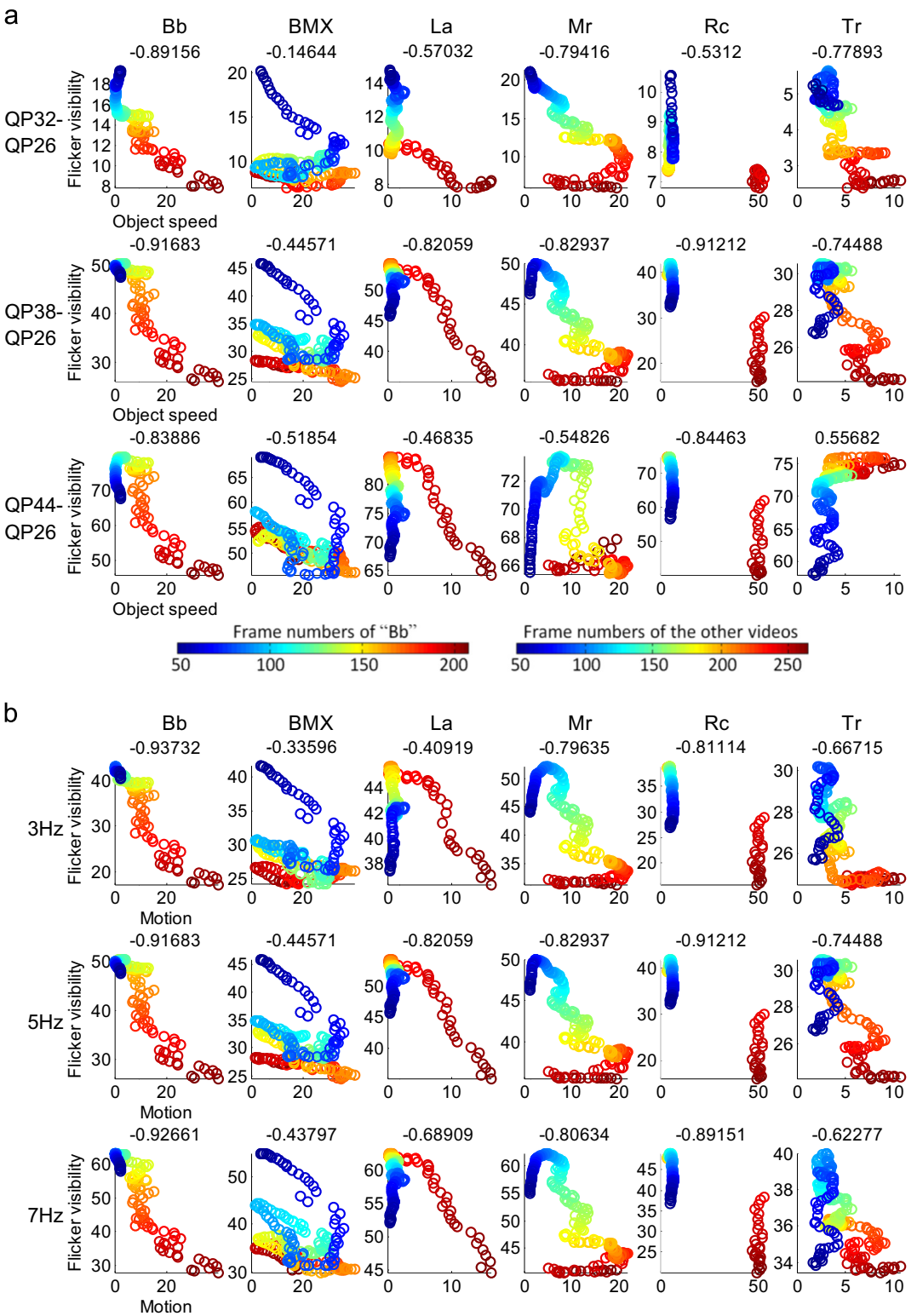

Fig. 5. Correlation analysis between the measured flicker visibility scores and object speeds (pixel/frame): (a) QP alternations and (b) flicker frequencies. Time runs from cool colors (beginning) to hot colors (end). The correlation coefficient is displayed above each subplot.

and zoom may modify the projected object speed unpredictably. Regarding the results of flicker frequencies, the impact of motion on flicker visibility was generally similar to the results of QP level changes as can be seen in Fig. 5(b).

2.5.3. Gaze analysis

All subjects followed the pre-defined moving objects during Task 1. Fig. 6 illustrates the average of accumulated gazes from all subjects for each video. The gaze traces shown in Fig. 6(a) indicate that subjects correctly followed the moving object as instructed. The gaze traces in Task 2 shown in Fig. 6(b) cover a wider range than those from Task 1. In Task 2, subjects usually started to watch videos at the center and followed the main moving object most of the time. When objects started to move, or when a new object appeared, or when a large distortion occurred, the gazes were drawn to these events. On high quality video sequences, subjects tended to look primarily at the moving objects, and the motion impact was similar to the result of Task 1.

For the analysis of the results in Task 2, we first found the gaze positions on each frame, then computed the magnitude of the optical flow corresponding to the gaze, and then associated the speed of motion with the flicker visibility. Because gaze movements occurred much more quickly than the rating of flicker visibility using a mouse, it was difficult to analyze the impact of motion using the same method as in Task 1. However, we still noticed that the overall flicker visibility was more greatly reduced when subjects followed fast moving objects than slow motion.

3. Prediction of flicker visibility

3.1. Algorithm

Inspired by the preceding, we developed a method of predicting target-related activation levels in an excitatory layer of artificial neurons on displayed video frames (the target) against immediately following frames (the mask) via spatiotemporal backward masking. Next, the target-related activation level is adjusted based on the flicker intensity (magnitude). Finally, flicker accumulation and adaptation processes are applied to predict flicker visibility on naturalistic videos.

3.1.1. Target-related activation levels of neurons

Neurons at the retina collectively receive inputs from the photoreceptors such as rods and cones, then produce center-surround, excitatory-inhibitory, on–off responses to local cone or rod cell signals and their surrounding neighbors, yielding a reduced-entropy residual signal [6]. Each receptive field (RF) describing a neuron's response may be well modeled as having an excitatory layer and an inhibitory layer. An excitatory layer increases firing rates, while an inhibitory layer suppresses the spontaneous firing of the cells. The two layers interact in an antagonistic way, and thereby the overlap of RFs controls firing rates of neurons to increase spatial resolution. These excitatory-inhibitory interactions can be modeled using neural field equations [33].

We used Hermens et al.'s backward masking model [14] to predict the firing rate of neurons, or target-related activation levels, for the visibility of flicker distortions on naturalistic videos. Since backward masking is the most dominant temporal masking in naturalistic videos as suggested by interruption theory [43] and scene change experiments [44], we adopted Hermens et al.'s model. Broadly, the target-related activation level is monotonically related to the percentage of correct responses to neural stimuli. In our context, it represents the visibility of a target frame (e.g., a current frame masked by following frames). The target-related activation levels are computed using interactions between and within excitatory and inhibitory layers of neurons based on Wilson and Cowan's neural network [14,33].

First, we assigned each current frame as a target frame and each following two frames as masking frames, as illustrated in Fig. 7. Since we were interested in the impact of motion on the visibility of flicker distortions, we selected the moving object (marked regions in Fig. 1) as a region of interest (ROI) and ignored or zeroed other regions. A set of partial differential equations describing the population activities A_e and A_i were used following [14]:

$$\tau_e \frac{\partial A_e(x,t)}{\partial t} = -A_e(x,t) + h_e\{w_{ee}(A_e * W_e)(x,t) + w_{ie}(A_i * W_i)(x,t) + I(x,t)\}. \quad (4-1)$$

$$\tau_i \frac{\partial A_i(x,t)}{\partial t} = -A_i(x,t) + h_i\{w_{ei}(A_e * W_e)(x,t) + w_{ii}(A_i * W_i)(x,t) + I(x,t)\}. \quad (4-2)$$

The parameters τ_e and τ_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 is time. The symbol "$*$" denotes convolution. The functions h_e and

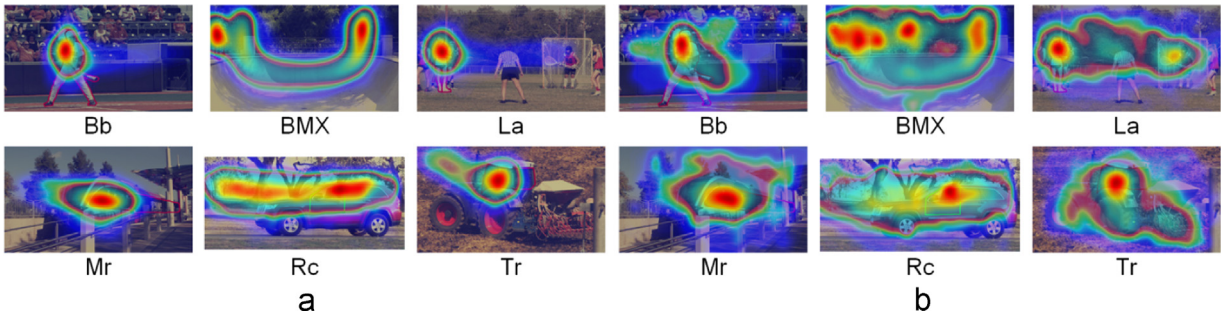

Fig. 6. The average of accumulated gazes from all subjects: (a) Task 1, "follow the moving object" and (b) Task 2, "view freely."

h_i are defined as

$$h_{e,i}(x) = \begin{cases} s_{e,i} \cdot x & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

with neuronal gain constants s_e and s_i. The coupling structures are

$$W_{e,i}(x-x') = \frac{1}{2\pi\sigma_{e,i}^2}\exp\left(-\frac{|x-x'|}{2\sigma_{e,i}^2}\right) \quad (6)$$

for the excitatory and inhibitory interactions, and $\sigma_{e,i}$ are the widths of the interaction kernels. The input to both populations is computed using

$$I(x,t) = (S*V)(x,t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(x',t)V(x-x')dx' \quad (7)$$

on a video frame S, while the input kernel is

$$V(x-x') = \frac{1}{2\pi\sigma_E^2}\exp\left(-\frac{|x-x'|^2}{2\sigma_E^2}\right) - \frac{1}{2\pi\sigma_I^2}\exp\left(-\frac{|x-x'|^2}{2\sigma_I^2}\right), \quad (8)$$

where σ_E and σ_I are the widths of the input kernel. We used the values of the different constants as described in [15]. The spatial kernels of these equations are arranged 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. Therefore, these equations describe a network of neurons whose connections describe the recurrent center-surround RF responses of each cell.

Finally, define the target activation level

$$T = \int_x A_e(x, r_0) \cdot S_T(x)dx, \quad (9)$$

where S_T is the ROI of the target frame, and r_0 is the duration of the excitatory activity (110 ms) after the target onset. We allowed the size of S_T to vary with the video content unlike Hermens et al.'s fixed constant target by dividing (normalizing) it by the size of the ROI.

The target activation levels are shown in Fig. 8 for the "Bb" and "BMX" test video sequences. One percent of the original T was rendered. Although T effectively captures the suppressed visibility of the target frame 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. Therefore, to quantify the visibility of flicker distortions, we defined a target and masks by controlling flicker intensity and object motions on naturalistic videos, respectively, with shifting and scaling factors, which were controlled based on quality level changes. 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.

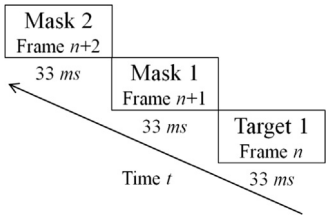

Fig. 7. The target and masks on naturalistic video stimuli.

3.1.2. Prediction of initial flicker visibility

Physiologically, in terms of visual masking, the target (flicker) in the test videos results from the intensity changes caused by periodic alterations of QP in the H.264 video compression (e.g., between QP44 and QP26), while the mask (object motion) comes from luminance changes created by natural object movements. Fig. 9 shows examples of the target and mask for different object speeds in the "Bb" video sequence. Fig. 9(a) shows frames containing different object motions. Fig. 9(b) shows the target and masks. The target is computed as the difference between the luminance frame of QP44 and QP26 at the same frame number, while the masks are computed as the frame differences of consecutive frames. In Fig. 9(b), intensities within the ROI are inversely rendered by magnifying the original intensities by a multiplicative factor of 10 to enable visual comparisons. While the target intensity is similar at different object motions, the intensities of the masks are quite different. The inset boxes in Fig. 9(c) and (d) show the object energies of frame differences when the object motion is small. QP changes between (QP32, QP26) and (QP44, QP26) caused different flicker intensities.

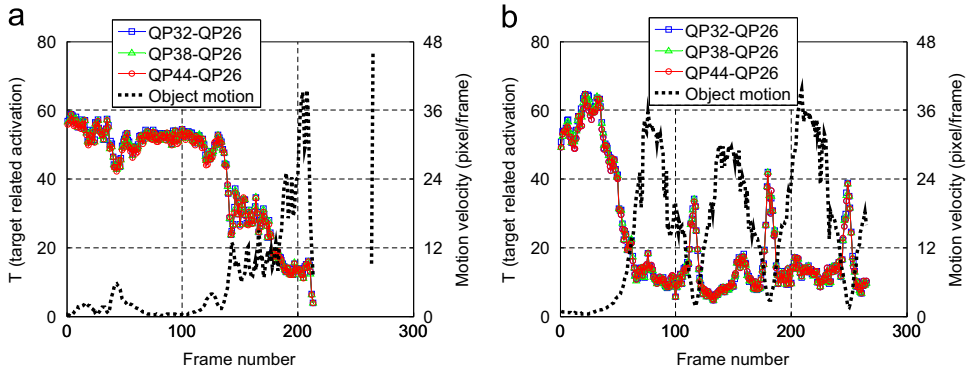

Fig. 8. The target related activation levels: (a) Bb and (b) BMX in spatiotemporal backward masking. Motion velocity is overlapped.

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 [45], then applied the logarithm of the quality level changes to compute a perceptual flicker quality factor [46]. Let $msssim_{jq}$ denote the MS-SSIM index result on video j at QP q (=26, 32, 38, and 44), $msssim_{ijq}$ be the MS-SSIM index result on video j at frame number i and QP q, and N be the total number of frames. The flicker quality factor, fi_{jp}, of video j for a QP change between QP26 and QP p (=32, 38 and 44) is then given by

$$msssim_{jq} = \frac{1}{N}\sum_{i=1}^{N} mssim_{ijq}, \qquad (10-1)$$

$$fi_{jp} = \log(msssim_{j26} - msssim_{jp}). \qquad (10-2)$$

Since video quality remains dependent 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 [47]. Each time we picked one video from the six videos in Fig. 1, trained the necessary parameters on the other five videos, then predicted the initial flicker visibility as follows:

$$z_{fi\_jp} = \frac{fi_{jp} - \mu(fi_j)}{\sigma(fi_j)}, \qquad (11-1)$$

$$z_{hfv\_jp} = \frac{hfv_{jp} - \mu(hfv_j)}{\sigma(hfv_j)}, \qquad (11-2)$$

$$\hat{z}_{hfv\_jp} = \beta_2 + \frac{\beta_1 - \beta_2}{1 + e^{-(z_{fi\_jp} - \beta_3)/|\beta_4|}}, \qquad (11-3)$$

$$\hat{\mu}(hfv_j) = \alpha_2 + \alpha_1 \times \hat{z}_{hfv\_jp}, \qquad (11-4)$$

$$\hat{\sigma}(hfv_j) = \frac{1}{M}\sigma(hfv_{\{j\}^c}), \quad j = 1, 2, \ldots 6, \qquad (11-5)$$

$$\widehat{hfv}_{jp} = \hat{z}_{hfv\_jp} \times \hat{\sigma}(hfv_j) + \hat{\mu}(hfv_j) \qquad (11-6)$$

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 "\widehat" means predicted quantities. We predicted the parameters β and α following [48] using a least squares fit of $\mu(hfv)$ at $\{j\}^c$, where $j = 1, 2, \ldots, 6$, respectively. Eq. (11-3) is a nonlinear regression on the z_{fi_jp} using a logistic function [48]. This well-accepted regression function works well although we also tried other nonlinear regression models which yielded similar results. Here $\hat{\sigma}(hfv_j)$ is the mean of $\sigma(hfv)$ at $\{j\}^c$. We used $\widehat{hfv}_{jp}$ as the initial flicker visibility at

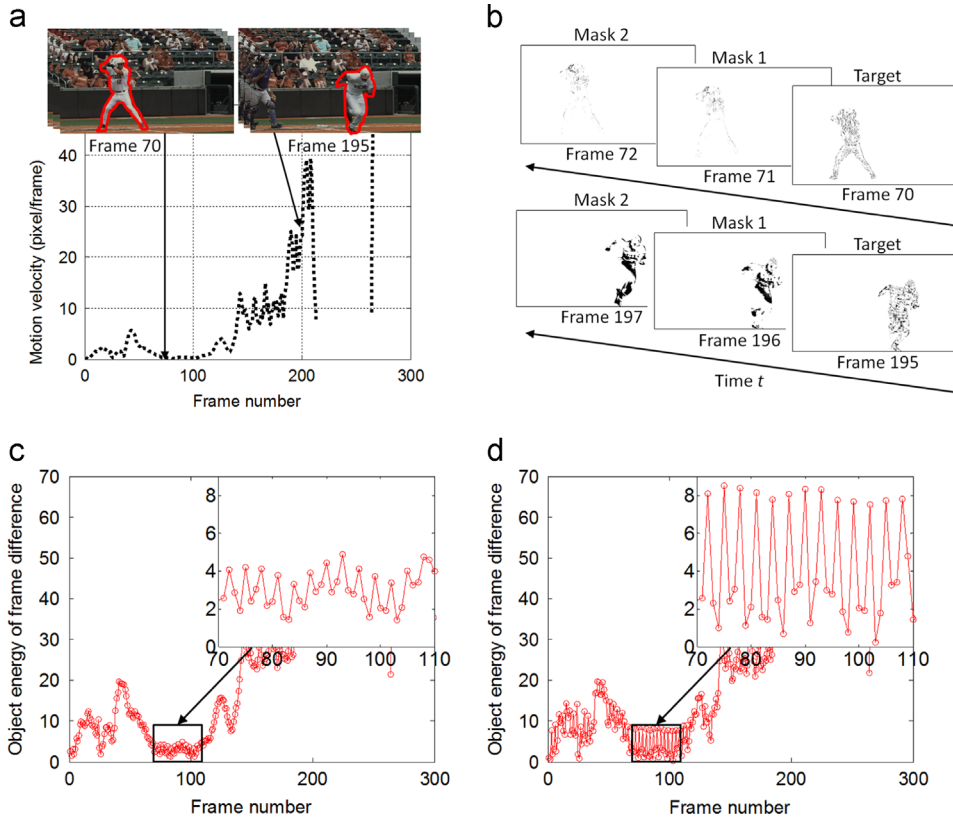

Fig. 9. The target (flicker) and the masks (object motion) in terms of visual masking: (a) frames at different object motionsand (b) magnitude of target and masks. Object energy of frame differences: (c) between QP32 and QP26 and (d) between QP44 and QP26 at the same object motion.

video j and QP p. Since we obtained the initial flicker visibility from the 51th frame due to the lag responses and judgment times in the human study, we used the value of hfv at the 51th frame to predict the initial flicker visibility. We shifted T to the values of $\widehat{hfv}_{jp}$ in each video j at QP p.

Next, since the range of flicker visibility also depends on the initial flicker visibility, we scaled the shifted T values as follows:

$$T_{ss\_jp} = (\widehat{hfv}_{jp}/100) \times (T - T^{51}) + \widehat{hfv}_{jp}, \tag{12}$$

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. T_{ss_jp} represents the initial flicker visibility and a general visibility change pattern.

3.1.3. Flicker adaptation and accumulation

The sensitivity to flicker distortions in the Human Visual System (HVS) is accumulated or attenuated after prolonged exposure to flickering stimuli [34]. Due to the limited dynamic range of a neuron, visual processing in the retina efficiently adjusts or adapts visual stimuli. Specifically, when a viewer is exposed to large flicker distortions for a longer period of time (e.g., 100 ms), flicker visibility may be affected by "visual persistence," [35] 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 HVS dynamically controls the flicker sensitivity and allocates a finite range of neural signaling, so an observer's flicker sensitivity may be attenuated [34]. Conceptually similar accumulation and adaptation processes may contribute to observed viewers' responses to time-varying video quality as a "recency effect" [36] or "temporal hysteresis" [49]. Here we extend these processes of flicker accumulation and adaptation by accounting for the impact of motion on flicker visibility.

We filtered T_{ss} and object motion using a temporal Gaussian weighting function [50] to remove noise, where the Gaussian window duration was one second. We then applied visual persistence and recency effects to adopt the influence of flicker accumulation and adaptation as a function of stimuli duration and object motion as described in Algorithm 1. Let FV^i represent the predicted flicker visibility at the ith frame and mv_{th} be a threshold velocity that leads to a strong decrease in the visibility of flicker distortions. We used the value mv_{th} (=10.1945) obtained from the human study results described earlier. The value of mv_{th} is computed as the average of all measured thresholds in the test videos. The results do not significantly differ when mv_{th} varies over 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 $mv^i \le mv_{th}$, the flicker adaptation effect on flicker visibility is large, while the influence of motion is small. Since visual persistence changes almost linearly with log stimulus duration at a given intensity, we suggest that FV^i may be accumulated or attenuated as a logarithmic function of time duration t_1 from a base-level, T^i_{ss}, as follows:

$$FV^i = T^i_{ss} + \gamma \log(t_1) \tag{13}$$

where γ is a rate parameter of accumulation or attenuation, where the value was set to -0.4356, 0.7599, and 3.3629 when the QP changes from 32, 38, 44, to 26 respectively. We obtained the value of γ using a least squares fit to the human experimental data.

When $mv^i \ge mv_{th}$, both the object motion and adaptation effect influence flicker visibility. To include the impact of object motion on flicker visibility, we use the measured difference of T_{ss} between frames (e.g., $T^{i+1}_{ss} - T^i_{ss}$) in order to estimate γ. To embrace the influence of adaptation effect on flicker visibility, we use the memory effect: first, we define a memory component $x(i)$ at each instant frame i by averaging the difference of T_{ss} over the previous time duration t_2 or t_3. Second, we construct a current flicker visibility change rate $y(i)$ at each frame instant i using the difference, $T^i_{ss} - T^{i-1}_{ss}$. We then linearly combine the past memory component $x(i)$ and the current change rate $y(i)$, thereby yielding an overall visibility adaptation rate that seeks to explain recency effects at frame i. In our tests, the linear factor $\lambda = 0.7$. We observed that the results do not vary significantly when λ is above 0.5, similar to [49]. 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}, $t_1=0$, $t_2=0$, $t_3=0$, $j=0$, γ, and T_{ss}

for $i=51$: N
  if $mv^i \le mv_{th}$ **and** $j=0$
    $t_1=t_1+1$, $t_2=0$, $t_3=0$
    $FV^i = T^i_{ss} + \gamma \times \log(t_1)$, 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 if
    $x(i)=\text{mean}[(T^{i-t_k}_{ss},\ldots,T^{i-1}_{ss})-(T^{i-t_k-1}_{ss},\ldots,T^{i-2}_{ss})]$
    $y(i)=T^i_{ss}-T^{i-1}_{ss}$
    $FV^i = FV^{i-1} + [\lambda y(i) + (1-\lambda)x(i)] \times \log(t_k)$, $\lambda=0.7$
  end if
end for

3.2. Evaluation of algorithm performance

We evaluated the performance of the proposed algorithm on 24 test videos, which contain the six reference videos and the corresponding 18 flicker distorted videos that were simulated by alternating QP pairs at (QP44, QP26), (QP38, QP26), and (QP32, QP26). As mentioned earlier, this type of flicker was used as it resembles distortions that can be caused by video rate adaptation algorithms.

The predicted flicker visibility from the algorithm and the 95% confidence interval (CI) of the measured flicker visibility by humans are plotted in Fig. 10. Each colored band indicates the 95% CI of the measured flicker visibility at different QP alternations – (QP44, QP26) (), (QP38, QP26) (), (QP32, QP26) (), and each solid line with different markers represents the predicted flicker visibility under the model at different QP alternations – (QP44, QP26) (), (QP38, QP26) (), (QP32, QP26) (). We compared the results in the frame intervals [51,208] for "Bb" and [51,265] for the other videos since 1) observers

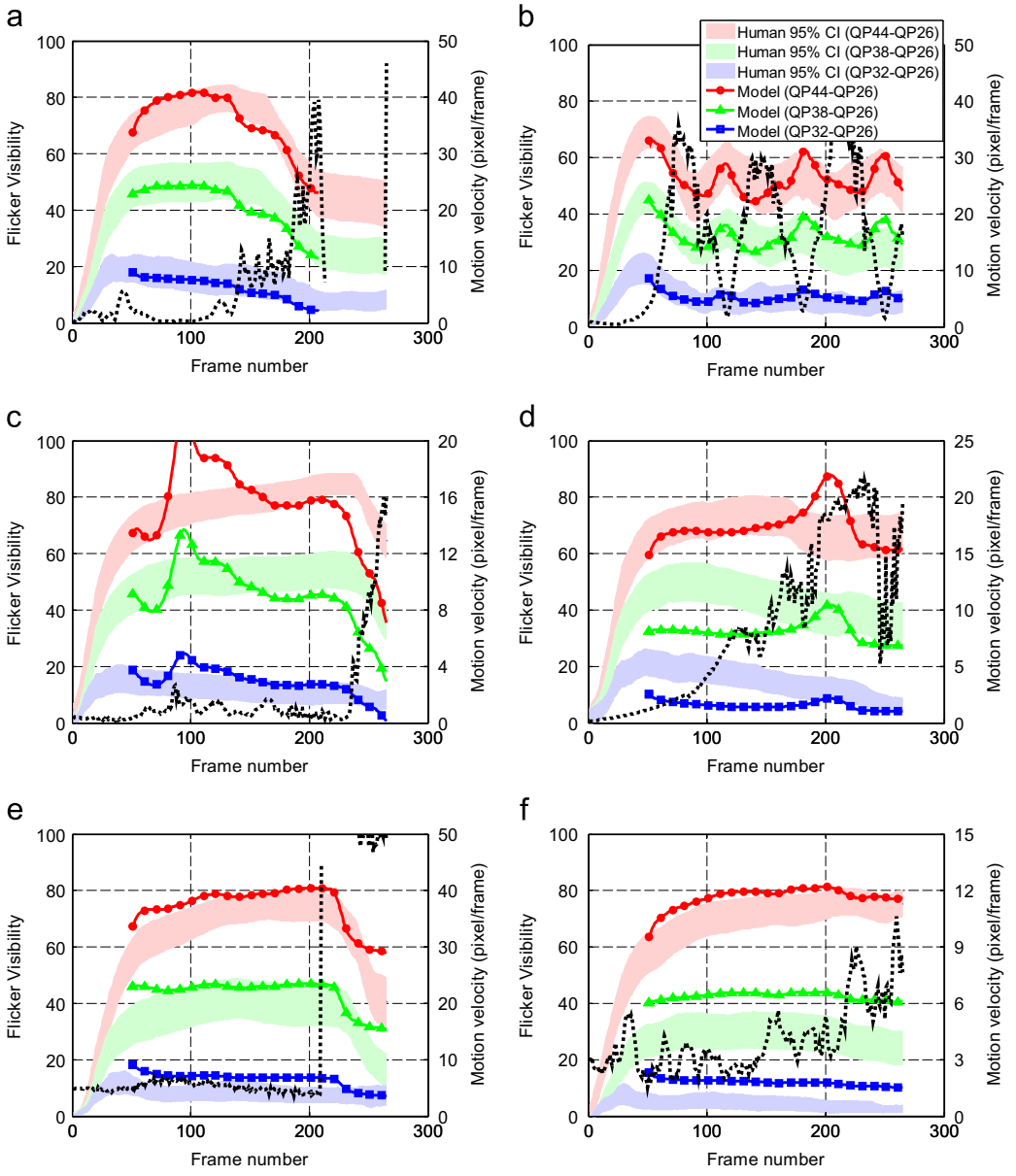

Fig. 10. The predicted visibility of flicker distortions from the proposed model and the 95% CI of the measured visibility of flicker distortions from the human experiments on test videos. Each colored band indicates the 95% CI of the measured flicker visibility at different QP alternations – (QP44, QP26) ( ), (QP38, QP26) ( ), (QP32, QP26) ( ), and each solid line with different markers represents the predicted flicker visibility from the model at different QP alternations – (QP44, QP26) ( ), (QP38, QP26) ( ), (QP32, QP26) ( ), respectively. (a) Bb, (b) BMX, (c) La, (d) Mr, (e) Rc, and (f) Tr.

Table 1
Pearson's Linear Correlation Coefficient (LCC) between the predicted visibility of flicker distortions from the proposed model and the measured visibility of flicker distortions from the human subjective study.

	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
Spearman's Rank Ordered Correlation Coefficient (SROCC) between the predicted visibility of flicker distortions from the proposed model and the measured visibility of flicker distortions from the human subjective study.

	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

needed at least 50 frames (1.67 s) to rate the initial flicker visibility after a test video begins; 2) when moving objects disappeared, no motion data was obtained; and 3) the last 35 frames were shifted to account for a lag response. As seen in Fig. 10, the proposed model can effectively predict variations of the perceived visibility of flicker distortions as reported by humans. When object motions were small, the visibility of flicker distortions was mainly influenced by flicker intensities (caused by QP alterations) and their persistent durations. On the other hand, when object motion was large, the visibility of flicker distortions was strongly reduced.

One interesting observation on the model predictions is that large jumps in 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 the referee in the "La" video sequence induced a flicker visibility score above 100 (around frame 100 at the QP change between QP44 and QP26 in Fig. 10c) in the model prediction, whereas the flicker visibility scores ranged from 0 to 100 in the human experiments following (3). The looming train in the "Mr" video sequence (around frame 200 in Fig. 10d) influenced the prediction of T and subsequently affected the prediction of flicker visibility. Overall, the proposed algorithm successfully captured the influence of object motion on the visibility of flicker distortions.

Tables 1 and 2 show the performance of the prediction method as assessed using the LCC and the Spearman's rank ordered correlation coefficient (SROCC) between the predicted visibility of flicker distortions and the perceived flicker visibility on the test videos after non-linear regression, as described in [51]. Similar to the results shown in Fig. 5, the LCC and SROCC are lower on the "La," "Mr" video sequences than on the others. Although the results vary depending on the video content and the QP alternation range, the average of LCC value was 0.8573. As can be seen in Fig. 10, Tables 1 and 2, the experimental results show that the prediction of flicker visibility using the proposed model correlates quite well with human perception of flicker distortions.

3.3. Possible extensions

In this subsection, we discuss flicker visibility with respect to blur and other related factors. Motion blur may influence flicker visibility. As shown in the human psychophysical experiments on naturalistic videos, when motion increased, flicker visibility decreased. In general, motion also produces blur due to rapid movement of objects or of the camera during the recording of a single frame. Given that motion introduces blur, does flicker masking occur below the threshold of motion blur? Flicker masking and motion blur may be considered as falling within the "window of visibility" of Watson et al. [52]. When flicker and blur signals occur on signals containing primarily low spatial and temporal frequency, both might be visible, whereas on signals exhibiting high spatial and temporal frequency, either signal might be invisible. Temporal luminance changes on moving objects are determined as a product of the flicker frequency and the spatial frequency on the object. When the temporal frequency of the signal is low enough, the HVS adequately samples the luminance changes, so that the changing luminances can be seen. On the other hand, at higher stimulus temporal frequencies under-sampling by the HVS may cause reduced visibility of luminance changes [32]. Since motion blur implies the loss of high spatial frequencies, motion blur may also affect flicker masking. To better understand the combined effect of motion blur on flicker masking, it would be of interest to perform a subjective study extending the one here. For example, it would be helpful to examine whether observers notice any blur while tracking moving objects.

While we used a variety of natural videos in this work, real content is very diverse. It is possible that some kinds of content could adversely affect model predictions of human percepts of flicker visibility, e.g., by conforming to the experimentally obtained model parameters (e.g., mv_{th} and γ in Algorithm 1). It is possible that nighttime conditions, variable weather, or unusual scenes (e.g., an Olympic ice-skating event consisting mostly of ice views) could produce less accurate predictions. Naturally, this work is just a start on this problem and certainly further research on the flicker visibility prediction problem is needed.

4. Conclusion

We analyzed the influence of object motion on the visibility of flicker distortions on naturalistic videos and developed a model of flicker visibility. The results of a human subjective study revealed that flicker visibility on natural videos strongly relies on the speed of object motion, where less flicker was seen on fast-moving objects although subjects held their gaze on the moving objects. The influence of object motion on flicker visibility is more pronounced when video quality is poor. Based on these observations, we believe that sufficiently fast, coherent motion silences the awareness of flicker distortions on naturalistic videos.

In addition, we presented a model of flicker visibility on naturalistic videos using spatiotemporal backward masking and neural adaptation processes. The experimental results show that the proposed model accurately predicts flicker visibility as perceived by humans. In the presence of large coherent object motions, the predicted suppression of flicker visibility on naturalistic videos is in agreement with a recently observed "motion silencing" illusion. Future work will involve connecting the visibility of flicker distortions to perceptual VQA models that can predict "silenced" temporal distortions.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.image.2015.03.006.

References

[1] Cisco Corporation, Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018, 2014. Available:

⟨http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html⟩.
[2] A.C. Bovik, Automatic prediction of perceptual image and video quality, Proc. IEEE 101 (9) (2013) 2008–2024.
[3] L.K. Choi, Y. Liao, A.C. Bovik, Video QoE metrics for the compute continuum, IEEE Commun. Soc. Multimed. Tech. Comm. (MMTC) E-Lett. 8 (5) (2013) 26–29.
[4] M. Yuen, H. Wu, A survey of hybrid MC/DPCM/DCT video coding distortions, Signal Process. 70 (3) (1998) 247–278.
[5] B. Breitmeyer, H. Öğmen, Visual Masking: Time Slices Through Conscious and Unconscious Vision, Oxford University Press, New York, NY, 2006.
[6] S.E. Palmer, Vision Science, MIT Press, Cambridge, MA, 1999.
[7] G.E. Legge, J.M. Foley, Contrast masking in human vision, J. Opt. Soc. Am. A 70 (12) (1980) 1458–1470.
[8] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[9] S.J. Daly, Visible differences predictor: an algorithm for the assessment of image fidelity, in: Proceedings of SPIE 1666 Human Vision Visual Processing and Digital Display III, 1992, pp. 2–15.
[10] A.N. Netravali, B. Prasada, Adaptive quantization of picture signals using spatial masking, Proc. IEEE 65 (4) (1997) 536–548.
[11] M.D. Swanson, B. Zhu, A.H. Tewfik, Multiresolution scene-based watermarking using perceptual models, IEEE J. Sel. Areas Commun. 16 (4) (1998) 540–550.
[12] G. Sperling, Temporal and spatial visual masking I. Masking by impulse flashes, J. Opt. Soc. Am. A 55 (5) (1965) 541–559.
[13] B.E. Rogowitz, Spatial/temporal interactions: backward and forward metacontrast masking with sine-wave gratings, Vis. Res. 23 (10) (1983) 1057–1073.
[14] F. Hermens, G. Luksys, W. Gerstner, M. Herzog, U. Ernst, Modeling spatial and temporal aspects of visual backward masking, Psychol. Rev. 225 (1) (2008) 83–100.
[15] A.J. Seyler, Z. Budrikis, Detail perception after scene changes in television image presentations, IEEE Trans. Inf. Theory 11 (1) (1965) 31–43.
[16] B. Girod, The information theoretical significance of spatial and temporal masking in video signals, C 1077, in: Proceedings of Human Vision Visual Processing and Digital Display, 1989, pp. 178–187.
[17] J.D. Johnston, S.C. Knauer, K.N. Matthews, A.N. Netravali, E.D. Petajan, R.J. Safranek, P.H. Westerink, Adaptive non-linear quantizer, U.S. Patent 5,136,377, August 4, 1992.
[18] A. Puri, R. Aravind, Motion-compensated video with adaptive perceptual quantization, IEEE Trans. Circuits Syst. Video Technol. 1 (1991) 351–378.
[19] B.G. Haskell, F.W. Mounts, J.C. Candy, Interframe coding of videotelephone pictures, Proc. IEEE 60 (1972) 792–800.
[20] I. Hontsch, L.J. Karam, Adaptive image coding with perceptual distortion control, IEEE Trans. Image Process. 11 (3) (2002) 213–222.
[21] Y. Zhao, L. Yu, Z. Chen, C. Zhu, Video quality assessment based on measuring perceptual noise from spatial and temporal perspectives, IEEE Trans. Circuits Syst. Video Technol. 21 (12) (2011) 1890–1902.
[22] A.M. Tekalp, Digital Video Processing, Prentice-Hall PTR, Upper Saddle River, NJ, 1995.
[23] R.D. Kell, A.V. Bedford, M.A. Trainer, Scanning sequence and repetition rate of television image, Proc. Inst. Radio Eng. 24 (4) (1936) 559–576.
[24] G. de Hann, E.B. Bellers, Deinterlacing – an overview, Proc. IEEE 86 (9) (1998) 1839–1857.
[25] X. Fan, W. Gao, Y. Lu, D. Zhao, Flickering reduction in all intra frame coding, JVT-E070, in: Proceedings of the Joint Video Team of ISO/IEC MPEG & ITU-T VCEG Meeting, 2002.
[26] Y. Kuszpet, D. Kletsel, Y. Moshe, A. Levy, Post-processing for flicker reduction in H.264/AVC, in: Proceedings of the Picture Coding Symposium, 2007.
[27] J. Yang, H.R. Wu, Robust filtering technique for reduction of temporal fluctuation in H.264 video sequences, IEEE Trans. Circuits Syst. Video Technol. 20 (3) (2010) 458–462.
[28] P. Ni, R. Eg, A. Eichhorn, C. Griwodz, P. Halvorsen, Flicker effects in adaptive video streaming to handheld devices, in: Proceedings of the ACM International Conference on Multimedia, 2011, pp. 463–472.
[29] J.W. Suchow, G.A. Alvarez, Motion silences awareness of visual change, Curr. Biol. 21 (2) (2011) 140–143.
[30] R.L. Gregory, Eye and Brain: the Psychology of Seeing, Princeton, University Press, Princeton, NJ, 1997.
[31] L.K. Choi, A.C. Bovik, L.K. Cormack, A flicker detector model of the motion silencing illusion, J. Vis. 12 (9) (2012) 777.
[32] L.K. Choi, A.C. Bovik, L.K. Cormack, Spatiotemporal flicker detector model of motion silencing, Perception 43 (12) (2014) 1286–1302.
[33] H.R. Wilson, J.D. Cowan, A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue, Kybernetik 13 (2) (1973) 55–80.
[34] S. Shady, D.I.A. MacLeod, H.S. Fisher, Adaptation from invisible flicker, Proc. Natl. Acad. Sci. USA 101 (14) (2004) 5170–5173.
[35] R.W. Bowen, J. Pola, L. Matin, Visual persistence: effects of flash luminance, duration and energy, Vis. Res. 14 (4) (1974) 295–303.
[36] J.R. Anderson, M. Matessa, A production system theory of serial memory, Psychol. Rev. 104 (4) (1997) 728–748.
[37] L.K. Choi, L.K. Cormack, A.C. Bovik, On the visibility of flicker distortions in naturalistic videos, in: Proceedings of IEEE International Workshop on Quality of Multimedia Experience, 2013, pp. 164–169.
[38] L.K. Choi, L.K. Cormack, A.C. Bovik, Visibility prediction of flicker distortions on naturalistic videos, in: Proceedings of IEEE Asilomar Conference on Signals, Systems and Computers, 2014, pp. 1542-1546.
[39] H.264/AVC software coordination. Available: http://iphome.hhi.de/suehring/tml/.
[40] ITU-R Rec. B.T. 500-11, Methodology for the Subjective Assessment of the Quality of Television Pictures, 2002.
[41] The XGL Toolbox, 2008. Available: ⟨http://svi.cps.utexas.edu/software.shtml⟩.
[42] D. Sun, S. Roth, M.J. Black, Secrets of optical flow estimation and their principles, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 2432–2439.
[43] T.J. Spencer, R. Shuntich, Evidence for an interruption theory of backward masking, J. Exp. Psychol. 85 (2) (1970) 198–203.
[44] A.R. Reibman and D. Poole, Predicting packet-loss visibility using scene characteristics, in: Proceedings of IEEE International Packet Video Workshop, 2007, pp. 308–317.
[45] Z. Wang, L. Lu, A.C. Bovik, Foveation scalable video coding with automatic fixation selection, IEEE Trans. Image Process. 12 (2) (2003) 243–254.
[46] J.E. Farrell, B.L. Benson, C.R. Haynie, Prediction flickering thresholds for video display terminals, Proc. SID 28 (4) (1987).
[47] L. Breiman, P. Spector, Submodel selection and evaluation in regression. The x-random case, Int. Stat. Rev. 60 (3) (1992) 291–319.
[48] Final Report from the Video Quality Experts Group on the Validation of Objective Quality Metrics for Video Quality Assessment Phase I, Video Quality Experts Group (VQEG), 2000. Available: ⟨http://www.its.bldrdoc.gov/vqeg/projects/frtv-phase-i/frtv-phase-i.aspx⟩.
[49] K. Seshadrinathan, A. Bovik, Temporal hysteresis model of time varying subjective video quality, in: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2011, pp. 1153–1156.
[50] H.G. Longbothan, A.C. Bovik, Theory of order statistic filters and their relationship to linear FIR filters, IEEE Trans. Acoust. Speech Signal Process. 37 (2) (1989) 275–287.
[51] H.R. Sheikh, M.F. Sabir, A.C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Trans. Image Process. 15 (11) (2006) 3440–3451.
[52] A.B. Watson, A.J. Ahumada Jr., J.E. Farrell, Window of visibility – a psychophysical theory of fidelity in time-sampled visual motion displays, J. Opt. Soc. Am. A 3 (3) (1986) 300–307.