Assessment of image sensor performance with statistical perception performance analysis

Stefan Franz[a], Dieter Willersinn[b], Kristian Kroschel[b]

[a]Daimler AG, Ulm, Germany;
[b]Fraunhofer IITB, Karlsruhe, Germany

ABSTRACT

The performance of perceptive systems depends on a large number of factors. The practical problem during development is, that this dependency is very often not explicitly known. In this contribution we address this problem and present an approach to evaluate perception performance, as a function of e.g. quality of the sensor data. The approach is to use standardized quality metrics for imaging sensors, and to relate them to the observed performance of the environment perception. During our experiments, several imaging setups were analyzed. The output of each setup is processed offline to track down performance differences with respect to the quality of sensor data. We show how and to what extend the measurement of the Modulation Transfer Function (MTF) using standardized tests can be applied to evaluate the performance of imaging systems. The influence of the MTF on the signal-to-noise ratio can be used to evaluate the performance on a recognition task. We assess the measured performance by processing the data of different, simultaneously recorded imaging setups for the task of lane recognition.

Keywords: performance evaluation, perceptive components, sensor evaluation

1. INTRODUCTION

Current driver assistance systems have only various warning methods or supporting actuation components to improve reaction times in critical situations. They never override actions of the driver. Future systems will reach one step further, by steering the vehicle autonomously and initiating emergency breaking to prevent critical situations. In order to do so, the performance and reliability of the perceptive components of such systems need to be improved, as well as the testing and measuring methods for these components. The improvement and optimization of a system component while the overall performance of a system is not affected negatively can only be reached,if the performance-critical parameters are known and their effect is well-understood . The problem is, that there is no common methodology or standard available, how to evaluate and compare perceptive components on a mathematical basis or which include tools to assess their performance, depending on functional parameters and data quality. Most tests are rather focused on the overall system performance, than on evaluating dependencies of the involved components. In the field of sensors, different associations[3, 6] have proposed concepts for standardized tests and measurements for sensor performance evaluation. What is still missing, are models that describe the relationship between sensor performance and the performance of the processing components.

We consider perceptive components, including sensors and processing units, which are responsible for acquiring and processing data, as well as extracting features of the environment. In combination with corresponding computational environment models, object generation and situation analysis is performed, to pose the final perception result (Fig. 1).

This approach helps with the identification of critical parameters, that have a direct influence on perception performance. With a concentrated attention on these parameters, a fast and efficient identification of performance breakdown points can be achieved, together with the possibility to balance the algorithm requirements and the sensor performance.[1]

Further author information: (Send correspondence to Stefan Franz)
S.F.: E-mail: stefan.franz@daimler.com
D.W.: E-mail: dieter.willersinn@iitb.fraunhofer.de
K.K.: E-Mail: kristian.kroschel@iitb-extern.fraunhofer.de

Intelligent Robots and Computer Vision XXVII: Algorithms and Techniques, edited by
David P. Casasent, Ernest L. Hall, Juha Röning, Proc. of SPIE-IS&T Electronic Imaging,
SPIE Vol. 7539, 75390I © 2010 SPIE-IS&T · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.838969

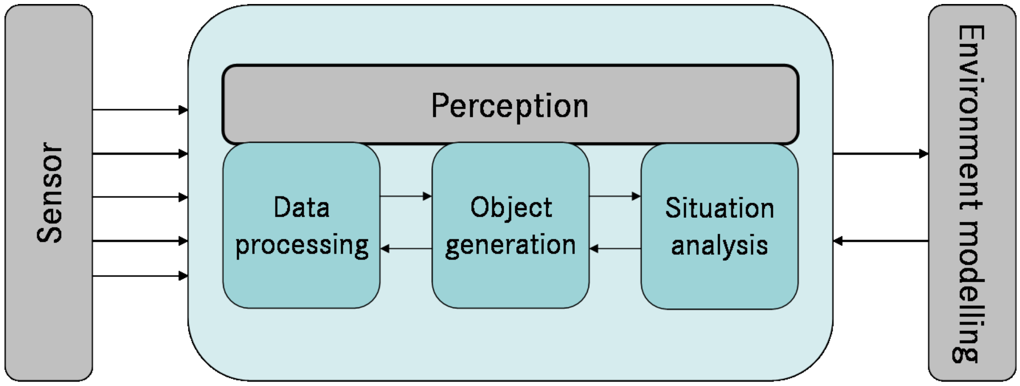

Figure 1. Structural layout of the perceptive components of a driver assistance system

This paper presents an approach to perception performance evaluation using available standards to measure and quantify the performance of imaging components and assess these results by computing a statistical measure for a given perception task.

2. DATA QUALITY OF IMAGING SENSORS

The identification of the set of parameters, that relate to data quality of a sensor setup, is the first step for this task. The focus of our work is in the field of camera based systems. For an imaging system, the reasonable characterization would be image quality. Its performance is perceived, as the accuracy to represent scene detail in the resulting image data. The level of detail or the unambiguousness and distinguishability of intensity and contrast, is designated to be expressed in terms of sharpness and resolution. Intensity and contrast in a digital image decrease, as the distance between 2 contrast edges decreases, or in terms of signal analysis, the spatial frequency increases. The capability of a system to capture the information content of an object as a function of spatial frequencies is given by the MTF of the system. High spatial frequencies correspond to fine image detail. Therefore, the more detailed an object and the lower the MTF of a system at high frequencies, the lower the ability to resolve these details in the image. From the MTF the limit of resolution can be determined. This limit depends on the characteristics of all the involved optoelectronic components. The system MTF integrates the modulations occurring in all the elements affecting the digital image.

$$MTF_{sys} = MTF_{lens} \cdot MTF_{sensor} \cdot MTF_{noise} \tag{1}$$

Image quality is one of the most important characteristics for all image sensing systems and the MTF is a common metric, used to quantify it. Nowadays the MTF is widely used throughout the optical industry as the objective way to clearly represent and evaluate the performance of optical systems.

Several measurement methods have been described, based on periodic signals, random noise, and other inputs. Edge-gradient methods provide the advantage of simple measurement targets.[2] This way the measurement process is detached from cost and time intensive measurement procedures, that need fixed and complex environment conditions. The ISO12233 methodology[3] has been established in order to provide a fast MTF measurement method based on only one image. In such a standardized way, the MTF data from various digital input devices, may be easily reproduced and compared. The MTF, in general, is a graph of the intensity measure in gray level percentage versus spatial frequency, resulting in a map of image intensities for various frequencies.

Spatial frequency is typically measured in lp/mm, describing the ability to discern the number of contrast pairs of lines, appearing within a millimeter. For digital cameras the normalized unit cycles per pixel (c/p) is more appropriate to account for the variety of sensor sizes. High spatial frequencies correspond to fine image detail. The more extended the response, the finer the detail and in the end, the sharper the image. Hertel and Chang[4] propose a sensor performance and image quality measure that takes the wide range of influences of automotive video footage into account to achieve a minimum quality criterion for camera based active safety systems. An overview of the state of the art in image quality assessment is given by Angelis.[5] Manufacturers of cameras for machine vision are also aware of their customers demanding a common framework and criteria to describe, test and compare products resulting in the proposal of the EMVA Standard 1288[6] which is already widely accepted. Using this framework several other topics for the evaluation can be covered. Our ambition is to study the effects of the performance measurement results of an imaging component on a perception task in real world scenarios, which has not been evaluated in detail yet.

2.1 Sensor Performance Measurement

The chosen procedure for the sensor performance measurement follows the Slant Edge Feature measurement,[3] which provides a fast and robust MTF estimation based on gradient edge evaluation of one image. The process itself tolerates a small misalignment of the target and is robust to different contrast ratios, as long as the camera's exposure limit is not reached.[7] The alignment of the camera can be measured by hand and the correct centering was checked with the help of a cross mark in the center of the image. To achieve a stable illumination level for all measurements and to prevent saturation, the histogram of the scene was used as input to control the imager. Once assured that all parameters met the requirements, they were fixed for the complete series of measurements. Image enhancement features of the sensors, like sharpening edges, were deactivated to provide unbiased results. No compression or other conversions were applied to the data. The capturing requirements for the measuring target were maintained using standard components for lighting and fixation and assured similar conditions for later measurements.

sensor	width	height	pixel size	$f_{Nyquist}$
1	$640px$	$480px$	$7.4\mu m^2$	$68lp/mm$
2	$1024px$	$768px$	$4.6\mu m^2$	$108lp/mm$

Table 1. Parameters of the used grayscale imaging devices

The used cameras have grayscale imagers with different resolution and pixel sizes (Table 1). Each camera was equipped with different lenses by turn and the MTF of the imaging system was measured. The lenses vary in focal length, except for lens 1 and 3, which are structurally identical, but from a different production line (Table 2). Their results provide an estimate of the variance of the process. Thus we get a measure for the effects

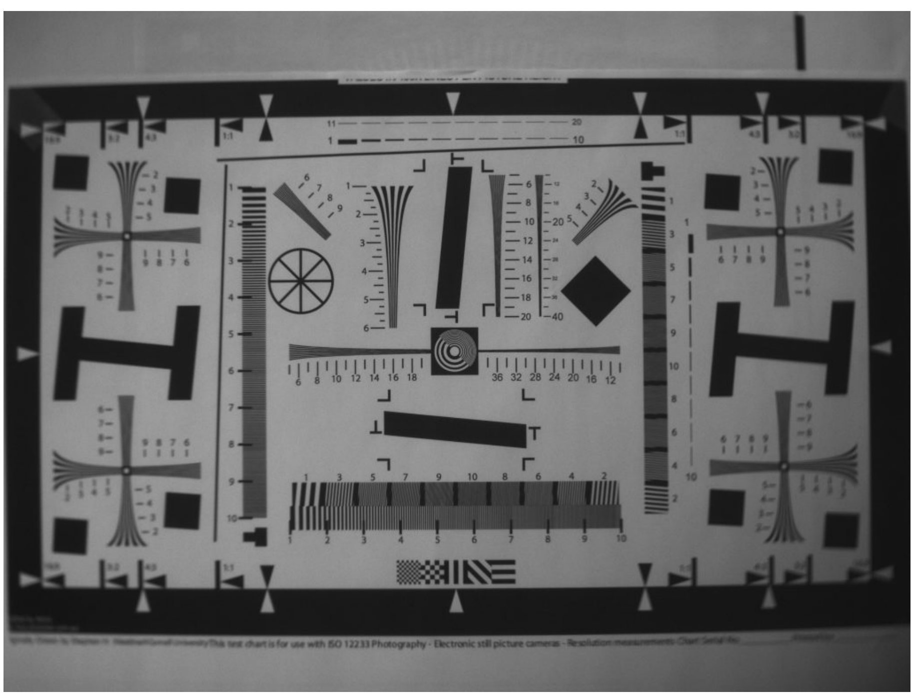

Figure 2. Example of one MTF measurement scene with ISO12233 target

lens	focal length	type
1	$8mm$	normal
2	$12.1mm$	tele
3	$8mm$	normal
4	$13mm$	tele
5	$12.6mm$	tele
6	$0.04mm$	wide

Table 2. Parameters of the used lenses

of different lenses on the imagers and a rating for the overall performance of one combination. A fine print ISO12233 test target was used for the slant-edge analysis, implemented in Matlab, according to the standard (Fig. 2).

Fig. 3 and Fig. 4 show the normalized MTF for each lens over the corresponding spatial frequency measured in cycles per pixel, with respect to $F_{Nyquist}$. At the frequency where the MTF is 0.5, the contrast has only half its original value. For all investigated systems, lens 4 poses the best performance, with almost linear behavior. Lens no. 6, a wide angle lens with short focal length, shows the lowest performance, with even dropping under a critical 0.1 MTF before reaching half of the sampling rate ($F_{Nyquist}$) and thus limiting the sensors performance. We expect this lens to degrade the performance of the imaging system noticeably. Also it is evident, that sensor 2 will show better performance as sensor 1 because of its superior resolution.

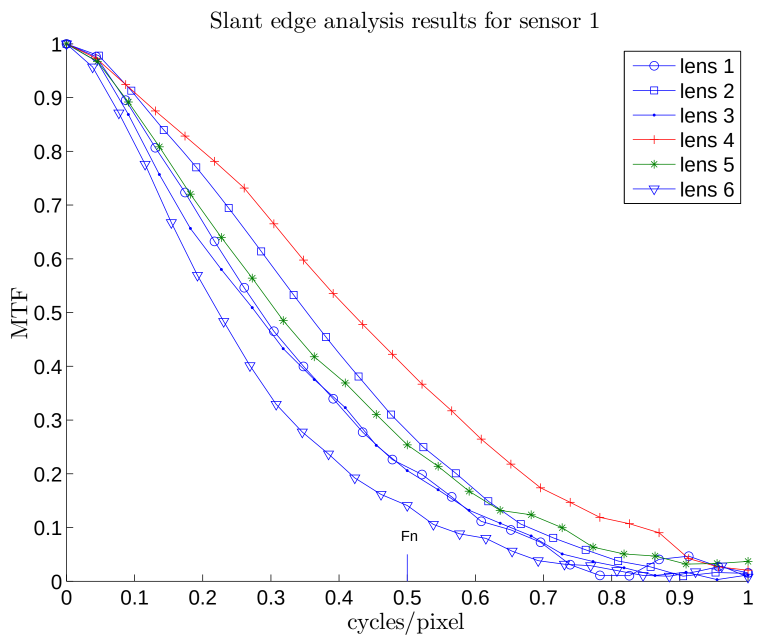

Figure 3. Normalised MTF for sensor 1 over cycles per pixel (F_n = Nyquist Frequency). Lens 4 shows the best performance and lens 6 the lowest. The structurally identical lenses 1 and 3 show the expected equal performance.

A comparison of the two graphs shows that the lower resolution of sensor 1 leads to less specifiable graphs and an overall lower performance. Lenses 1 and 3 show comparable MTF performance on both sensors with a small divergence for sensor 2, which we relate to a higher sensitivity for measurement errors. This leads to the assumption, that small differences between lenses can not be resolved sufficiently with this simple measurement setup. However, the measurement bears a high potential for fast and easy performance estimation and a first impression of how the overall system affects contrast and resolution which directly influences the performance of a given recognition task.

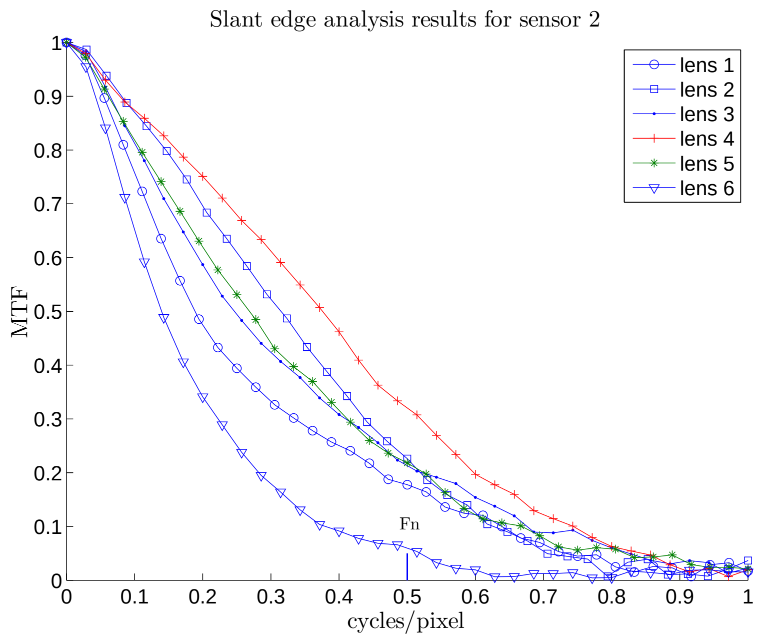

Figure 4. Normalized MTF for sensor 2 over cycles per pixel (F_n = Nyquist Frequency). Lens 4 shows the best performance, while the low performance of lens 6 even limits the sensors performance.

3. REAL WORLD EXPERIMENTS

This section describes the experimental work on the subject. The influence of the different imaging setups is evaluated on the scenery in front of a moving truck. The sensed images are processed offline to extract features for building up knowledge about the position and specificity of road lanes. The camera and lens combinations

Figure 5. Scene from real world experimental setup (left: sensor 1, right: sensor 2)

were placed behind the windshield at the center of the vehicle. It is important to align the field of view, to have the effects of different imaging systems pictured in comparable data and not to introduce more scenery than needed. Sequences of two different combinations are then recorded in parallel for offline processing, allowing for a direct comparison (Fig. 5).

Road lanes represent vertically oriented contrast edges in the data and can be filtered with classical edge detection algorithms, like sobel filtering or enhanced variants, with respect to orientation and magnitude.[8]

To relate the measured imaging system performance to this real world scenario, a measure from statistical signal analysis can be applied. In one line of an image, a road marking is described as a contrast step from the lower intensity of the road, to the higher intensity of the marking. This information is transformed into a deterministic signal $x(t) = x_0 \cdot rect(t/T)$ with signal length T defined by the width of the road marking itself. Each imaging system has a different response to the carrier signal, based on the influence of the different MTF. This directly affects the error probability and performance of a filter, designed for signal detection. The quality of the transmitted signal, is described by the signal-to-noise ratio (SNR). A matched filter with the impulse function $h(t) = c \cdot x(T - t)$ is able to produce the maximum SNR for a given input signal at time $t = T$. The output of a filter with response function $h(t)$ excited by $x(t)$ is

$$y(t) = \int_{-\infty}^{\infty} h(t-\tau)x(\tau)d\tau \tag{2}$$

The signal is assumed to be superimposed by a white noise process $N(t)$ with the sample function $n(t)$ and the auto covariance function $c_{NN}(\tau) = N_w \cdot \delta_0(\tau)$. The SNR at the output of the filter at time $t = T$ is described with the ratio of the squared source signal and the square mean of the noise process, which equals its variance.[9]

This leads to

$$SNR = \frac{\left(h(t) * x(t)\right)^2 \|_{t=T}}{N_w \cdot \int_{-\infty}^{\infty} h^2(t)dt} \tag{3}$$

With this measure it is possible to specify the detection probability of a given signal for the correlator $h(t)$. The higher the SNR at the receiver input, the higher the detection probability of the signal. The marking width in an image depends on the distance from the vehicle. With growing distance the number of pixels describing the marking decreases and thus the length of the input signal. The matched filter is adapted for each image line to manage the decreasing number of pixels to compute the maximum SNR. For stabilization of the measurement in each image line, the mean of the signal and noise power is computed in a small patch along the marking and scaled with the marking width of the according viewing distance. The needed information about position and width of the marking features are provided by the lane recognition system that is used to process the data and based on the estimations for its road model.

The data for all sensor lens combination was recorded on about $100km$ of freeway and driven several times to provide comparable input. The data is then processed offline by the lane recognition system and the SNR measurement is applied on the detected lanes.

In Fig. 6 we can see how the SNR for the two sensors compared to each other behave over a long time measuring campaign. Both equipped with the same lens model (i.e. lens 1&3), a distinguishable SNR from each other for the right lane marking is achieved. The performance difference which could be seen in the MTF measurement is visible with sensor 2 being superior.

The high variation at the beginning is due to very poor quality of the marking itself (Fig. 7), resulting in unbalanced signal response. For normal quality, a stable average is maintained with a visible correlation for performance drops, while passing dashed markings of exit lanes.

Fig. 8 and Fig. 9 show the results for 250 consecutive mean filtered SNR measurements at 15 frames per second of the right lane marking with each lens (lens 3 is the same model as lens 1 and therefor omitted). The used sequences show scenes with normal marking quality and were not recorded at the same time.

Though the impact of external conditions has to be considered, when comparing sequences that are not recorded at the same time, differences in the average performance are apparent. It is possible to rediscover the MTF performance, with lens 4 having an overall high value, lens 6 at the bottom and the other lenses separated in between. The tendency for the low performance of lens 6 and the overall high performance with lens 4 can be found for both sensors. Sensor 2 shows the expected better performance, with an apparently better overall SNR

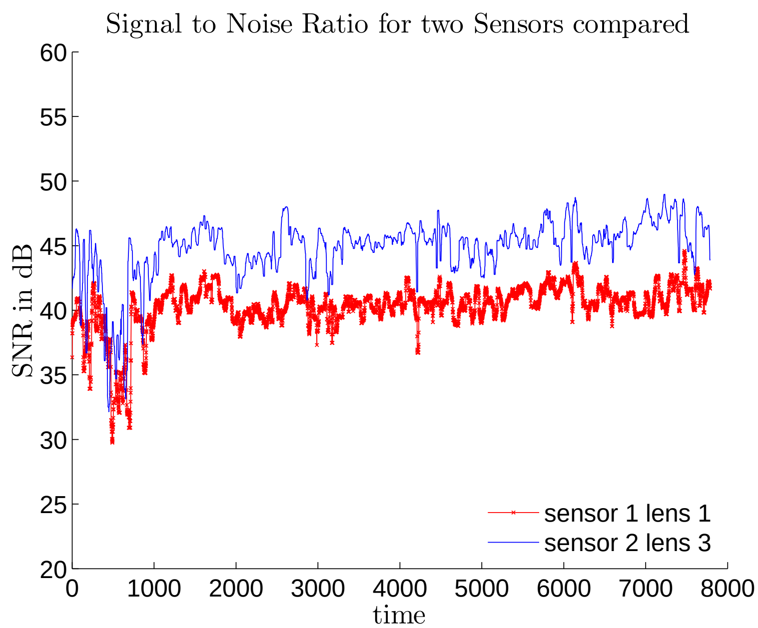

Figure 6. SNR results of the right lane for sensor 1 and sensor 2 equipped with the same lens for a long sequence of one measuring campaign (ca. $9min$ at 15 frames per second).

Figure 7. Example for poor marking quality, which results in lower SNR

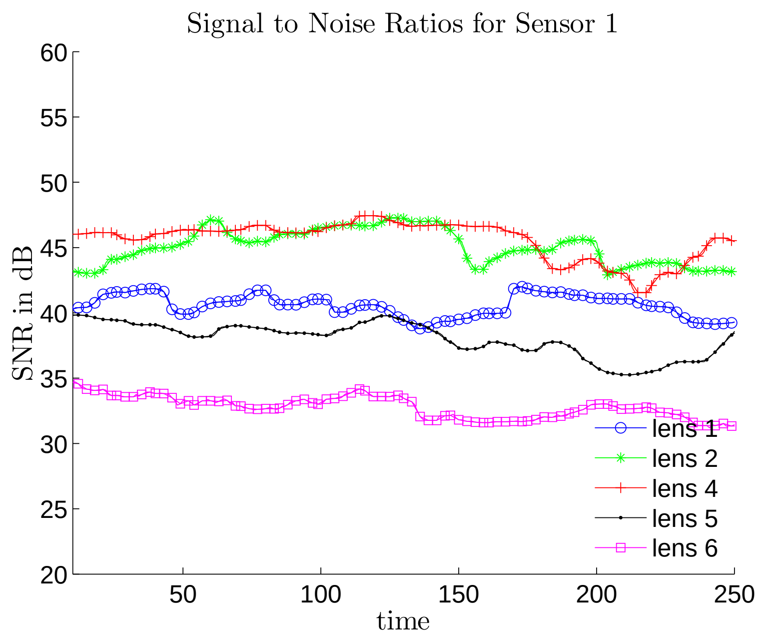

Figure 8. SNR results of sensor 1 for different lenses (250 consecutive measurements at 15fps, lens 3 omitted)

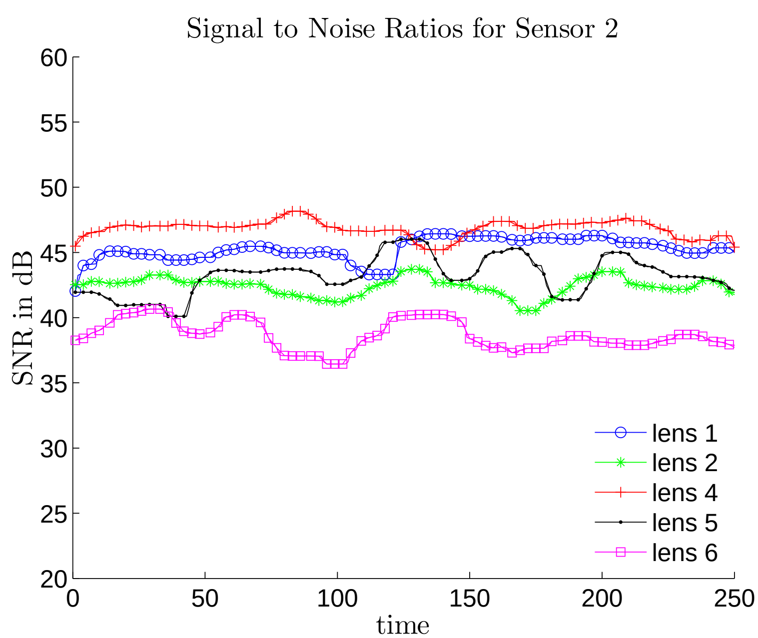

Figure 9. SNR results of sensor 2 for different lenses (250 consecutive measurements at 15fps, lens 3 omitted)

for lens 4 and lens 6 and a less spread range of values due to less quantization noise. As assumed, lenses that are not well separated in the MTF graph, can not be separated well in these measurements at different times either. The variation of the environment is higher than the variance of the lenses.

With a focus on distance in the computation process the SNR measurement can provide more significant information on the impact of the different sensor lens combinations on the system performance. By dividing the sensed road scene into segments and computing the mean SNR for each signal segment, it is possible to relate the sensor performance to distance in front of the vehicle. The average of all SNR measurements inside a segment of $5m$ size represents its performance. These segments are then spread from $5m$ to $35m$ to cover the complete region of interest for the recognition task.

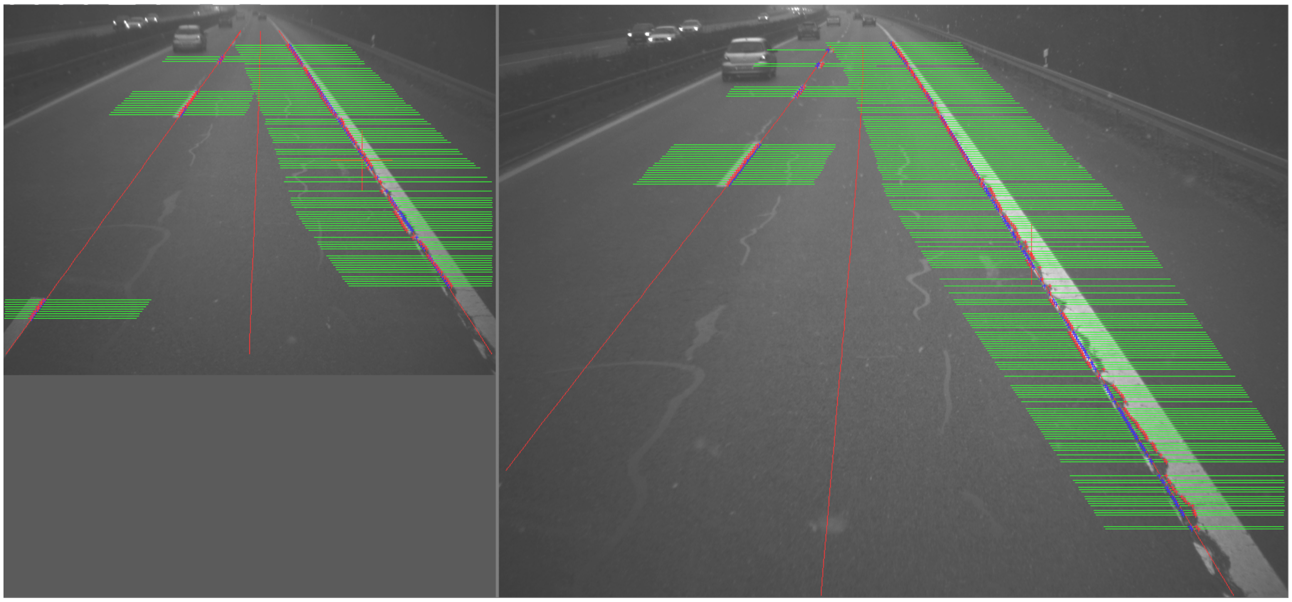

Figure 10. Lane feature extraction result for a dark scene with low marking quality. Green lines represent successful lane detections and the road model state is colored red. Sensor 1 with lens 1 on the left shows a significant performance loss against sensor 2 with lens 3 at far distance. In the near distance both sensors show incomplete detection of the lane due to bad quality of the markings.

Comparing the sensor 1 with lens 1 and sensor 2 with lens 3 on a dark scene with bad road marking quality shows, that sensor 1 fails to resolve the right lane marking above $35m$, whereas sensor 2 is capable of resolving the lane features in the complete region of interest up to $40m$ (Fig. 10).

Looking at the SNR for each segment shows that the ratio drops below $30dB$ in distances above $35m$ for sensor 1, and the perceptive component is not able to produce successful feature extraction (Fig. 11). In addition the poor quality of the road markings is visualized by the low measurements in $5m$ and $10m$ distance, where the ratio is even lower than the next further segments where the cracks are not resolvable for the sensors and thus a better signal is perceived.

Fig. 12 shows a scene with wet road, where the lane recognition fails on sensor 2 with lens 6 in far distances. Sensor 1 with lens 5 shows better performance in the distance but with the loss of information in the near due to the different perspective. In the SNR evaluation (Fig. 13) it is visible, that the measurements on sensor 2 drop under $35dB$ above $30m$. With lens 5 sensor 1 the ratio stays above $35dB$, but the loss of information in the near distance has to be considered for the task at hand. The measurement helps to determine and investigate the performance differences and the limit range for the imaging setups and the connected perception task.

4. CONCLUSION

The measurement of the MTF using a proposed standard can be applied to evaluate the performance of imaging systems, without the need of cost intensive measurement processes and tools. We assessed the measured performance by processing the data of different imaging setups for the task of lane recognition. The presented measurement of the SNR of an optimal feature signal, is able to rate the performance of different imaging systems when compared to each other and resolves the same performance relations of the MTF analysis of the imaging systems. By computing the SNR for different segments of the data it is possible to evaluate the performance differences dependent on parameters like viewing distance and further distinguish the different components.

The SNR provides a relative performance measure for the significance of an explicit feature for captured scenes and conditions. Influences on the feature quality, lightning and changing weather conditions, distort the

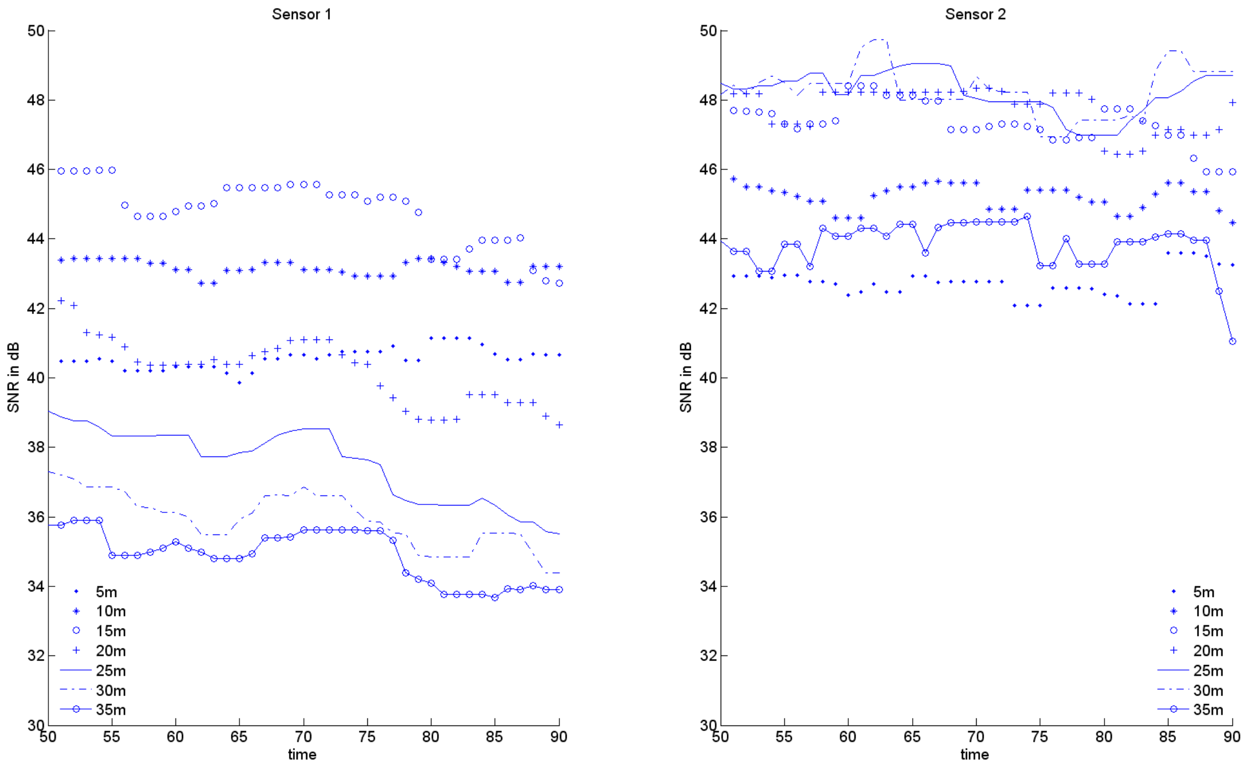

Figure 11. Comparison of the SNR evaluation for different distance segments of the sequence shown in Fig. 10. The low performance of sensor 1 with growing distance is represented by a low SNR for the segments above $30m$. Sensor 2 shows superior performance for all segments. The bad quality of the road markings results in a moderate ratio for the near distance.

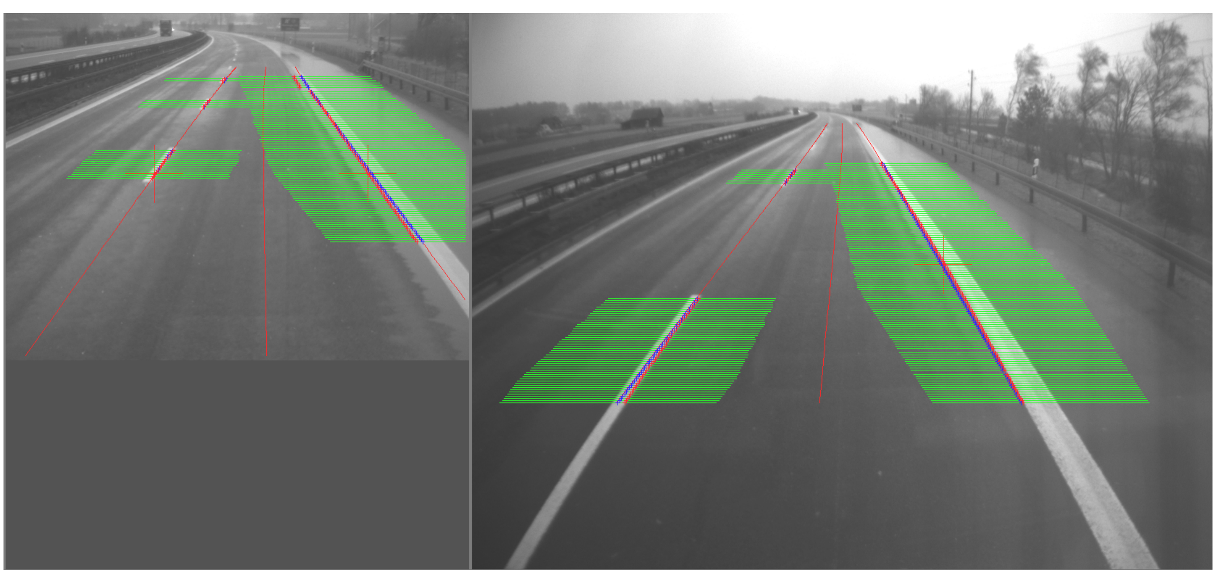

Figure 12. Lane feature extraction result with different lenses on wet road. Due to the different perspective, Sensor 1 with lens 5 (left) is capable of resolving features in the far distance, where sensor 2 with lens 6 fails to detect anything for the far part of the viewing area. This gain in distance for lens 5 comes with the loss of data in near distances, which are simply not covered by the field of view.

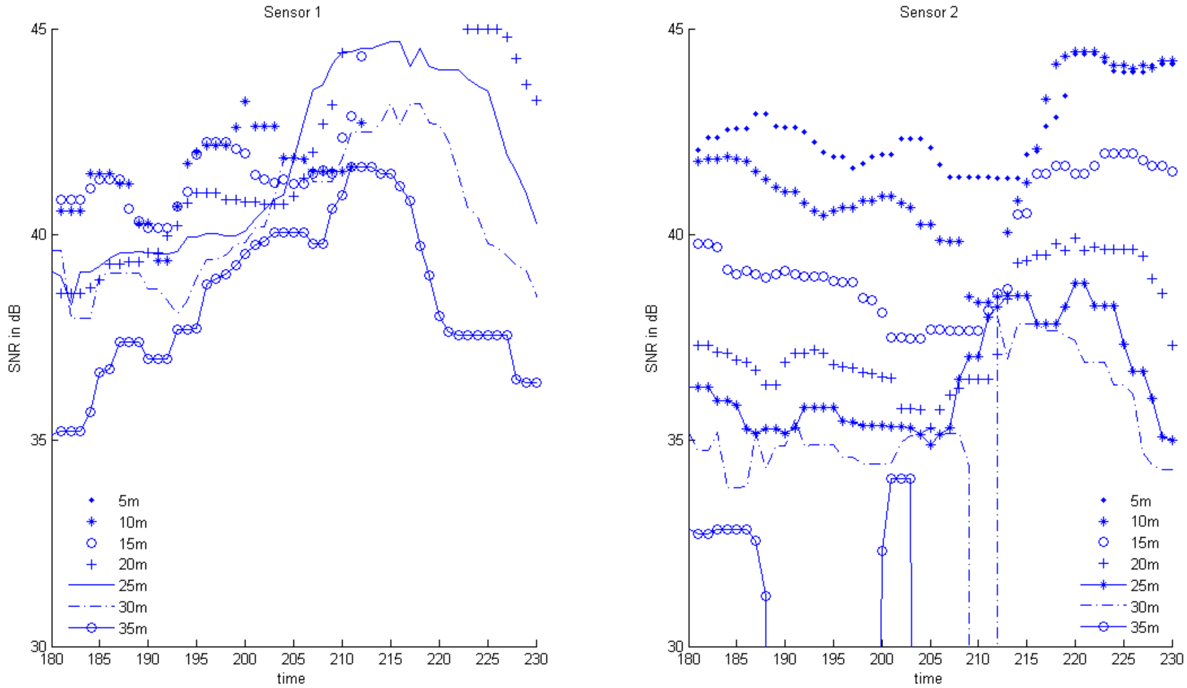

Figure 13. SNR evaluation for different distance segments of the scene shown in Fir. 12. Sensor 1 with lens 5 outperforms sensor 2 with lens 6 in distances above 25m but with the loss of information for road features under 10m due to the different perspective. The constant good quality of the markings is visible in the stepwise decreasing ratio from 5m on in sensor 2.

results respectively. To get precise and reliable results it is advised to compare systems that are recording in parallel, or use a large quantity of data from each imaging component, that covers enough scenarios to compute a fair average result.

5. FUTURE WORK

Improving the accuracy and reliability of the measurements is part of the current work in progress. Another step is to expand the experiments to other perception tasks and recognition systems, as well as different sensor classes. The MTF measurement information can be improved by introducing interocular analysis to account the varying performance of a sensor lens combination over the complete field of view. There are several proposals and solutions for improved analysis, but these also introduce more complex measurement processes which we try to keep simple. To map this more detailed performance information into measurements in real world is another task at hand.

REFERENCES

[1] Franz, S., Willersinn, D., and Kroschel, K., "A Performance Evaluation Concept for Perceptive Components of Driver Assistance Systems," in [*Proceedings 23rd IAR Workshop on Advanced Control and Diagnosis*], (2008).

[2] Williams, D. and Burns, P., "Low-frequency MTF estimation for digital imaging devices using slanted-edge analysis," *Proceedings of SPIE* **5294**(1), 93–101 (2003).

[3] International-Standards-Organization, "ISO12233 Photography - Electronic still-picture cameras - Resolution measurements," (2000).

[4] Hertel, D. W. and Chang, E., "Image Quality Standards in Automotive Vision Applications," in [*Proceedings of the IEEE Intelligent Vehicles Symposium*], *Intelligent Vehicles Symposium, IEEE*, 404–409 (2007).

[5] Angelis, A., Moschitta, A., Russo, F., and Carbone, P., "Image Quality Assessment: an Overview and some Metrological Considerations," in [*IEEE International Workshop on Advanced Methods for Uncertainty Estimation in Measurement*], *IEEE International Workshop on Advanced Methods for Uncertainty Estimation in Measurement*, 47–52 (2007).

[6] European Machine Vision Association, "EMVA Standard 1288: Standard for Measurement and Presentation of Specifications for Machine Vision Sensors and Cameras," (2007).

[7] Estribeau, M. and Magnan, P., "Fast MTF measurement of CMOS imagers using ISO 12333 slanted-edge methodology," *Proceedings of SPIE* **5251**(1), 243–252 (2004).

[8] Lee, J. W., "A Machine Vision System for Lane-Departure Warning," *Computer Vision and Image Understanding* **86**(1), 52–78 (2002).

[9] Kroschel, K., [*Statistische Informationstechnik: Signal- und Mustererkennung, Parameter- und Signalschätzung*], Springer, Berlin, 4., neubearb. aufl., studienausgabe. ed. (2004).