https://doi.org/10.2352/EI.2022.34.16.AVM-108

A review of IEEE P2020 flicker metrics

Brian Deegan[1]
[1] ***Valeo Vision Systems, Tuam, Ireland***

Abstract

In this paper, we review the LED flicker metrics as defined by the IEEE P2020 working group. The goal of these metrics is to quantify the flicker behaviour of a camera system, to enable engineers to quantify flicker mitigation, and to identify and explore challenging flicker use cases and system limitations. In brief, Flicker Modulation Index quantifies the modulation of a flickering light source, and is particularly useful for quantifying banding effects in rolling shutter cameras. Flicker Detection Index quantifies the ability of a camera system to distinguish a flickering light source from the background signal level. Modulation Mitigation Probably quantifies the ability of a camera system to mitigate modulation of a flickering light source. This paper explores various use cases of flicker, how the IEEE P2020 metrics can be used to quantify camera system performance in these use cases, and discusses measurement and reporting considerations for lab based flicker assessment.

Introduction

Since 2016, the IEEE P2020 working group have been developing a number of new standards for automotive camera image quality assessment. This effort has included the development of new metrics and measurement procedures for LED flicker. The root cause and manifestations of flicker have been described extensively in the literature [1, 2, 3, 5]. In brief, flicker is an artifact observed in digital imaging, where a light source or region of an imaged scene appears to flicker (i.e. the light may appear to switch on and off, or modulate in terms of brightness or color), even though the light source appears constant when viewed by a human observer. As previously described [5], flicker can be an annoyance or distraction to a human observer, or it may impact the performance of Advanced Driver Assistance Systems (ADAS) algorithms (e.g. traffic light detection, electronic road signs etc).

P2020 Flicker metrics

A full description of P2020 flicker metrics has been defined previously [5, 7]. The following section provides a brief review and update of metrics proposed by P2020, which are due for prerelease publication in 2022.

The following P2020 metrics are calculated from a time series signal $s(n)$, which is calculated from the average of an ROI defined within the captured image test sequence.

Flicker Modulation Index (FMI):

$$\mathrm{FMI} = 100 \times \frac{X_{\max} - X_{\min}}{X_{\max} + X_{\min}} \tag{1}$$

where $X_{\max}$ is the maximum measured signal, $X_{\min}$ is the minimum measured signal of the PWM light source for the entire captured video sequence. FMI is essentially a Michelson Contrast based approach for measuring the magnitude of flicker within a time series $s(n)$. A lower number indicates less flicker in the output image.

Flicker Detection Index (FDI):

$$\mathrm{FDI} = \mathrm{Prob}\left(\frac{X_{\mathrm{meas}} - X_{\mathrm{ref,off}}}{X_{\mathrm{ref,off}}} \geq \tau\right) \tag{2}$$

where X_{meas} is the instantaneous measured flickering signal level, $X_{\mathrm{ref,off}}$ is the reference background light level or "off" light level, and τ is the flicker threshold i.e. the minimum defined acceptable Weber Contrast level. FDI is essentially a metric used to determine the likelihood that a flickering light source can be distinguishable from a background signal level. A higher number indicates better flicker mitigation, with a value of 1.0 indicating that in all frames measured, the flickering light source can be distinguished from the background/off light level. Note that the value of τ is defined by the user, as the required contrast level may vary, depending on the application. The value of τ is required to be reported as part of P2020 flicker test reports.

Modulation Mitigation Probability - reference ($\mathrm{MMP}_{\mathrm{reference}}$):

$$\begin{aligned}\mathrm{MMP}_{\mathrm{reference}} = \mathrm{Prob}\big((\overline{X_{\mathrm{ref,on}}} - \delta \cdot \overline{X_{\mathrm{ref,on}}}) \leq X_{\mathrm{meas}} \\ \leq (\overline{X_{\mathrm{ref,on}}} + \delta \cdot \overline{X_{\mathrm{ref,on}}})\big)\end{aligned} \tag{3}$$

where X_{meas} is the measured flickering signal level, $X_{\mathrm{ref,on}}$ is the reference expected light level, and δ is the defined acceptable threshold level. A higher number indicates better flicker mitigation, with a value of 1.0 indicating the light level was measured within a target threshold for all video frames measured.

Modulation Mitigation Probability - mean ($\mathrm{MMP}_{\mathrm{mean}}$):

$$\begin{aligned}\mathrm{MMP}_{\mathrm{mean}} = \mathrm{Prob}\big((\overline{X_{\mathrm{ref,on}}} - \delta \cdot \overline{X_{\mathrm{ref,on}}}) \leq X_{\mathrm{meas}} \\ \leq (\overline{X_{\mathrm{ref,on}}} + \delta \cdot \overline{X_{\mathrm{ref,on}}})\big)\end{aligned} \tag{4}$$

where X_{meas} is the measured flickering signal level, $X_{\mathrm{ref,on}}$ is the average signal level for $s(n)$, and δ is the defined acceptable threshold level. A higher number indicates better flicker mitigation, with a value of 1.0 indicating the light level was measured within a target threshold for all video frames measured.

Flicker Beat Frequency(FBF):

$$\mathrm{FBF} = \min[\mathit{mod}(f_{\mathrm{scene}}, f_{\mathrm{camera}}), \mathit{mod}(-f_{\mathrm{scene}}, f_{\mathrm{camera}})] \tag{5}$$

where f_{scene} is the frequency of the light source under test, and f_{camera} is the frequency of the camera (i.e. the frame rate). FBF is a measure of the beat frequency of the flickering light source when imaged by a camera system.

Flicker metrics and use cases

As previously highlighted, flicker has multiple manifestations, and the impact on image quality varies, depending on the use case. With this in mind, the P2020 working group have defined multiple flicker metrics in an attempt to fully characterize flicker. One of the key findings from the P2020 working group has been that no single metric can fully characterize the performance of a camera under test for a given lighting/flicker condition. However, by using a combination of P2020 flicker metrics, it is possible to sufficiently characterize a camera system for a given application. The following explores each of the P2020 flicker metrics and gives an overview of their strengths and limitations.

FMI is a relatively straight forward metric. It is based primarily on Michelson Contrast, measured over a flicker time series. It essentially reports the contrast of the flickering effects, and is particularly useful for measuring reflectance flicker phenomena (e.g. banding in the case of rolling shutter pixel architecture). However, FMI by itself does not quantify frequency of imaged flicker, which has been demonstrated previously to be highly correlated with perceived annoyance to human observers [8, 9]. A combination of FMI and FBF measurements may likely provide a better overall indication of the visual disturbance of flicker. Also, FMI does not indicate if a signal will be "detectable" for machine vision algorithms.

In contrast, FDI quantifies the likelihood that a flickering light is distinguishable from a background. FDI is largely intended for CV applications, e.g. traffic light detection. FDI only considers detectability versus background level – FDI does not quantify magnitude or frequency of flicker. Also, FDI as a single metric does not quantify the variation in detectablity over time. For example, in the case of a traffic light detection algorithm, a camera system may report an FDI of 0.5 (i.e. the traffic light is detectable for 50% of frames) over a 10 second interval. However, depending on the camera configuration and signal driving the traffic light, 0.5 may indicate every second frame is detectable, or it may be the case that the light is not detectable for 5 consecutive seconds. The implications for a traffic light detection algorithm are very different for each scenario. To address this, the P2020 working group have additional recommendations for reporting of FDI, which are explored later in this paper.

MMP quantifies how well a camera system mitigates magnitude of flicker. It is useful for both human viewing and machine vision applications. However, by itself it is not an indicator of signal detectivity. For CV applications, it should therefore be used in combination with FDI. MMP is also sensitive to signal saturation. This is explored further in this paper. The main difference between $\text{MMP}_{\text{reference}}$ and MMP_{mean} is the $\text{MMP}_{\text{reference}}$ quantifies modulation in relation to a target signal level, rather than the average signal level. For certain applications (e.g. traffic light recognition) this may be more beneficial, as it quantifies the likelihood that the traffic light will be reproduced "correctly". However, establishing the reference light level can be challenging from a test setup and measurement point of view.

MMP_{mean} in contrast, is relatively straight forward to measure. It is important to ensure sufficient phase sampling during testing to ensure accurate results. This topic will be explored in this paper. MMP_{mean} can report relatively low values even for LED Flicker Mitigating (LFM) sensors. This can occur if magnitude of modulation of the flickering light source is high. It is therefore necessary to carefully control and report the lighting configuration used during testing, to avoid reporting mis-leading results.

Flicker study - considerations for flicker measurement and reporting

As part of the development and validation of the P2020 flicker metrics, several challenges regarding test methodology, and interpretation and reporting of results have been raised and explored by the P2020 working group. This paper explorers several of these challenges.

Previous work by the P2020 working group has identified that the size of the measurement ROI and test duration affect P2020 flicker measurements [10]. The impact of measurement ROI and recording duration is explored further in this paper. As previously mentioned, MMP can be affected by signal saturation (e.g. if the pixel is overexposed). This paper demonstrates the impact of saturation on MMP, and includes recommendations for reporting when saturation occurs. These recommendations will be included in the upcoming P2020 release. Finally, this paper also explores variations within FDI measurements, and provides recommendations for reporting.

Test Methodology

Flicker simulator

As described in previous work [7], the test space for flicker is too large to feasibly validate P2020 flicker metrics in all measurement conditions using a physical test setup (i.e. with cameras and test charts). To that end, as part of the flicker metric validation efforts, a flicker simulator has been developed by the P2020 working group [7].The P2020 flicker simulator was used to generate the input data used in this study. The flicker simulator can simulate flicker scenarios with configurable light source wave-forms and intensities, as well as camera parameters, including exposure time, pixel saturation and sensitivities. The output from the flicker simulator is a video file of configurable length and resolution. Examples of output videos are shown in Figure 1.

Measurement ROI Size

To explore the impact of ROI size on flicker measurements, two exemplary scenarios are explored. In the first scenario, an ROI of 30×1 (width$\times$height) is used, and in the second scenario, an ROI of 30×300 is used (see Figure 2). The input flicker video was generated by using a 160 Hz, 45% duty cycle PWM illuminant, and a camera with a 30fps frame rate and 3ms exposure time. MMP_{mean} was then calculated from this video sequence, for both ROI sizes.

Measurement duration

Previous work by the P2020 working group [10] has demonstrated that in order to get accurate flicker measurements, sufficient phase coverage between the camera and flickering light source is required. In the absence of direct control over camera to light source phase, the best way to ensure sufficient phase coverage is to have a sufficiently long sampling period. A full exploration of the relationship between test duration and flicker metrics measurement is beyond the scope of this paper. However, an illustrative example of the impact of measurement duration is explored in this paper. Table 1 defines the study, PWM and sen-

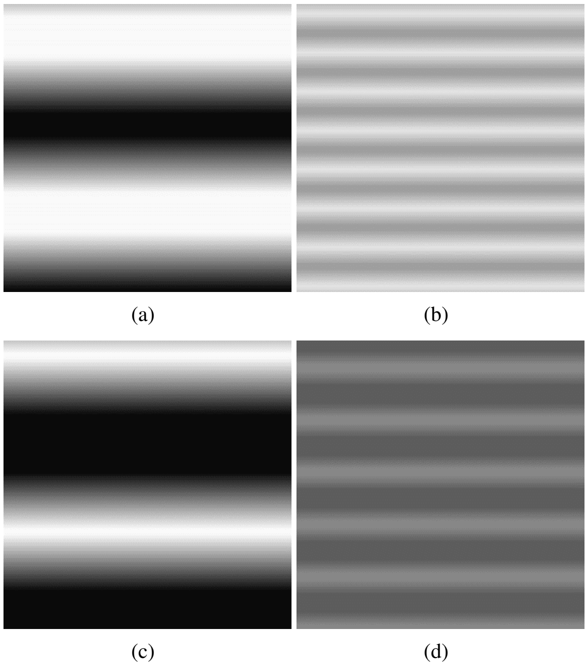

(a) (b) (c) (d)

Figure 1: Example outputs from P2020 flicker simulator. The flicker simulator allows generation of videos of arbitrary size, duration, frequency, duty cycle, exposure time etc.

sor model parameters used to explore the impact of measurement duration on MMP_{mean} results. Briefly, 60 seconds of input video was generated for multiple PWM frequency and camera exposure times. MMP_{mean} was then calculated for different measurement durations, and the stability of measurements assessed.

Impact of signal saturation

As mentioned previously, signal saturation can affect flicker metrics. To demonstrate this, two video sequences were generated with the flicker simulator. In both cases, a frequency of 125 Hz and duty cycle of 55% was used. For the sensor model, a frame rate of 30fps and exposure time of 5ms was used. In the first video, the offset and contrast were selected such that the simulated pixel would saturate. In the second video, lower values of offset and contrast were used, to ensure the pixel did not saturate. MMP_{mean} was measured for both video sequences, and the results compared.

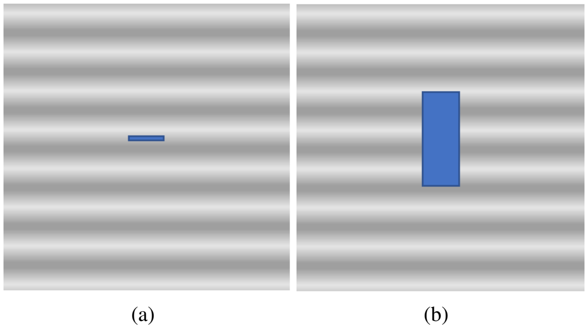

(a) (b)

Figure 2: (a) 30×1 ROI, (b) 30×300 ROI. MMP_{mean} is measured for both ROI sizes

Table 1: Measurement duration study parameters

Study model Parameters	
Parameter	**Description**
Frequency (Hz)	120.1, 130.1
Offset	0.1 a u
Contrast	0.5 a u
Phase offset (degrees) φ	0
PWM model	
Duty cycle (%)	20
Rise/Fall curve	Capacitance loading curve
Half load cycle HLC	0.001
Sensor model	
Exposure time (ms)	LFM mode: 11 ms. Non LFM mode: 3 ms
Measurement duration details	
Input video duration (s)	60 s
Measurement duration (s)	5 s, 10 s

Reporting of FDI

A full exploration of the variation in FDI measurement is beyond the scope of this work. However, to demonstrate that the distribution of FDI detection varies depending on the test scenario, three illustrate example scenarios were explored. For all three scenarios, a sensor model with a 30fps frame rate and 2 ms was used. In the first scenario, a PWM model with 121 Hz and 25% duty cycle was selected. In the second scenario, a PWM model with a 129 Hz and 25% duty cycle was selected. Finally, for the third scenario, a PWM model with a 135 Hz 25% duty cycle was selected. The impact on FDI detection for each scenario was then reported.

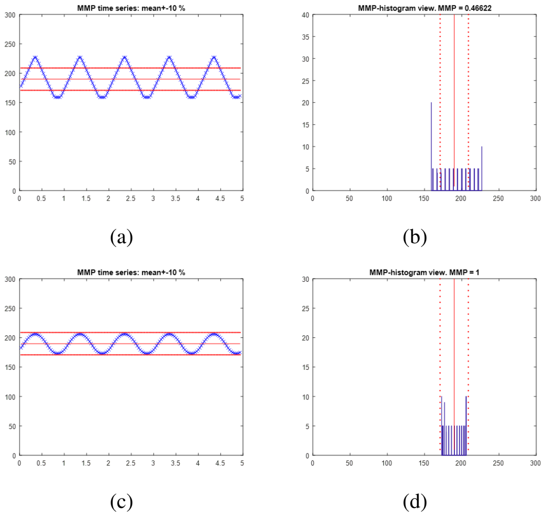
Results & Discussion

Impact of ROI size

The results of different ROI sizes on MMP_{mean} calculation is shown in Figure 3. In the case of a rolling shutter sensor model (as shown in this example), an ROI with a higher number of rows has the effect of smoothing the time series of the flicker signal. This has a direct impact on the MMP_{mean} score; with a 30×1 ROI, the reported MMP_{mean} is 0.46, whereas with a 30×300 ROI, the reported MMP_{mean} is 1.0. In the absence of other considerations, an ROI of a single row is generally recommended, as it avoids temporal smoothing effects and related underestimates in flicker metrics.

Impact of measurement duration

Figure 4 shows the result of measurement duration on MMP_{mean} calculation for a 120.1 Hz 25% duty cycle PWM signal and 3ms exposure time. For this test, MMP_{mean} is unstable when the duration of testing is less than 10 seconds. However, if the measurement duration is 10 seconds or longer, the result is quite stable This holds true even if the LED frequency is fairly close to the beat frequency. More testing is required, but if MMP_{mean}



(a) (b) (c) (d)

Figure 3: (a) MMP_{mean} time series and (b) MMP_{mean} histogram results calculated with an ROI of 30×1, (c) MMP_{mean} time series and (d) MMP_{mean} histogram results calculated with an ROI of 30×300

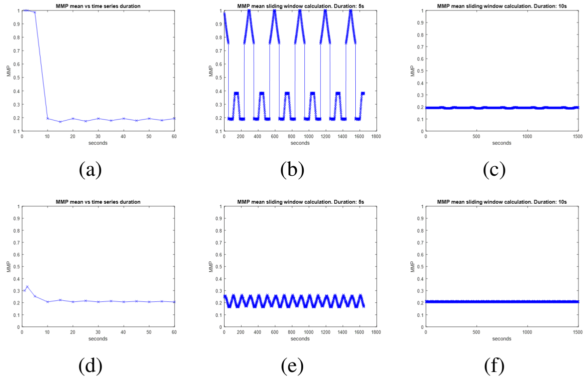
is calculated over 60 seconds or so, it looks like the result will be stable, unless the LED frequency is exactly a multiple of the frame rate (i.e. in this scenario, no temporal modulations in brightness will occur).

Figure 5 shows the results using the same PWM configuration as Figure 4, but with a sensor exposure time of 11ms. With an 11ms exposure time, at least one pulse of light is always captured. This was confirmed by reviewing the time series trace. So, one would expect MMP_{mean} to be closer to 1.0.

Reviewing the data in Figure 6, the root cause became clear: the background level was low, and the contrast of the LED (i.e. how bright the pulses are) is relatively high. As a result, there is a significant amount of modulation, even though there is flicker mitigation and pulses are not missed. In other words, flicker is mitigated, and MMP_{mean} is low, because there is a lot of modulation. So, the result is technically correct, but might be confusing. This scenario is likely to occur in dark background, low frequency, high intensity LEDs. Care should therefore be taken when interpreting the results from MMP_{mean}, as the degree of modulation of the light source can affect the MMP_{mean} result independently of camera configuration. It is therefore critical the a full description of the test lighting configuration be included in flicker test reports, to allow meaningful interpretation and comparison of MMP_{mean} results.

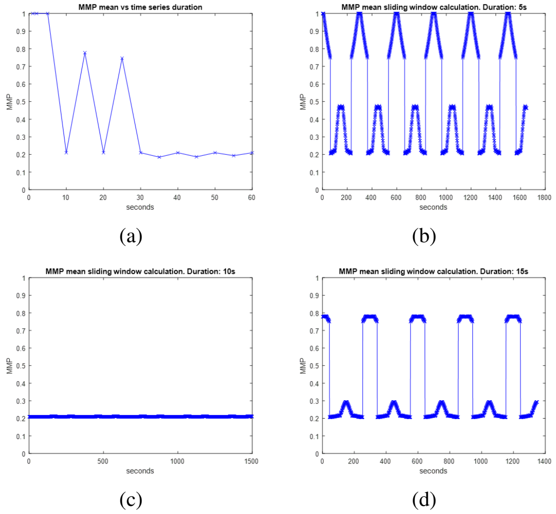
Impact of signal saturation

Figure 7 shows the impact of signal saturation on flicker metrics. If the flicker time series signal saturates, the modulation due to flicker is lower (somewhat similar to the way SNR changes close to pixel saturation). The measured MMP_{mean} in this scenario is 1.0. However, if the signal does not saturate, the reported MMP_{mean} is 0.338. This result highlights a concern regarding MMP_{mean} calculation and reporting. Take a use case where two sensors are being benchmarked. Both sensors have similar flicker mitigation characteristics, but one sensor has a higher dynamic range than the other. If care is not taken in reporting of results,



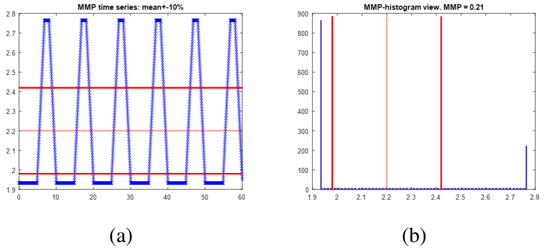
(a) (b) (c) (d) (e) (f)

Figure 4: (a) MMP_{mean} versus time series measurement duration (b) MMP_{mean} measurement, 5 second sliding window (c) MMP_{mean} measurement, 10 second sliding window. (d) MMP_{mean} versus time series measurement duration (e) MMP_{mean} measurement, 5 second sliding window (f) MMP_{mean} measurement, 10 second sliding window. PWM Model: 120.1 Hz, 20% duty cycle, Sensor Model: 30fps, 3ms exposure time



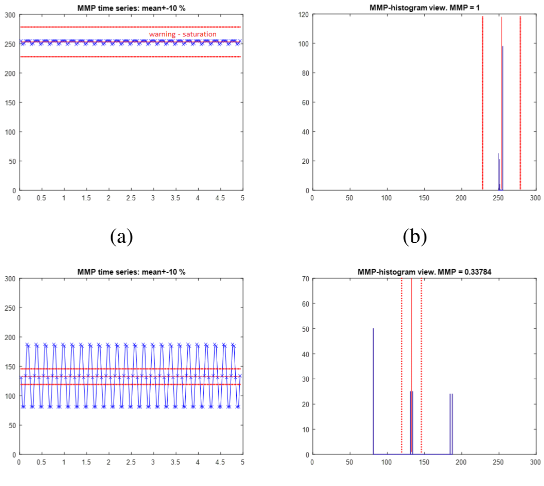
(a) (b) (c) (d)

Figure 5: (a) MMP_{mean} versus time series measurement duration (b) MMP_{mean} measurement, 5 second sliding window (c) MMP_{mean} measurement, 10 second sliding window. (d) MMP_{mean} measurement, 15 second sliding window. PWM Model: SI120.1Hz, 20% duty cycle, Sensor Model: 30fps, 11ms exposure time



(a) (b)

Figure 6: (a) MMP_{mean} time series (b) MMP_{mean} histogram. PWM Model: 120.1 Hz, 20% duty cycle, Sensor Model: 30fps, 11ms exposure time



(a) (b)

(c) (d)

Figure 7: (a) MMP_{mean} time series and (b) MMP_{mean} histogram results, signal saturated. (c) MMP_{mean} time series and (d) MMP_{mean} histogram results calculated, signal not saturated.PWM model: 125 Hz, 55% duty cycle. Sensor model: frame rate = 30fps, exposure time = 5ms.

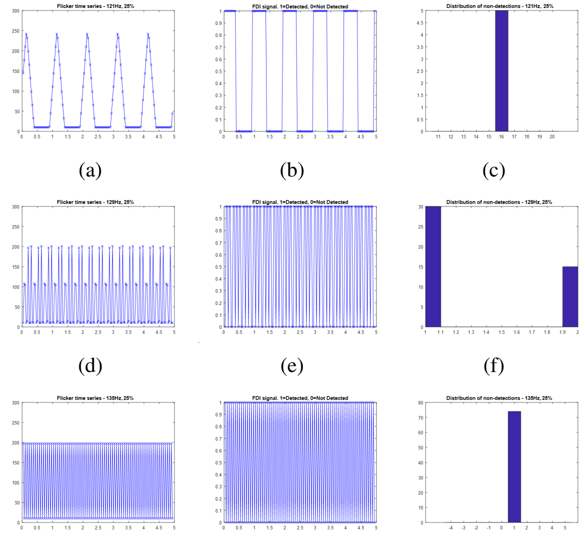
the sensor with the higher dynamic range may report a worse MMP_{mean} score. This is a misleading result and contrary to the intention of the MMP_{mean} metric. P2020 will therefore require that saturation be reported in test results when it occurs.

Reporting FDI results

The results of the FDI analysis are shown in Figure 8. For a sensor frame rate and 121 Hz frequency, the FPF will be 1 Hz. As a result, for a 5 second duration capture, there are 5 instances, each approx. 0.5 seconds where the FDI signal is zero (i.e. the light would not be distinguishable from the background). At 129 Hz, the FBF is 9 Hz. In this scenario, there are 30 instances where the FDI detection is zero for 1 consecutive frame, and 15 instances where the FDI detection is zero for 2 consecutive frames. Finally, at 135 Hz, the FPF is 15 Hz. In this scenario, all non-detections are of a single frame duration only (i.e. the light would be not detected every second frame). Note that in all three scenarios, the FDI for the entire 5 second test duration is quite similar (0.53, 0.4 and 0.5 for 121 Hz, 129 Hz and 135 Hz respectively). However, as previously mentioned, the impact for a machine vision algorithm may be quite different in each example. Depending on the algorithm implementation, it may be acceptable to capture the light on every second input frame, in which case the 135 Hz scenario may cause no issues. However, the 121 Hz scenario, where the light is missed for 0.5 second intervals may cause a significant problem. P2020 will therefore be recommending that the distribution of detections for FDI be included in test reports.

Conclusion

No single flicker metric proposed by P2020 gives a full picture of the characteristics of the camera under test for a given lighting configuration. However, by using a combination of the flicker metrics, particularly for a target application, a camera system can be sufficiently characterized. For accurate results, ensure metrics are calculated over a sufficiently long video sample, otherwise, measurements will be unreliable. At least one phase cycle is required to ensure a reasonable level of accuracy. Further work is required to fully determine the impact of measurement duration on measurement accuracy and precision. ROI size can affect metric calculation. It is generally recommended to use an ROI one single row high, to avoid temporal smoothing effects. Saturation of signal also affects metrics. Saturation is not always avoidable. When reporting flicker results, the P2020 standard will require that instances of signal saturation are reported in test results. From the results in this and previous publications, it is clear that great care is required during measurement and reporting of results to ensure that accurate flicker measurements are captured, and misleading results are avoided. To that end, the P2020 working group are planning to include a reporting template for flicker as part of the official P2020 release. This will contain many of the recommendations included in this paper, including ROI size, reporting of saturation/clipping, recording duration, and FDI distributions.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 8: (a) Time series (b) FDI detections and (c) FDI distribution, 121 Hz 25% duty cycle. (d) Time series (e) FDI detections and (f) FDI distribution, 129 Hz 25% duty cycle. (g) Time series (h) FDI detections and (i) FDI distribution, 135 Hz 25% duty cycle. Sensor model: frame rate = 30fps, exposure time = 2ms.

References

[1] Brian Deegan, LED flicker: Root cause, impact and measurement for automotive imaging applications, Electronic Imaging, vol.2018, pp.146-1, 2018.

[2] IEEE Working Group, IEEE P2020 Automotive Imaging White Paper, IEEE, 2018.

[3] Tarek Lule, Challenges of HDR imaging in Automotive Environment, AutoSens, Brussels, 2017.

[4] EN12966:2014, European Standard for Variable Message Traffic Signs, 2014.

[5] Brian Deegan, LED flicker measurement: Challenges, considerations, and updates from IEEE P2020 working group, Electronic Imaging, vol.2020, pp. 1-1-1-6(6), 2020

[6] Isao Takayanagi, Ken Miyauchi,Shunsuke Okura,Kazuya Mori, Junichi Nakamura and Shigetoshi Sugawa: A 120-ke$^-$ Full-Well Capacity 160-µV/e$^-$ Conversion Gain 2.8-µm Backside-Illuminated Pixel with a Lateral Overflow Integration Capacitor, MDPI, 17 December 2019

[7] Christian Wittpahl, Brian Deegan, Bob Black, Alexander Braun, An analytic-numerical image flicker study to test novel flicker metrics, Electronic Imageing, vol.2021, pp.183-1, 2021.

[8] Nicolai Behmann, Sousa Weddige, Holger Blume, Psychophysical Study of Human Visual Perception of Flicker Artifacts in Automotive Digital Mirror Replacement Systems, Journal of Perceptual Imaging, 2021.4.1.010401, 2021.

[9] Nicolai Behmann and Holger Blume, Psychophysics Study on LED Flicker Artefacts for Automotive Digital Mirror Replacement Systems, Electronic Imaging, vol.2020,pp234-1, 2020

[10] Robert C Sumner, Describing and Sampling the LED Flicker Signal, Electronic Imaging, vol.2020,pp.38-1-38, 2020

Author Biography

Brian Deegan received his Bachelor's Degree in Computer Engineering from the University of Limerick in 2004, an MSc In Biomedical Engineering from the University of Limerick in 2005 and PhD in Biomedical Engineering from the National University of Ireland, Galway in 2011. Since 2011 Brian has been working in Valeo Vision Systems as a Vision Research Engineer focusing on Image Quality. His main research focus is on high dynamic range imaging, LED flicker, topview harmonization algorithms, and the relationship between image quality and machine vision.