

# Predicting visible flicker in temporally changing images

Gyorgy Denes and Rafal K. Mantiuk

Department of Computer Science and Technology, University of Cambridge, United Kingdom

## Abstract

*Novel display algorithms such as low-persistence displays, black frame insertion, and temporal resolution multiplexing introduce temporal change into images at 40-180 Hz, on the boundary of the temporal integration of the visual system. This can lead to flicker, a highly-objectionable artifact known to induce viewer discomfort. The critical flicker frequency (CFF) alone does not model this phenomenon well, as flicker sensitivity varies with contrast, and spatial frequency; a content-aware model is required. In this paper, we introduce a visual model for predicting flicker visibility in temporally changing images. The model performs a multi-scale analysis on the difference between consecutive frames, normalizing values with the spatio-temporal contrast sensitivity function as approximated by the pyramid of visibility. The output of the model is a 2D detection probability map. We ran a subjective flicker marking experiment to fit the model parameters, then analyze the difference between two display algorithms, black frame insertion and temporal resolution multiplexing, to demonstrate the application of our model.*

## Introduction

Displays are known to produce temporally changing signals, relying on the limited integration time of human visual system which then fuses this into a steady image. Well-established technologies, such as digital micromirror devices and projector colorwheels operate at such high frequencies that the change is indeed unperceivable. However, novel algorithms introduce temporal changes at 40-180 Hz, on the boundary of temporal integration of the visual system. This can lead to highly-objectionable flicker.

The visibility of flicker is commonly associated with the critical flicker frequency (CFF). However, many other factors such as the adaptation luminance and the spatial frequency of the stimulus have a significant impact. Therefore, a content-dependent model is needed to produce accurate predictions. Conventional video quality metrics are designed for lower temporal frequencies, and hence do not model high-frequency flicker detection accurately.

In this paper, we introduce a visual model for predicting flicker visibility in temporally changing images. The model performs a multi-scale analysis of the difference image between consecutive frames of the animation, normalizing contrast values with the spatio-temporal contrast sensitivity function from the pyramid of visibility [1]. To calibrate the model parameters, we conduct a psychophysical experiment, collecting user markings for a wide range of content across different refresh rates.

The rest of the paper is structured as follows: first, we review the relevant publications on flicker sensitivity, temporal display algorithms and quality metrics. Then, we introduce our flicker model, describe the psychophysical experiment and the parameter fitting process. Finally, we demonstrate the application of our model by comparing flicker in two display algorithms.

## Flicker sensitivity

Most artificial light sources do not produce a temporally-stable amount of light; they are known to vary with time [2]. The minimum frequency at which a light is perfectly fused and is perceived as steady is the flicker fusion threshold, or as also known: the critical flicker frequency (CFF). CFF depends on multiple factors: it is known to increase with log-luminance (Ferry-Porter Law) and angular size; to vary with spatial frequency and eccentricity – the most sensitive being the para-fovea [3, 2].

CFF is typically measured for stimuli with full-on, full-off cycles; however, the amplitude of luminance change in display algorithms is often small. Hence, flicker visibility is better captured by the spatio-temporal contrast sensitivity function (CSF). CSF is usually characterized as a function of spatial frequency, temporal frequency, and background luminance. Studies suggest that contrast sensitivity across the retina is homogeneous when the stimulus size increases with the cortical magnification factor [2, 4]. On the other hand, Peli et al. argues that the contrast threshold increases exponentially with eccentricity, and the CSF should be scaled accordingly. In this paper, we assume a worst-case scenario, a free-viewing set-up with flicker attended by the most sensitive part of the retina. Sensitivity values can be interpreted as the inverse of the contrast thresholds at which flicker is observed by 50% probability. We use a slightly different definition of sensitivity, as the inverse of the contrast threshold at which flicker is observed by 50% of the population. Contrast in this context is commonly defined as Michelson contrast:

$$C = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}}, \quad (1)$$

where  $L_{\min}$  and  $L_{\max}$  are the minimum and maximum luminance values of a periodically changing signal.

The pyramid of visibility [1] offers an approximate model of the CSF, matching previous foveal measurements for medium-to-high spatio-temporal frequencies. However, model parameters were not fully consistent when fitted to different datasets. Furthermore, the pyramid of visibility can explain flicker visibility for single spatial frequencies, but not necessarily for complex images. We apply this model in a novel way to multiple spatial frequency bands, refitting the parameters to our proposed application.

Flicker artifacts discussed in this paper are around the detection threshold, and as such, we focus on their detectability. We have not conducted extensive studies on long-term fatigue, and do not consider photosensitive epilepsy [5].

## Display algorithms

There are a number of computer graphics and display algorithms that introduce temporal change into images [6, 7, 8, 9, 10], also referred to as temporal multiplexing algorithms. Commonly

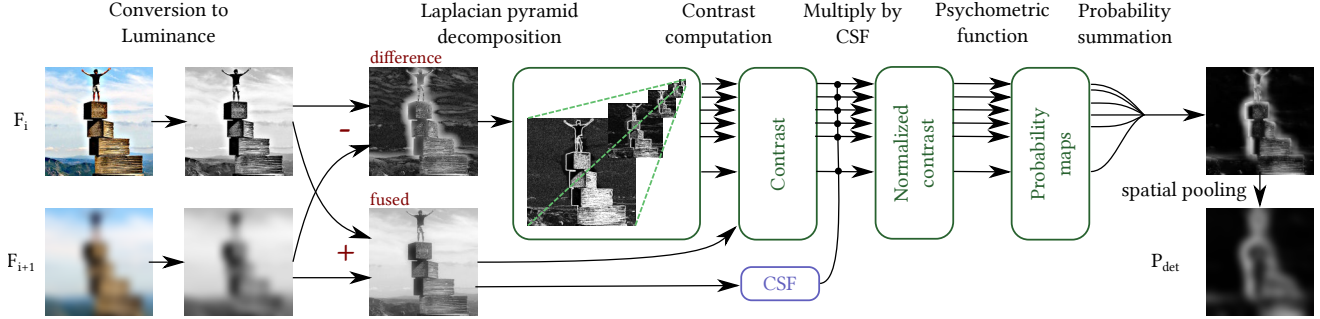


Figure 1: Overview of the flicker predictor model. The input is a pair of color frames; the result of the model is a 2D probability of detection ( $P_{det}$ ) map.

these modify a few consecutive frames in order to perceived resolution, reduce the computational or transmission cost.

Chen et al. [6] adjusts pairs of frames to reduce motion blur on a 120 Hz display. Didyk et al. [8] modifies up to three frames to boost the apparent resolution of moving images. The latter introduces low-contrast flicker at 40 Hz – well below the CFF; authors mitigate this with a multi-scale CFF predictor. While our model is conceptually similar, we rely on a multi-scale CSF model, with psychophysical calibration for robustness.

Other algorithms introduce temporal change to reduce the computational cost of image generation. In virtual reality, Asynchronous SpaceWarp (ASW) re-projects every other frame based on image-space motion information [9]; black frame insertion (BFI) assumes that every other frame is black, boosting display luminance to match the target content. More recently, temporal resolution multiplexing (TRM) produces pairs of blurred and sharpened frames that are fused on the retina [10]. In the paper we argued that TRM introduces less flicker than BFI. We now demonstrate how our flicker model can be utilized for validation.

## Image and video metrics

Image and video metrics can be broadly categorized as quality metrics and visibility metrics. Full-reference quality metrics, such as PSNR, output a single quality value for the entire image, whereas visibility metrics, such as VDP, HDR-VDP [11, 12] produce a distortion map, providing spatial information on the probability of detecting artifacts. Some metrics, such as SSIM produce a distortion map, but it is only the mean value over the image that is shown to correlate with subjective scores [13, 14], SSIM distortion maps are not calibrated to psychophysical data. Our flicker prediction model is inspired by the visual difference predictors in the sense that it also outputs psychophysically calibrated detection probability maps rather than a single quality value.

Video quality is often evaluated by executing an image quality metric frame-by-frame and pooling quality values over time. Some video quality metrics, such as VQM and MOVIE [15], analyze temporal changes in videos. However, these metrics are designed for film content with frame rates typically below 50 Hz, and lack a robust flicker model for  $>50$  Hz content. Furthermore, many novel display algorithms that we target are designed specifically for real-time computer generated (CG) content. Video quality metrics are not immediately applicable to CG, as CG content has a subjective perception unaffected by the well-established rules of the film industry such as the soap opera effect [16].

## Flicker model

Temporal multiplexing algorithms often manipulate pairs of frames. Let us denote two consecutive frames as  $F_i$  and  $F_{i+1}$  that could be, for instance, the reduced-resolution and sharpened frames of TRM; or a black frame and a luminance-boosted frame of BFI. The proposed visual model predicts whether displaying  $F_i$  and  $F_{i+1}$  alternately at refresh rate  $R$  would result in perceivable flicker. The output is a  $P_{det}(x, y)$  map corresponding to the percentage of the population detecting flicker at a pixel  $(x, y)$ .

Our flicker predictor utilizes a spatio-temporal CSF in a multi-scale model with probability summation along the spatial frequency bands. For an overview of the pipeline, please see Figure 1. The model does not distinguish orientation-sensitive bands, often found in masking models [17]. The model also assumes that eye movements have been already accounted for in image-space; i.e., the same pixels on  $F_i$  and  $F_{i+1}$  will correspond to roughly the same photoreceptors on the retina.

As the source of most flicker is the change in luminance, we do not consider chromaticity here. First, we compute respective luminance values ( $Y_i$  and  $Y_{i+1}$ ) of consecutive frames ( $F_i$  and  $F_{i+1}$ ) based on a calibrated display model. Then, to find contrast, we compute the difference image:

$$\Delta(x, y) = Y_i(x, y) - Y_{i+1}(x, y), \quad (2)$$

where  $x$  and  $y$  describe pixel location, and  $Y_i(x, y)$  is the luminance of pixel  $(x, y)$  of the frame  $F_i$ . The summed luminance of the consecutive frames can be similarly defined as:

$$\bar{Y}(x, y) = Y_i(x, y) + Y_{i+1}(x, y) \quad (3)$$

As flicker sensitivity varies with spatial frequency, the difference image ( $\Delta(x, y)$ ) is decomposed into a Laplacian pyramid. Each layer of the pyramid is half the spatial resolution of the one above; the bottom layer capturing 2 cycles per visual degree (cpd) resolution or just below – e.g. for a 52 pixel-per-degree image the mid points of the spatial frequency bands are  $S_i = \{26, 13, 6.5, 3.25, 1.625\}$  cpd. We use an undecimated pyramid, in which each band has the same resolution.

In each layer, Michelson contrast can be then computed as:

$$C(x, y, l) = \frac{|\Delta(x, y, l)|}{\bar{Y}(x, y)}, \quad (4)$$

where  $l$  is the Laplacian pyramid layer. To account for contrast sensitivity, we normalize contrast at each layer by a the spatio-

temporal CSF ( $\rho$ ). We use the pyramid of visibility, as it is parametric, and has been shown to provide a good fit to previous CSF measurements.

$$\rho(W, F, L) = \exp(c_0 + c_W W + c_F F + c_L \log Y), \quad (5)$$

where  $W$  and  $F$  are the spatial and temporal frequencies as in the original paper,  $Y$  is the adapting luminance, and  $(c_0, c_W, c_F, c_L)$  are parameters that we keep as free variables in our model. We assume that the mean fused image  $0.5\bar{Y}(x, y)$  provides a good estimate of the local adapting luminance. The normalized contrast is then:

$$\hat{C}(x, y, l) = C(x, y, l) \rho(R/2, S_l, 0.5\bar{Y}(x, y)), \quad (6)$$

where  $S_l$  is the spatial frequency of the layer, and  $R$  is the display refresh rate. We use  $R/2$  to sample the temporal dimension of the CSF, as when modifying pairs of frames, this is the highest temporal frequency according to the Nyquist limit.

Next, to transform the normalized contrast into probabilities of detection, we use a Weibull psychometric function:

$$P(x, y, l) = 1 - \frac{\exp(-\hat{C}^\beta(x, y, l))}{2}, \quad (7)$$

where  $\beta$  controls the slope of the psychometric function, a free parameter in our model. In order to pool the probabilities across all layers, we use probability summation:

$$P(x, y) = 1 - \prod_l (1 - P(x, y, l)). \quad (8)$$

Finally, to account for spatial pooling, we further convolve the probability map with a small Gaussian filter:

$$P_{\text{det}}(x, y) = P(x, y) * G_{\sigma_{sp}}, \quad (9)$$

where  $*$  denotes convolution, and  $G_{\sigma_{sp}}$  is a Gaussian kernel with  $\sigma_{sp}$  being a free parameter in our model.

## Flicker marking experiment

To tune the model parameters, ground-truth data is required on flicker perception in complex images. As flicker is often perceived in multiple parts of the image, and location information is crucial for  $P_{\text{det}}(x, y)$ , we designed a marking experiment. Such experiments have been utilized to calibrate similar metrics for image difference predictors [18, 19].

## Participants

Nineteen participants aged 18-40 with normal or corrected-to-normal vision took part in the experiment. As the frequency of flicker was above 3 Hz, we ensured that no participant reported a history of photosensitive epilepsy. Informed consent was acquired before the beginning of the experiment, which involved briefing participants on the aim, the procedure, and potential risks of the experiment both verbally and in writing. Participants were offered a small financial compensation in the form of gift cards.

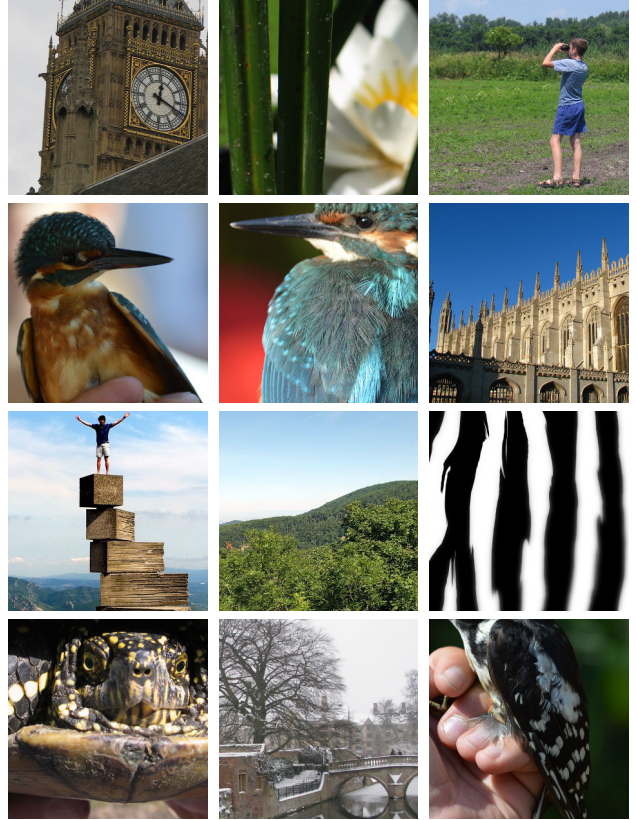


Figure 2: Pool of reference images used for the flicker marking experiment with a range of content.

## Setup

Participants were shown a  $512 \times 512$  pixel flickering photograph in the center of a G-Sync capable ASUS ROG Swift PG279Q 28" monitor. The viewing distance was fixed at 65 cm, yielding an angular resolution of 52 pixels per degree (ppd). Images hence had a field of view of  $9.85^\circ$ ; the rest of the monitor was filled with a gray background of  $36 \text{ cd/m}^2$ . Accurate refresh rates were achieved with custom C++/OpenGL software and G-Sync.

## Stimuli

Eighteen stimuli were created by flickering twelve color photographs (see Figure 2). The photographs provided a range of content from primitive zebra stripes, photographs of birds, buildings and people. For each trial, a spatially band-limited flicker was introduced at temporal frequency  $R$ . This was achieved by displaying a pair of images ( $F_i$  and  $F_{i+1}$ ) alternating at  $R$  Hz such that,

$$F_i(x, y) = F_{\text{ref}}(x, y) * g_\sigma(x, y), \quad (10)$$

where  $F_{\text{ref}}$  denotes the original reference image in gamma-compressed  $rgb$  color space (rec.709 primaries),  $*$  stands for 2D convolution, and  $g_\sigma(x, y)$  is an isotropic Gaussian blur kernel with a standard deviation of  $\sigma$ . Inspired by the temporal resolution multiplexing algorithm [10],  $F_{i+1}$  was computed as

$$F_{i+1}(x, y) = \xi^{-1} \left( 2\xi(F_{\text{ref}}(x, y)) - \xi(F_i(x, y)) \right), \quad (11)$$



Figure 3: Example stimulus pair. Reference frame is low-pass filtered ( $F_i$ ), and sharpened ( $F_{i+1}$ ) to produce band-limited flicker.

where  $\xi()$  is a gain-offset-gamma display model [20] transforming gamma-compressed  $rgb$  to linear  $rgb$  values. For stimulus generation we used a crude approximation of the display model of the ASUS monitor with

$$r'(x,y) = 0.99917r(x,y)^{2.15} + 0.000825, \quad (12)$$

where  $r$  is the red channel. The same formula was applied to all color channels. For an example image pair, see Figure 3. Note that we do not use an accurate display model here, neither do we account for the limited dynamic range and color gamut of the display. This might introduce some spatial artifacts to the viewer, but this experiment did not attempt to establish overall visual quality, and such artifacts did not impede flicker detection. For a summary of the stimuli images,  $\sigma$  and  $R$  values, refer to Figure 4.

For model calibration, specifically for the first step of conversion to luminance, we measured a more accurate display model with a Specbos 1211 spectroradiometer.

### Task

Participants were asked to “mark (or *paint*) any part of the image where flicker is visible” – quoted from the briefing form. Flickering areas could be marked by holding down the left mouse button and moving the pointer around. Previous markings could be deleted with the right mouse button in a similar fashion. A circular mouse pointer was used with the diameter adjustable from  $0.15^\circ$  (8 pixels) to  $2^\circ$  (104 pixels) using the mouse wheel. Any marked area was highlighted and immediately stopped flickering. At the beginning of each trial, the mask was cleared. Participants were specifically asked to first mark (and hence remove) the strongest flicker first to minimize the effect of masking. During briefing and training it was highlighted that flicker might be more visible in the parafoveal area of vision, and hence looking at objects slightly off-center might reveal more flicker. This was to ensure that all participants utilize the free-viewing setup equally.

Each participant created a marking map for each stimulus three times, yielding 54 trials in total. The order of the trials was randomized.

### Results

Figure 4 shows the flicker marking maps averaged over nineteen observers and three repetitions. As expected, flicker perception degrades with increasing refresh rates and increases with the blur  $\sigma$ . The markings, however, cannot be considered ground truth data for two reasons: (1) participants might make mistakes producing mis-markings, and (2) the finite size of the brush allows for limited precision.

Following the analysis in [18], markings can be considered the output of a stochastic process, where observers attend

to a distortion with  $P_{att}$ , and mis-mark a pixel with probability  $P_{mis} = 0.01$ . Due to the small image size ( $9.85^\circ \times 9.85^\circ$ ), and the characteristics of temporal sensitivity, we assumed  $P_{att} = 1$  for this experiment.

For each image in each of the 57 marking maps each  $(x,y)$  pixel takes a binary  $\{0, 1\}$  value depending on whether the participant marked it with the mouse. Assuming a detection probability  $P_{det}(x,y)$ , the data can be modeled as a binomial distribution. Accounting for the mis-markings, the likelihood of observing the collected data given a model is:

$$\Lambda(x,y) = P_{mis} + (1 - P_{mis}) \binom{n}{k} P_{det}(x,y)^k [1 - P_{det}(x,y)]^{n-k}, \quad (13)$$

where  $n = 57$  is the number of all collected markings for an image,  $k$  is the number of trials where the pixel is marked to flicker, and  $P_{det}(x,y)$  is the predicted detection probability of the model.

### Parameter fitting

We posed the task of finding the best model parameters as a non-linear optimization problem, maximizing the average log-likelihood over the images. However, we observed that the effects of spatial pooling were masked by the finite paint brush size; therefore we decided to fix this parameter to a value comparable to the brush sizes ( $\sigma = 0.36^\circ$ ). Variable slope values in the psychometric function were also expected to create a range of local minima, hence we selected a single likely candidate  $\beta = 2$ . The remaining parameters are parameters from the pyramid of visibility which we restricted to physically sensible ranges ( $c_W < 0$  for decreasing sensitivity with temporal frequency;  $c_F < 0$  for decreasing sensitivity with spatial frequency;  $c_L > 0$  for increasing sensitivity with background luminance).

Our results are summarized in Table 1. When refitting the parameters from the original pyramid of visibility, the mean log-likelihood increases, as expected. Parameters show some deviation from the values fitted to the Robson measurements. While  $c_0$  is comparable, increasing temporal frequencies attenuate sensitivity faster (lower  $c_W$  values), increasing spatial frequencies attenuate sensitivity slower, and luminance amplifies sensitivity faster. Such deviations are to be expected due to the significantly more complex nature of the task presented in the marking experiment.

To analyze the possibility of over-fitting the model parameters to our dataset, we also executed a 3-fold cross-validation. For this, the dataset was randomly split into three 6-element groups. The models was fit to each two groups (training), then performance was evaluated on the third groups (test). Results in Table 1 indicate that the training and test likelihoods were comparable, and the optimum parameters did not differ significantly from the scenario when all 18 images were included in the training dataset.

Qualitatively we observed that model predictions for the experiment stimuli capture the flickering details well. The user-produced and the predicted markings are as shown in Figure 4.

### Application

To demonstrate the utility of our model, we analyze the amount of flicker introduced by two state-of-the-art display algorithms: temporal resolution multiplexing (TRM) and black frame insertion (BFI). For TRM we assume the worst-case scenario and ignore the residual buffer (for full details, see [10]). For black



	training	test	$c_0$	$c_W$	$c_F$	$c_L$
Pyramid of visibility	-	-3.204	2.1900	-0.0600	-0.0650	0.3880
Fit to all	-2.849	-	1.9993	-0.1059	-0.0242	0.9102
cross 1	-2.829	-2.933	2.1946	-0.1025	-0.0362	0.8581
cross 2	-2.903	-2.779	1.9982	-0.1146	-0.0113	0.9782
cross 3	-2.782	-3.009	1.5741	-0.1006	-0.0128	0.9265

Table 1: Parameters, training and test fitness (measured as mean log-likelihood). Pyramid of visibility (first row) uses the parameters from the Robson fit from [1]. When fitting to our dataset, the log likelihood increases (Fit to all). Cross # are indicate results of the 3-fold cross-validation.

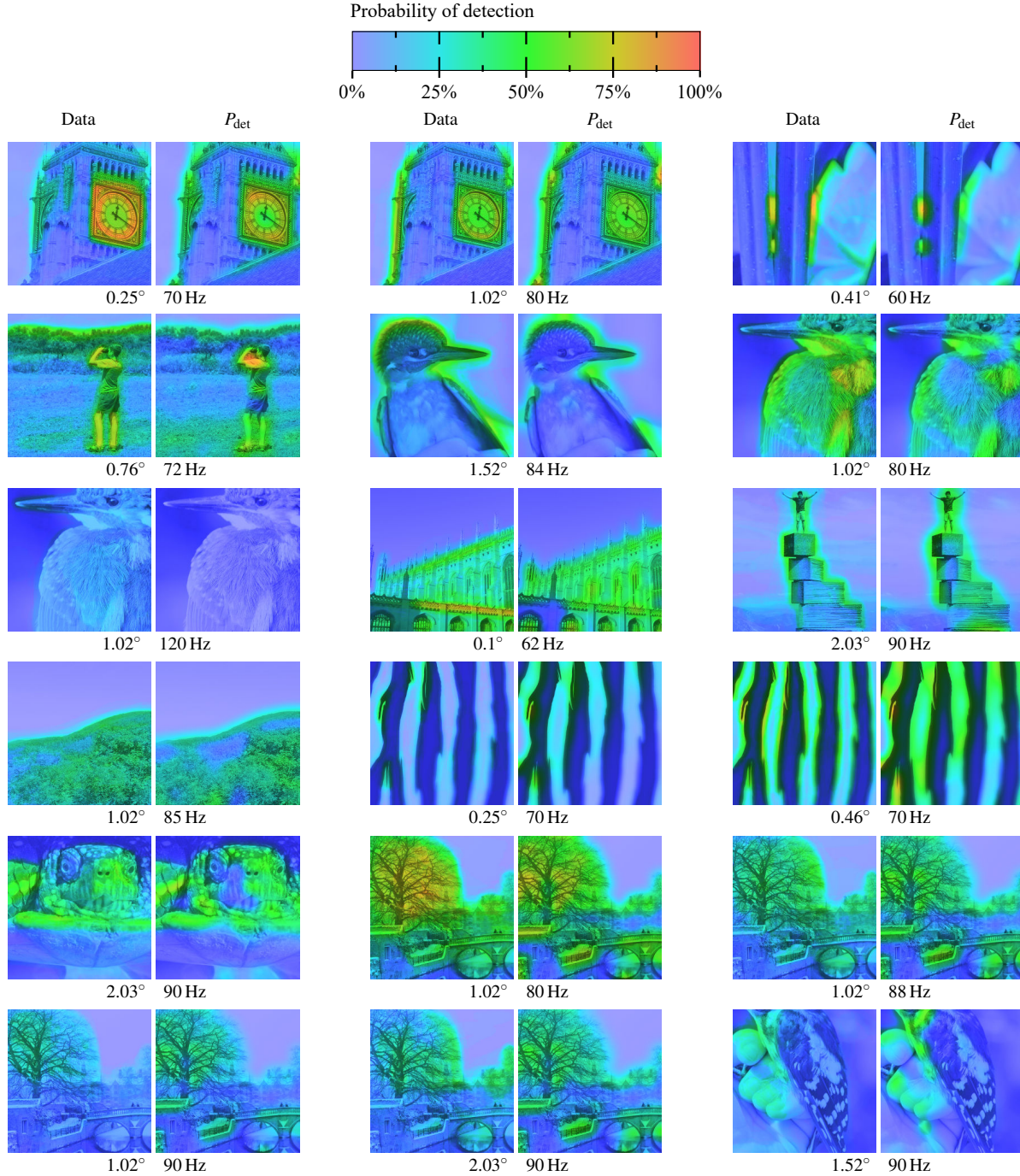


Figure 4: Flicker markings (Data) and model predictions ( $P_{\text{det}}$ ) overlaid on the 18 reference images. Blue indicates no flicker ( $P_{\text{det}} \approx 0$ ), green to orange indicates strong perceivable flicker ( $P_{\text{det}} \rightarrow 1$ ). Sub-captions state the standard deviation of the Gaussian kernel (in visual degrees), and the refresh rate at which  $F_i$  and  $F_{i+1}$  were alternated (in Hz).

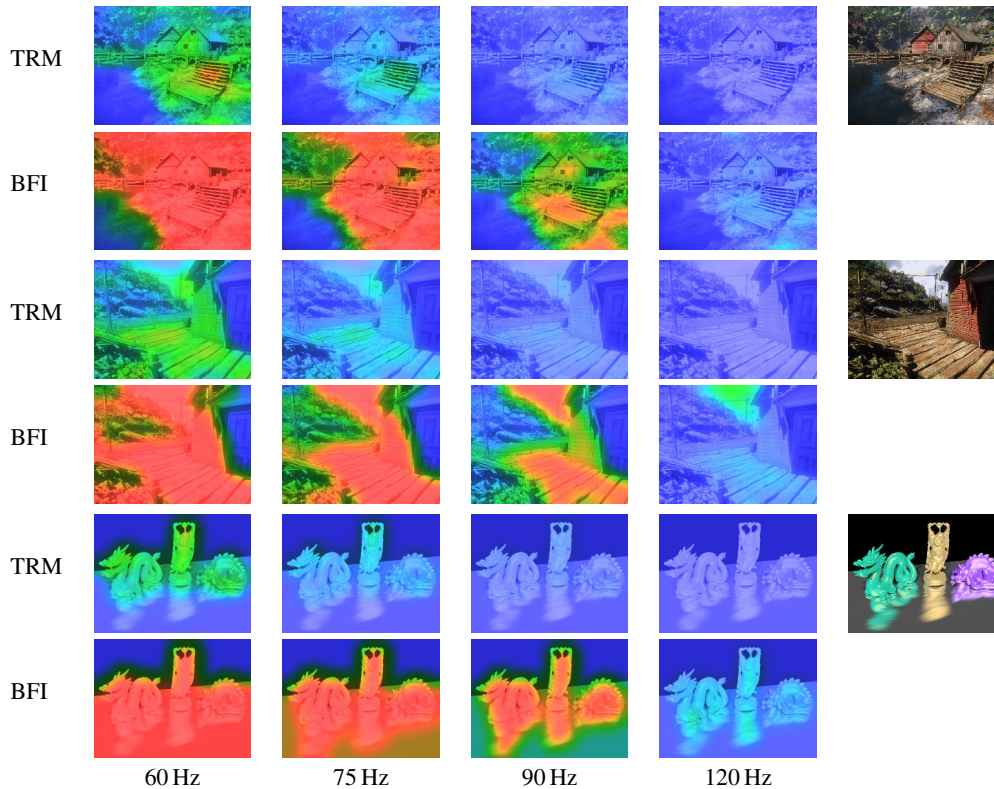


Table 2: Analysis of flicker artifacts in black frame insertion (BFI) and temporal resolution multiplexing (TRM) across refresh rates. Red regions indicate that BFI produces noticeable flicker artifacts at bright regions of the image even on 120 Hz. TRM, on the other hand, does not flicker above 90 Hz on the analyzed content for our display with 156 cd/m<sup>2</sup> peak luminance.

frame insertion we assume that  $F_i$  is completely black, while  $F_{i+1}$  is boosted to double the luminance. For representative content we selected three computer-generated images.

Our model can be used to establish the minimum refresh rate at which flicker is no longer perceivable when using either temporal techniques. As shown in Table 2, TRM generally requires lower refresh rates; it is unlikely to be perceived as flickery on 90 Hz, while BFI causes minor distortions even on 120 Hz. This is consistent with previous observations [10].

## Conclusion

We presented a multi-scale visual model for predicting flicker, one of the most objectionable artifacts in display algorithms that introduce temporal change. Specifically, we assumed that such algorithms produce consecutive frames that are fused on the retina. Our model takes spatio-temporal sensitivity into account to predict if perfect fusion is possible, and outputs a detection probability map showing a 2D image of flicker visibility. Model parameters were fitted to the results of a subjective marking experiment. Our model provided a good fit to the observed data. We demonstrate how the model can be utilized to analyze existing multiplexing algorithms: temporal resolution multiplexing, and black frame insertion. The main limitation of the current model is the lack of masking effects, including cross-channel spatial sensitivity masking and motion masking. In future work we wish to collect further data to address these issues.

## Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement N° 725253–EyeCode), and EPSRC EP/N509620/1.

## References

- [1] Andrew B Watson and Albert J Ahumada. The pyramid of visibility. In *Human Vision and Electronic Imaging*, volume 2016, pages 1–6, feb 2016.
- [2] Andrew B. Watson. Temporal sensitivity. In *Handbook of perception and human performance*, pages 6–1. 1986.
- [3] E. Hartmann, B. Lachenmayr, and H. Brettel. The peripheral critical flicker frequency. *Vision Research*, 19(9):1019–1023, 1979.
- [4] Veijo Virsu, Jyrki Rovamo, Pentti Laurinen, and Risto Näsänen. Temporal contrast sensitivity and cortical magnification. *Vision Research*, 22(9):1211–1217, jan 1982.
- [5] CD Binnie, J Emmett, P Gardiner, GFA Harding, D Harrison, and AJ Wilkins. Characterizing the flashing television images that precipitate seizures. *SMPTE journal*, 111(6-7):323–329, 2002.
- [6] Hanfeng Chen, Sung-soo Kim, Sung-hee Lee, Oh-jae Kwon, and Jun-ho Sung. Nonlinearity compensated smooth frame insertion for motion-blur reduction in LCD. In *2005 IEEE 7th Workshop on Multimedia Signal Processing*, pages 1–4. IEEE, oct 2005.
- [7] Piotr Didyk, Elmar Eisemann, Tobias Ritschel, Karol Myszkowski, and Hans Peter Seidel. Perceptually-motivated real-time temporal upsampling of 3D content for high-refresh-rate displays. *Computer*

- Graphics Forum*, 29(2):713–722, 2010.
- [8] Piotr Didyk, Elmar Eisemann, Tobias Ritschel, Karol Myszkowski, and Hans-Peter Seidel. Apparent display resolution enhancement for moving images. *ACM Transactions on Graphics*, 29(4):1, jul 2010.
  - [9] Dean Beeler, Ed Hutchins, and Paul Pedriana. Asynchronous spacewarp. <https://developer.oculus.com/blog/asynchronous-spacewarp/>, 2016. Accessed: 2018-05-02.
  - [10] Gyorgy Denes, Kuba Maruszczyk, George Ash, and Rafał K. Mantiuk. Temporal Resolution Multiplexing: Exploiting the limitations of spatio-temporal vision for more efficient VR rendering. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):2072–2082, may 2019.
  - [11] Scott J. Daly. Visible differences predictor: an algorithm for the assessment of image fidelity. page 2, 1992.
  - [12] Rafał K. Mantiuk, Kil Joong Kim, Allan G. Rempel, and Wolfgang Heidrich. HDR-VDP-2. *ACM Transactions on Graphics*, 30(4):1, jul 2011.
  - [13] Zhou Wang, Eero P. Simoncelli, and Alan C. Bovik. Multi-scale structural similarity for image quality assessment. *Conference Record of the Asilomar Conference on Signals, Systems and Computers*, 2:1398–1402, 2003.
  - [14] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, apr 2004.
  - [15] K. Seshadrinathan and A.C. Bovik. Motion Tuned Spatio-Temporal Quality Assessment of Natural Videos. *IEEE Transactions on Image Processing*, 19(2):335–350, feb 2010.
  - [16] Krzysztof Templin, Piotr Didyk, Karol Myszkowski, and Hans-Peter Seidel. Emulating displays with continuously varying frame rates. *ACM Transactions on Graphics*, 35(4):1–11, jul 2016.
  - [17] Andrew B. Watson and JA Solomon. Model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A*, 14(9):2379–2391, 1997.
  - [18] Krzysztof Wolski, Daniele Giunchi, Nanyang Ye, Piotr Didyk, Karol Myszkowski, Radosław Mantiuk, Hans Peter Seidel, Anthony Steed, and Rafał K. Mantiuk. Dataset and metrics for predicting local visible differences. *ACM Transactions on Graphics*, 37(5), 2018.
  - [19] Nanyang Ye, Krzysztof Wolski, and Rafał K. Mantiuk. Predicting visible image differences under varying display brightness and viewing distance. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition 2019 (CVPR 2019)*, 2019.
  - [20] Roy S. Berns. Methods for characterizing CRT displays. *Displays*, 16(4 SPEC. ISS.):173–182, may 1996.

## Author Biography

*Gyorgy Denes is a PhD student at the University of Cambridge, having previously received his BA and MA degrees from this same institution. He has also worked as a Software Engineer at Microsoft, and as a research intern at Oculus (Facebook Reality Labs).*

*Rafał K. Mantiuk is a Reader in Computer Graphics and Displays at the University of Cambridge. Previously, he worked at Bangor University, the University of British Columbia and the Max-Planck-Institut (MPI) for Computer Science.*