https://doi.org/10.2352/EI.2023.35.8.IQSP-309

Evaluation of image quality metrics designed for DRI tasks with automotive cameras

Valentine Klein, Theophanis Eleftheriou, Yiqi Li, Emilie Baudin, Claudio Greco, Laurent Chanas, Frédéric Guichard
DXOMARK, Boulogne-Billancourt, France

Abstract

Nowadays, cameras are widely used to detect potential obstacles for driving assistance. The safety challenges have pushed the automotive industry to develop a set of image quality metrics to measure the intrinsic camera performances and degradations. However more metrics are needed to correctly estimate computer vision algorithms performance, which depends on environmental conditions. In this article we consider several metrics that have been proposed in the literature: CDP, CSNR and FCR. We show a test protocol and promising results for the ability of these metrics to predict the performance of a reference computer vision algorithm that was chosen for the study.

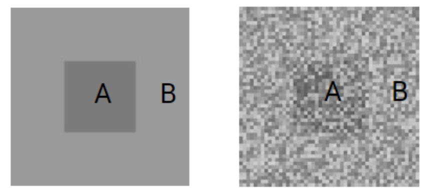
Introduction

Nowadays, cameras are widely used to detect potential obstacles for driving assistance. Image quality metrics are currently being standardized for automotive applications [1]. These metrics measure individual camera degradations intrinsic to the camera.

Computer vision algorithms performance is also linked to environmental conditions, for example fog can reduce the contrast. There is a need to develop and test metrics specific to computer vision algorithms. Two new metrics have been proposed and are under discussion in the IEEE-P2020 working group, Contrast Detection Probability (CDP) [2, 3] and Contrast Signal To Noise Ratio (CSNR) [4]. These two KPIs evaluate the camera performance in terms of object detection, based on the contrast probability distribution function.

While detecting objects is rather recent in the automotive world, it has existed for a long time for surveillance cameras and military applications. In these domains, the Jonhson criterion [5] is widely used to evaluate the range capacity of a camera, in the Detection Recognition Identification (DRI) approach. This criterion is based on resolution and environmental conditions. Detection means that the camera is able to detect that an object is present in the scene. Recognition means that the camera is able to find the class of object (is it a person or a car). Identification means that the camera can precisely identify the object (model of the car). A new metric, Frequency of Correct Resolution (FCR), has been proposed based on this DRI approach [6, 7].

In this article we evaluate the performance of the above listed KPIs for computer vision applications. We define a computer vision metric evaluation and validation protocol, in the same way as we perform perceptual analysis for human viewing applications. We choose a common computer vision application: license plate detection [8, 9]. The goal of the study is to determine whether each metric can predict the success rate of the license plate recognition algorithm on a given imaging system. This will allow us to compare the metrics and determine which ones are best suited for qualifing devices for computer vision applications.



(a) Ideal Case (b) With Noise

Figure 1. *Two gray patches, with and without noise.*

Computer vision metrics

CDP and CSNR

Contrast Probability Distribution Function

Contrast is an important parameter for object detection. It measures a luminance change between two regions of interest (ROI), relative to the average luminance. We consider two ideal uniform patches A and B, as presented on Figure 1(a). There are several ways to define the contrast between A and B in the literature. Two common definitions are Michelson Contrast and Weber Contrast:

- Weber Contrast is often used when detecting a small object A on a uniform background B. In this case, the average luminance of the scene is approximately the luminance of the background. Weber contrast can take very large values when x_B gets close to 0.
$$\text{Weber Contrast} = \frac{|x_A - x_B|}{x_B} \tag{1}$$
- Michelson Contrast is used when comparing two objects of the approximate same size. Unlike Weber Contrast, Michelson contrast is symmetrical:
$$\text{Michelson Contrast} = \frac{|x_A - x_B|}{x_A + x_B} \tag{2}$$

On real images, due to noise or other effects, pixels corresponding to a given luminance value in the scene do not all have the same gray level value. As seen in Figure 1(b), noise makes it harder to distinguish between the two patches compared to Figure 1(a), even though the average value has not changed. Given this, instead of a single contrast value between the two patches, we can compute a different contrast value for each pairs of pixels. Thus we define the contrast Probability Density Function (PDF)(Figure 2). The PDF can be computed using either Michelson or Weber contrast.

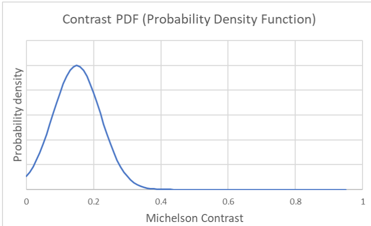


Figure 2. *Contrast Probability Density Function (PDF) computed from two gray patches with noise*

CDP

Contrast Detection Probability (CDP) quantifies the ability of a camera to preserve accurately an input contrast in an image. Mathematically, CDP is defined as the probability that the measured contrast falls within a confidence interval centered around the nominal input contrast:

$$\mathrm{CDP} = \Pr\left[C_{\mathrm{in}} \cdot (1-\delta_-) \leq C_{\mathrm{meas}} \leq C_{\mathrm{in}} \cdot (1+\delta_+)\right] \tag{3}$$

where C_{in} and C_{meas} are respectively the input contrast in the scene and the measured contrast in the image. And δ_- and δ_+ are the parameters defining the lower and upper bounds of the confidence interval presented in Figure 3, δ_- and δ_+ can be the same.

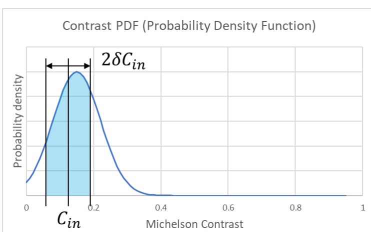


Figure 3. *Confidence interval of the contrast PDF*

CSNR

Contrast Signal-to-Noise Ratio (CSNR) quantifies the ability of a camera to distinguish two objects, or distinguish an object from its background. Its computation is similar to that of SNR and is based on the measured contrast probability statistics:

$$\mathrm{CSNR} = \frac{\overline{C}}{\sigma_C} \tag{4}$$

where $\overline{C}$ and σ_C are respectively the mean and the standard deviation of the contrast between two patches as shown in Figure 4.

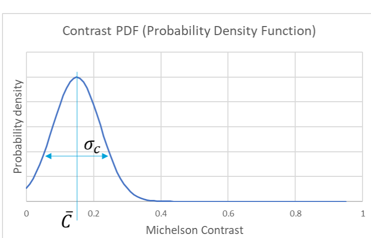


Figure 4. *Mean and standard deviation of the Contrast PDF*

Theoretical observations of CDP and CSNR

CDP and CSNR are two quality metrics, based on contrast and noise. CDP considers that contrast fidelity is based on the contrast of the image being close to the contrast on the scene, within a confidence interval. CSNR assumes that object separability is based on a low level of noise compared to the contrast in the image.

Since CDP is a probability value, it saturates at unity. CSNR does not saturate and can provide information when patches are close to saturation. However CSNR is not defined when two patches are both fully saturated. Unlike CDP, CSNR is not defined or can take very large values when the standard deviation of the contrast PDF is close to zero. If we consider a patch that is saturated to the dark level, and a patch that is saturated to the white level, the contrast between the two patches is large and they can easily be separated by a computer vision detection task. However CSNR cannot be measured between these two patches.

CSNR describes noise processes using the mean and the standard deviation that are a synthetic description of the contrast PDF. The contrast PDF may not always be Gaussian, in this case the mean and standard deviation are not sufficient to describe accurately the noise process. CDP uses directly the entire contrast PDF in order to use the real dispersion due to noise. The bounds of the confidence interval should be fixed for CDP by the user and take into account the tolerance of a given application of the camera under test. No confidence interval on contrast PDF is needed for the CSNR computation.

CDP compares the measured contrast to the input contrast. Therefore performing image linearization is mandatory before the metric computation, to inverse all non-linear operations performed by the ISP. CSNR can be computed without image linearization since CSNR does not evaluate the fidelity of the reproduced contrast to the input contrast. Thus, CSNR can be used to test ISP performance.

FCR

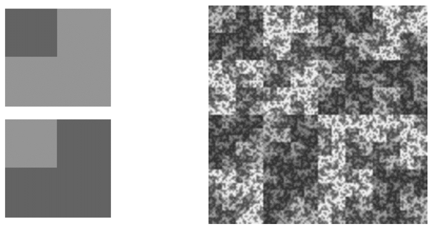
CSNR and CDP do not include the frequential aspect of the detection problematic, thus we introduce Frequency of Correct Resolution (FCR) proposed by Landeau [6, 7] that examines the contrast degradation as a function of the size of the object. FCR reflects the ability of the camera to reproduce details and textures with fidelity.

FCR computation is based on a multi-scale fractal test target shown in Figure 5(b). Each scale represents a spatial frequency and is made of randomly rotated Corner-Point patterns shown in Figure 5(a). We can distinguish two types of Corner-Point with four possible orientations for each type.

The goal of the FCR computation is to find the maximum scale correctly recorded by the camera. To do so, we compare scale by scale the test image to the ground truth. For each scale, a score called Probability of Correct Response (PCR) is computed to quantify the similarity of the recorded Corner-Point patterns with the reference ones:

$$\mathrm{PCR}(s) = \frac{1}{2^s} \cdot \sum_{i=0}^{2^s-1} \mathrm{score}(i) \tag{5}$$

where s is the scale and *score* is a function that outputs in the set $\{-1, -1/3, 1/3, 1\}$ depending of the level of resemblance of the



(a) 2 types of Corner-Point pattern (b) Fractal chart

Figure 5. *Multi-scale fractal chart generation from randomly rotated patterns*

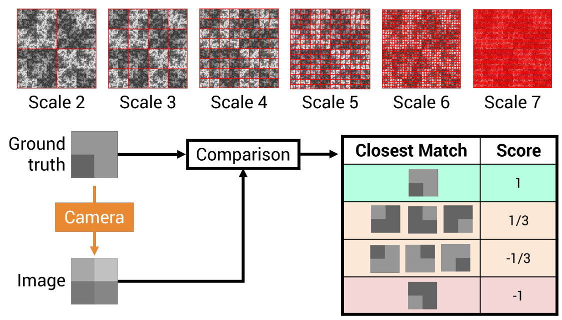


Figure 6. *PCR score computation*

test patterns with the reference ones (see Figure 6). We notice that a PCR score of 100% is attributed to a scale if all Corner-Point patterns of the considered scale are correctly recorded by the camera.

For each scale s from 2 to the maximum scale, we can define a corresponding spatial frequency in cycles per pixel, depending on the total width W of the chart in the image:

$$f = \frac{2^{(s-1)}}{W}, \text{ cy/pix} \tag{6}$$

The multi-scale analysis leads to a PCR curve in Figure 7 that represents the ability of the camera to preserve the input contrast as a function of spatial frequencies. From the PCR curve, we can define the Frequency of Correct Resolution (FCR) (see Figure 7) that is a spatial frequency characterizing the resolution limit of a camera in terms of contrast detection, with respect to a detection threshold. A PCR threshold of 50% was chosen by the original author, so FCR is the maximum frequency that gives a PCR of 50%:

$$\text{FCR} = \max\{f | PCR(f) \geq 50\%\} \tag{7}$$

Reference Computer Vision Application

If we want to validate a full system with a camera and a computer vision algorithm, and find its limits, a very large number of images of natural scenes is required. When testing the same algorithm with a different camera, the whole work would have to be repeated. The goal of our study is to find an easier way to validate this kind of systems. We want to link the performance of computer vision algorithms to metrics that are easy to measure from a single target chart in a laboratory environment.

To do so, we have to select a reference CV algorithm that represents the final DRI task, then study the correlation between

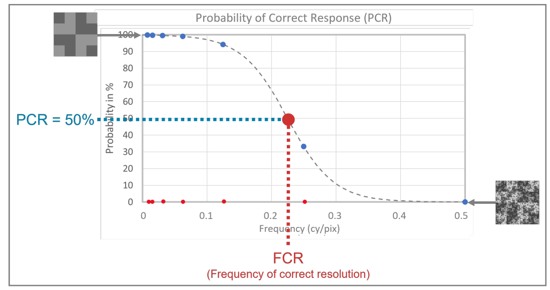


Figure 7. *PCR curve with the defintion of FCR.*

the metrics and the performance of the full system (camera and computer vision algorithm). The algorithm should be generic, commonly used, public and free to use, easy to shoot in laboratory, and with a well defined ground truth.

We chose an Automatic License Plate Recognition (ALPR) algorithm proposed by Silva *et al.* [8, 9]. It is a modern and common application, using neural networks, and easy to test in a laboratory, with images of license plates. The complete ALPR pipeline includes several steps: car detection, license plate detection and Optical Character Recognition (OCR). The final purpose of this study is to discriminate among imaging systems based on their performance on computer vision algorithms. Now, the task of detecting and isolating the license plate in the scene is in general a function more of the shooting conditions (distance of the plate from the camera, perspective, illumination, etc.) and not very discriminating to qualify the imaging system itself. We therefore decided to focus our attention on the more discriminating aspect of the pipeline, i.e., reading the characters on the plate. We assume that the position of the license plate in the image is known, and only focus on the OCR.

The OCR algorithm uses a modified YOLO network. Since YOLO is aimed to detect any kind of object, it had to be tuned by the authors [9] to adapt it to license plate (LP) recognition use cases. In particular, the input and output aspect ratio and granularity are adjusted to work with license plates. It has been trained on license plates from different regions around the world (Brazil, Europe, United States, and Taiwan). In this article we will use simple license plates following the French license plates format, with 7 characters in total (two letters, three digits, two letters). An example can be seen on Figure 9

The OCR algorithm returns the list of detected characters from an input image of a license plate. For our application, we consider that a license plate has been correctly recognized if all characters from the input image have been correctly recognized. We use the recognition rate as the indicator of performance of the OCR, defined for a set of different license plates as:

$$\text{Recognition Rate (\%)} = \frac{\text{Number of recognized LPs}}{\text{Total number of tested LPs}} \tag{8}$$

Experimental analysis

The goal of this experimental analysis is to compare the metrics previously defined (CDP, CSNR, FCR), and see which one can better show the limits of the tested camera from the point of view of the chosen computer vision algorithm. This will allow us to know which metric is best at predicting the performance of a full system, with a camera and a computer vision algorithm.

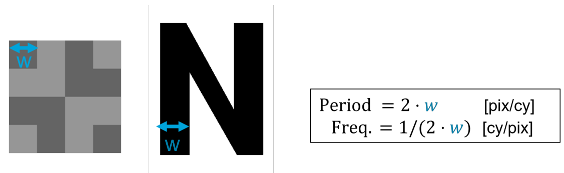


Figure 8. *Definition of period and frequency*

(a) Grayscale used for CDP and CSNR

(b) License plates with different luminance and contrast

Figure 9. *Example of images used for CDP, CSNR and OCR, with SNR max 28 dB.*

Simulation framework

This analysis requires to capture a very large number of images. For this reason, we have preferred to work only in simulation. Indeed, the computer vision algorithm needs to run on several images with different characters to get reliable statistics on results. Since we also need to test different resolutions, luminance levels and contrasts, this means that a very large number of images is required to test a single camera.

We have used a simple 8 bit single channel camera pipeline to generate different quality of images, applied in the following order:

1. **Framing:** Different size of the target in the image.
2. **Exposure:** Fixed exposure for all images.
3. **Lens:** MTF blur, with different cut-off frequencies, corresponding to different qualities of cameras, or different shooting conditions.
4. **Sensor**: White Gaussian Noise with different SNR curves, corresponding to different qualities of cameras, or different shooting conditions. In the results, the different noise curves are represented with their maximum SNR value.
5. **ISP:** Denoise (Gaussian kernel) and sharpen (unsharp masking).

The size of the target in the image can be defined in terms of spatial period. We define the period as the size of an FCR Corner-Point pattern at the considered scale, or twice the size of a line of an OCR character (see Figure 8). Note that CDP and CSNR are statistical measurements on nominally uniform gray patches, they are not impacted by resolution and MTF blur. Therefore for these measures the "Framing" and "Lens" steps have been omitted.

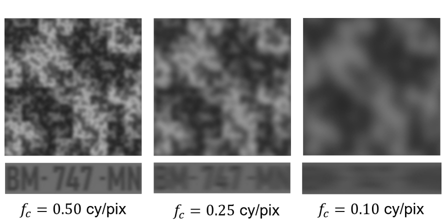


Figure 10. *FCR chart and license plates, with different MTF applied*

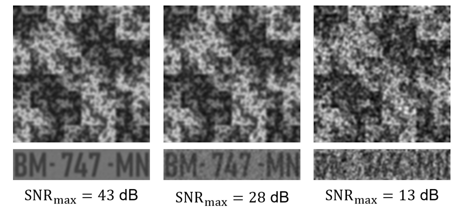
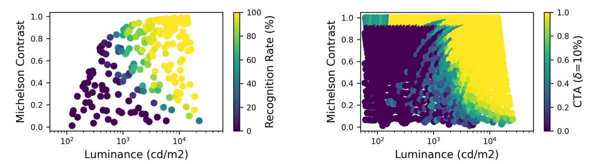
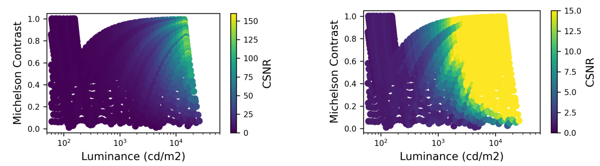


Figure 11. *FCR chart and license plates, with different noise levels applied*



(a) OCR (period = 8 pix/cy)

(b) CDP, with a 10% threshold



(c) CSNR

(d) Thresholded CSNR (threshold=15)

Figure 12. *Results of CDP, CSNR and recognition rate for a camera with maximum SNR 43dB*

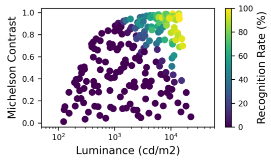
CDP and CSNR are measured on a grayscale chart with a constant Michelson Contrast of 6% between patches, and luminance from 50 to 30000 cd/m^2 which is enough to cover the whole dynamic range of the simulated 8 bit sensor. By taking patches two by two on this chart, we get 3000 different combinations of luminance and contrast.

To make a comparable database of license plate images, we had to generate license plates with different contrast between the background and the characters, and different average luminance. Thirty different license plates were used for each conditions, and 230 different combination of luminance and contrast have been used to compare with CDP and CSNR. This leads to a total of 7000 images of license plates needed to compare to a single image of grayscale chart. Examples of generated grayscale chart and license plates with different luminance and contrast can be seen on Figure 9.

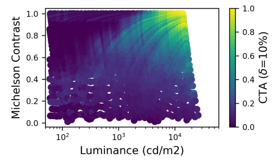
We have generated an FCR chart with a contrast value of 15%, and have generated a database of license plates with the same contrast. Figures 10 and 11 show examples of generated FCR charts and license plates with different image degradation.

Predicting the performance of OCR recognition rate with contrast based metrics (CDP and CSNR)

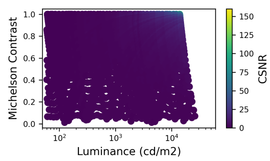
We used our simulation framework to determine whether CDP and CSNR can predict the performance of the reference computer vision algorithm. The results for the reference computer vision algorithm are recognition rate of the OCR. The results for CDP are computed with $\delta_+ = \delta_- = 0.1$ (see Equation 3). CSNR is presented both with original and thresholded values. Figure 12



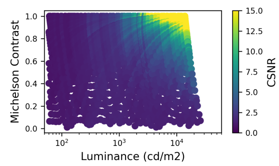
(a) OCR (period = 8 pix/cy)



(b) CDP, with a 10% threshold

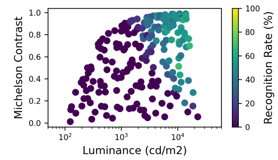


(c) CSNR

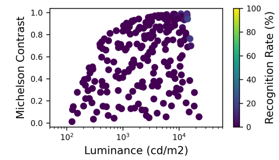


(d) Thresholded CSNR (threshold=15)

Figure 13. *Results of CDP, CSNR and recognition rate for a camera with maximum SNR 13dB*

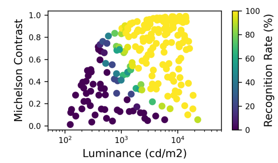


(a) SNR 43 dB

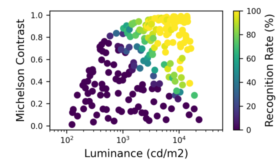


(b) SNR 13 dB

Figure 14. *OCR recognition rate maps, period 4 pix/cy*



(a) SNR 43 dB



(b) SNR 13 dB

Figure 15. *OCR recognition rate maps, period 16 pix/cy*

presents performance maps for each KPI, for a camera with high maximum SNR (43dB). Each dot on the map represents a probability of detection for a given luminance and contrast.

CDP, CSNR and OCR recogition rate show a similar behavior. As expected, they all show good performance for high contrast and high luminance values, and drop to zero for low contrast and luminance values. We can see some differences between the CDP and CSNR. CDP saturates to 1, it does not show the performance of the camera above a certain limit. Instead, it shows a large contrast and luminance area for which the camera behaves well. On the contrary CSNR does not saturate and highlights smaller areas that correspond to the very best behavior of the camera. The OCR recognition rate also saturates to 100% on a large area, therefore it looks more similar to CDP than CSNR, and we can say that CDP is better here at predicting the performance of the computer vision algorithm. Jenkin [4] has shown that when applying a threshold on CSNR, we can obtain an approximation of CDP. This is what we can see here, thresholded CSNR looks good at predicting the performance of the computer vision algorithm. However this has not been demonstrated theoretically and there is no easy way to choose a threshold.

Figure 13 presents results of the same metrics, with a camera with a lower maximum SNR (13dB). As expected, all metrics have worse performance for lower SNR values. All previous observations remain valid for the lower SNR.

Note that the OCR results presented here have been computed with a period of 8 pix/cy for the characters (see Figure 8 for the definition of period). With this size of characters there is 100% recognition rate for a large contrast and luminance area at maximum SNR. We have also tested other size of characters. In Figured 14 and 15 we can see that the behavior of the computer vision algorithm depends a lot on the size of characters, and does not follow as well the behavior of CDP and thresholded CSNR for all sizes of characters.

To make a more quantitative comparison, we can use the Pearson Correlation Coefficient [10], defined for two random variables X and Y as:

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} \tag{9}$$

Figure 16 presents the Pearson Correlation Coefficient between CDP and OCR recognition rate, and thresholded CSNR and OCR recognition rate, as a function of SNR values, for different size of license plate characters. We can first look at the results for a period of 8 or 16 pix/cy, since it is not too challenging for the OCR algorithm. We can see that all correlation values are above 80%, which means that CDP and thresholded CSNR have a good correlation with the OCR recognition rate. With a period of 4 pix/cy and low SNR, the correlation values are lower. This can be expected from the fact that this is a more difficult task for the OCR, due to the small size of the license plate characters.

CSNR allows to identify a small area with the best camera performance. If we know precisely the luminance and contrast of the objects we want to detect, CSNR is a good measure to know if a camera will work for this application. However in real scenes the luminance and contrast values are diverse. CDP and thresholded CSNR allow to identify a wide area where the camera can perform good for most applications, which seems more useful in practice.

Predicting the performance of OCR recognition rate with FCR

As seen in the previous section, resolution cannot be ignored when looking at the performance of a computer vision algorithm.

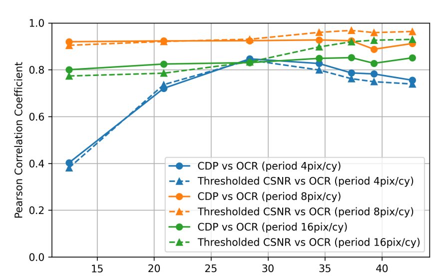
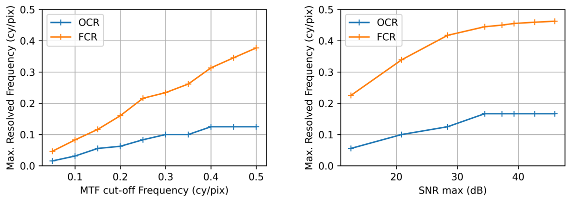
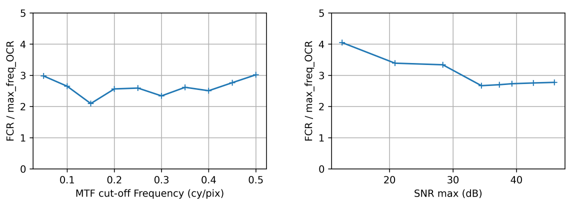


Figure 16. *Pearson Correlation Coefficient (9) between CDP and OCR recognition rate, and CSNR and OCR recognition rate, as a function of SNR values, for different size of license plate characters.*



(a) Variation with blur (b) Variation with noise

Figure 17. *Maximum resolved frequency for OCR and FCR, for different blur and noise levels*



(a) Variation with blur (b) Variation with noise

Figure 18. *Maximum resolved frequency for OCR and FCR, for different blur and noise levels*

We have defined the frequency of license plate characters in the same way as we have defined frequency for FCR (Figure 8). To have a metric comparable to FCR, we consider the maximum frequency where 50% of the license plates of the database can be correctly recognised. Figure 17 shows a comparison of the maximum resolved frequency both for FCR and OCR, for different noise and blur levels. Smaller maximum frequency means worse performance. The luminance and contrast is the same for the FCR chart and for the license plates.

As expected, both KPIs have better performance for lower noise and lower blur values. The FCR performs better than the OCR. This can be explained by the fact that in general several periods are needed to detect a character, whereas a single period is necessary to detect an FCR pattern. When looking at the ratio of the two metrics (Figure 18) we can see that the ratio is between 2 and 4, which seems a reasonable ratio when comparing the total size of a character to its line width. This shows good correlation between FCR and maximum resolved frequency of the OCR, meaning that FCR can predict the results of the reference computer vision algorithm.

Conclusion and future work

In this article we have presented a simulation environment that allowed us to test and compare CDP, CSNR and FCR, and see which one can best predict the performance of a full system, with a camera and a computer vision algorithm.

We have seen that CDP and CSNR show similar results, and the same limitations, though with small differences. CSNR has been linked to detection theory [4] and our results show that this metric can highlight the very best performance area of a camera, which can be useful for some difficult computer vision tasks. CDP is more empirical but it has shown to better follow the behavior of the tested computer vision task. CSNR can be thresholded to produce results similar to CDP. CDP seems easier to interpret because the CDP threshold used is based on the contrast in the scene. However, little study has been done on the choice of thresholds for CDP or CSNR.

Our results also show that frequency is important when evaluating a computer vision algorithm. For this reason we have started to test FCR, that include both contrast and frequency. We still have limited results but they show a good correlation with the reference computer vision algorithm.

The next step will be switching from simulation to real laboratory testing for the comparison work using several cameras. Many challenges are identified. Firstly, the FCR computation calls for a high precision image registration algorithm of the test image with the reference one. Secondly, an optimized test protocol needing a minimum number of images should be defined for the comparison, since many images of license plates are required to generate significant statistics for the reference computer vision.

Then the ultimate goal is to design a new set of metrics. On the one hand, CDP and CSNR are dedicated to study the contrast reproduction over the entire dynamic range of the camera, and on the other hand, FCR has the advantage to provide the ability of a camera to distinguish fine details and preserve texture. So a new metric seems to be necessary to gather all important aspects to evaluate the capability of a camera to perform a DRI task.

References

[1] "IEEE Draft Standard for Automotive System Image Quality," in IEEE P2020/D3, vol., no., pp.1-284, 1 (2022).

[2] M. Geese; U. Seger; A. Paolillo. "Detection Probabilities: Performance Prediction for Sensors of Autonomous Vehicles". Proc. IS&T Int'l. Symp. on Electronic Imaging: Autonomous Vehicles and Machines (2018).

[3] U. Artmann; M. Geese; M. Gäde. "Contrast Detection Probability – Implementation and use cases". Proc. IS&T Int'l. Symp. on Electronic Imaging: Autonomous Vehicles and Machines (2019).

[4] R. Jenkin. "Contrast Signal to Noise Ratio". Proc. IS&T Int'l. Symp. on Electronic Imaging: Autonomous Vehicles and Machines, pp 186-1 - 186-6 (2021).

[5] How to Assess Thermal Camera Range Capability for Site Design Purposes, Leonardo DRS, Electro-Optical & Infrared Systems (2014).

[6] S. Landeau; G. Swiathy. "Multi-level fractal test target: Corner-Point criterion for evaluating visible and digital fusion imagers". Proc. SPIE 11537, Electro-Optical and Infrared Systems: Technology and Applications XVII, 115370C (2020)

[7] S. Landeau. Fast Corner-Point Image Transformation: Applications to the Evaluation of Turbulence Mutigation Algorithms and Digital Fusion Systems (2022).

[8] Silva, S.M., Jung, C.R. "License Plate Detection and Recognition in Unconstrained Scenarios". In: 2018 European Conference on Computer Vision (ECCV). pp. 580–596 (2018). https://doi.org/10.1007/978-3-030-01258-8_36

[9] Silva, S.M., Jung, C.R. "Real-time Brazilian license plate detection and recognition using deep convolutional neural networks". In: 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images. pp. 55–62 (2017). https://doi.org/10.1109/SIBGRAPI.2017.14

[10] Pearson correlation coefficient, Wikipedia (2023). https://en.wikipedia.org/wiki/Pearson_correlation_coefficient