



# Exposure time and point cloud quality prediction for active 3D imaging sensors using Gaussian process regression

Alejandro Magaña<sup>1</sup> · Lukas Schneider<sup>1</sup> · Maximilian Benker<sup>1</sup> · Thomas Altmann<sup>1</sup> · Philipp Bauer<sup>1</sup> · Gunther Reinhart<sup>1</sup>

Received: 5 August 2023 / Accepted: 13 November 2023 / Published online: 26 December 2023  
© The Author(s) 2023

## Abstract

Setting an optimal image exposure is crucial for acquiring dense point clouds using 3D active optical sensor systems such as structured light sensors [structured light sensors (SLSs)] and active stereo sensors. One of the most common and seamless ways to optimize the image brightness of an image exposure is to adjust the camera's exposure time. However, optimizing the image exposure alone is ineffective for acquiring surfaces of large-scale objects with a complex topology if a spatial understanding of the scene is neglected. Hence, the present paper proposes a data-driven approach using two Gaussian processes [Gaussian processes (GPs)] regression models to select a proper exposure time considering the nonlinear correlations between image exposure and the scene spatial relationships. To model these correlations, our study introduces first the generic synthesization of seven inputs and two target variables. Then, based on these inputs, two independent GPs are designed: one for predicting the measurement quality and one for estimating the exposure time. The performance and generalizability of both models are thoroughly evaluated using an SLS and an active stereo sensor. The evaluation demonstrated that the point cloud quality models adequately matched observations with an R2 exceeding 90%. Specifically, the models predicted point cloud quality with an root mean square error (RMSE) of 10%. Additionally, the assessment of the performance of the exposure time models showed a model fit with an R2 above 97%. The exposure time prediction accuracy, as evidenced by the RMSE values, was within 10% of the corresponding exposure time range for each sensor. The present research shows the potential and effectiveness to completely automate the assessment of a point cloud quality and the selection of exposure times with the help of data-driven models.

**Keywords** 3D active optical sensor systems · Exposure time estimation · Point cloud quality prediction · Gaussian Process regression · Structured light sensor · Active stereo sensor

## 1 Introduction

The increasing performance of 2D and 3D sensors combined with the falling prices of electronic components over the last two decades have enabled researchers and industry to develop cost-effective solutions [1] for improving vision tasks, such as image-based quality inspection, digitization, reverse engineering, object detection, surveillance, and navigation. In particular active 3D imaging sensors, such as

structured light sensors (SLSs), comprising at least one camera and a light projecting source, have been established as the leading acquisition technology for high precision vision tasks due to their accuracy and large working volumes [2]. By combining optical triangulation and pattern illumination, active sensors can generate an accurate 3D sampled representation of a probing object's surface, known as a *point cloud*. The generated point clouds can then be interpreted and processed to solve different vision tasks.

However, obtaining a robust and reliable sampling of an object's surface, i.e., a high-density point cloud, requires that an adequate camera exposure has been previously selected. Correctly setting the camera exposure of imaging sensors is a non-trivial task that requires the consideration of several influencing variables, e.g., camera parameters (exposure time, lens aperture, sensitivity),

✉ Alejandro Magaña  
alejandro.magana@tum.de

<sup>1</sup> TUM School of Engineering and Design, Department of Mechanical Engineering, Institute for Machine Tools and Industrial Management, Technical University of Munich, Boltzmannstr. 15, 85748 Garching, Bavaria, Germany

projecting device parameters (light intensity, pattern) [3], probing object surface material [4]. Figure 1 exemplifies the complexity of this task for measuring highly reflective large-scale objects, such as a car door, using an SLS. The camera image (left) shows that the scene illumination is mostly inhomogeneous and that a globally optimized exposure time is likely ineffective to obtain a high-density point cloud (right) for acquiring all features. Furthermore, this example illustrates the importance of considering the spatial relationships between the features and the sensor in order to achieve an appropriate image exposure.

Due to the many factors influencing scene illumination, selecting a proper exposure time is still a challenging task generally performed manually, which can take up to several hours and strongly depends on the vision system characteristics, the object's geometric complexity and size, and user expertise. Therefore, automated optimization of image exposure to ensure dense point cloud quality is a key step toward effective and resource-efficient performance of vision tasks using active 3D optical sensors. For this reason, the present study proposes a data-driven, supervised learning approach using two independent Gaussian processes (GPs) regression models for predicting the local quality of a point cloud and for estimating the required camera exposure time. The introduced models were developed in the context of coordinate metrology applications, where the robust acquisition of individual *features* (related terms: inspection feature, artifact, region, area, or point of interest) has a higher priority than the entire object's surface. For this reason, the current research follows a local feature-driven optimization rather than a global approach. Both proposed data-driven models are trained using first the camera image's local light intensity around a feature and the spatial relationships between the features and the sensor. The models are iteratively developed and analyzed to gain a better general understanding of the model outputs and input spatial correlations. Our work aims at providing users and researchers with a novel method to assess and automate

the acquisition of robust point clouds without considering any hardware or software alteration of the vision system.

## 1.1 Related work

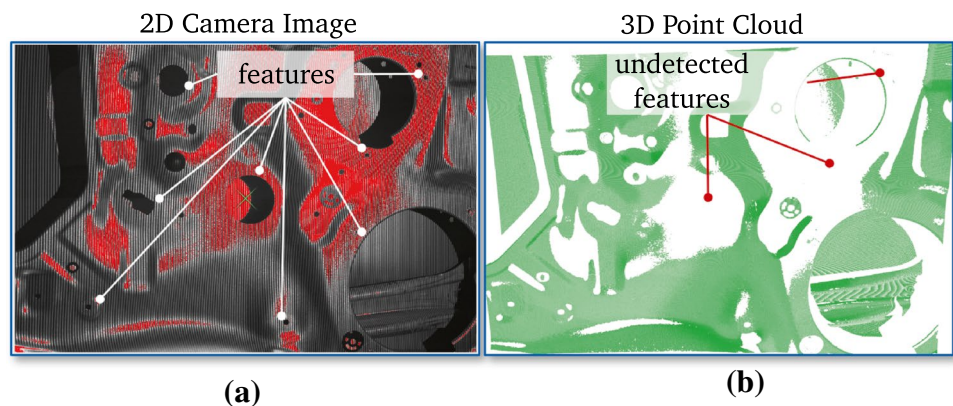
Effective control of the camera exposure for measuring high reflective surfaces is a well-known challenge faced by any active sensor. Hence, several researchers have analyzed this problem over the last three decades and proposed different solutions comprising hardware- and software-based techniques to address it.

One of the most common and pragmatic methods to obtain uniform reflectivity is matte coating the surface of the measuring object [5]. Although this technique is still considered to be highly effective for eliminating the overexposure of the camera image, this approach also has disadvantages, e.g., not all workpieces can be coated, the process is difficult to automate, and the measurement accuracy is affected by the powder thickness and its distribution homogeneity [6].

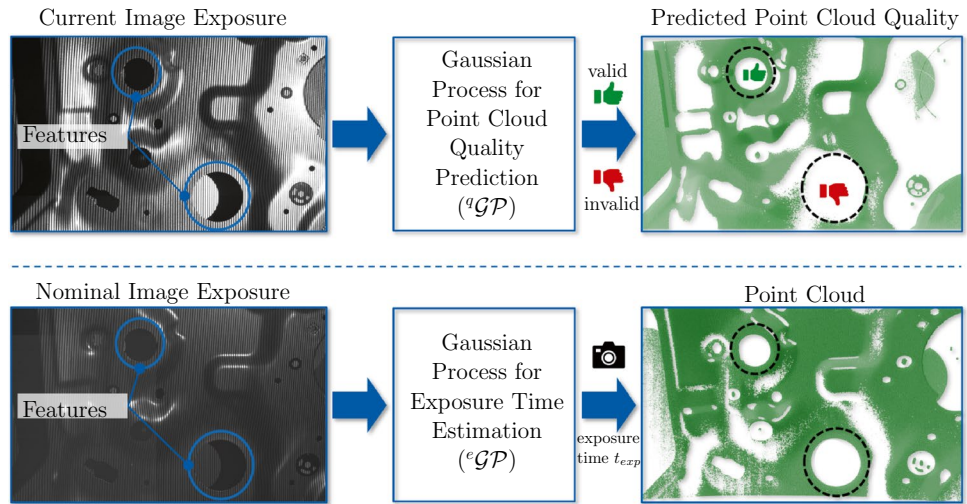
In addition, there are a handful of methods that combine multiple camera images with different exposures to optimize image intensity. Such approaches strive for targeted exposure control of individual pixels and fall under the category of High Dynamic Range (HDR). The targeted spatial image exposure can be achieved using different techniques, e.g., multiple exposure times [7], modulation of the fringe projector [3], pixel-wise adaptation of the fringe projection [8], or hybrid approaches using external filters [9] and multiple cameras [10] or projectors [11]. Moreover, [12] proposed a software-based solution that generates synthetic fringe images without considering any system adaptation. This study refers to the following review papers [13, 14], which offer an overview of the various HDR approaches and their applications.

Although HDR techniques offer robust solutions, such methods are impractical for commercial or industrial active systems that do not allow the required hardware or software modifications. Additionally, many of the HDR

**Fig. 1** Measurement of a car door using an SLS: **a** camera image and corresponding, **b** 3D point cloud. The 2D image provides a qualitative representation of the scene illumination. The red pixels indicate an over saturation of the camera impeding the acquisition of the surface points of the features in the point cloud (color figure online)



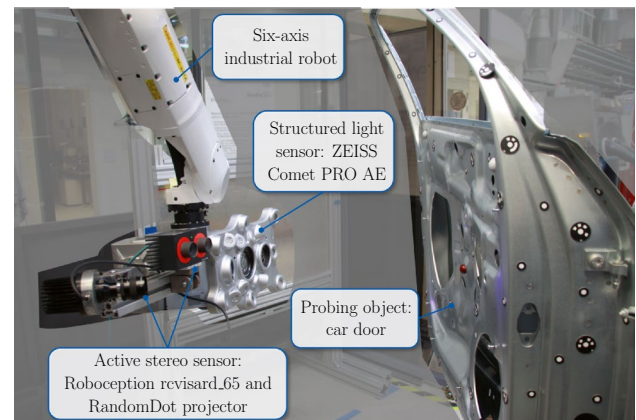
**Fig. 2** An approach involving two data-driven models is utilized to predict measurement quality in order to determine the ideal exposure time subsequently. The quality model (top) predicts the quantitative point cloud quality for a particular feature to evaluate the measurement’s validity. The second model (bottom) calculates the necessary exposure time to achieve a local image intensity for the same feature



techniques require an a priori image analysis, which affects the resource-efficiency (computational and time) of the whole process [15]. Hence, [15] proposed a global exposure time optimization using a single acquired fringe image and an approximated linear model of the exposure time and the image intensity. Although the study demonstrated the validity and effectiveness of an approximated model, the investigation of the spatial relationships between camera and object was not further addressed. More analytical methods were used by [16], which proposed an offline planning system for predicting a valid exposure time based on the reflectance simulation model framework of [4] and a 3D surface model of the probing object to control the camera exposure time. The study of [16] utilized the reflectance model to compute a global optimized exposure time for a given ideal image intensity. Their study showed that the expected surface area could be estimated with a deviation of 20% to predict the quality of the resulting point cloud. Furthermore, [17] argues the importance and relevance of quantifying the quality of point clouds. To this end, the authors introduce a comprehensive set of performance indicators to assess, among other metrics, the quality of point clouds. However, their work focuses on post-acquisition evaluation rather than deriving optimized parameters to ensure high point density point clouds.

**1.2 Approach**

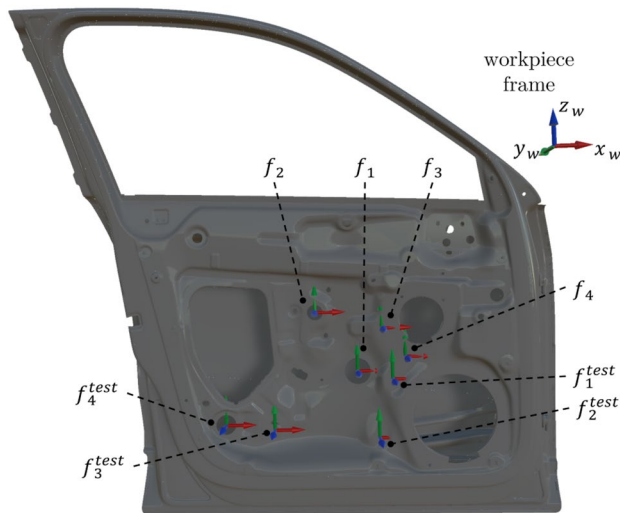
Based on the literature review, the automated optimization of exposure times for active 3D sensors remains an open issue within academia and industry. Moreover, the effective and efficient selection of exposure times can be regarded as a multi-dimensional problem that requires an imaging and spatial understanding of the acquiring scene taking into



**Fig. 3** Overview of the core components of the measurement cell AIBox: car door as probing object, SLS ZEISS Comet PRO AE, and six-axis industrial robot for automated sensor positioning. In addition, an alternative active stereo system, rc\_visard 65 and dot projector from Roboception, was attached to the SLS for evaluating the transferability of the models

account a variety of different the influencing factors. For this reason, a human operator is still required for setting adequately the camera exposure within complex vision applications such as depicted in Fig. 1. Motivated by the idea that an operator is capable of adjusting effectively the camera’s exposure time, where other approaches fail, our study first considers an analysis of the operator’s behavior based on an observation-action-assessment procedure:

- Observation (2D image exposure and spatial relations): Considering an initial exposure time, the operator observes the image intensity concentrating on the features to be acquired.



**Fig. 4** Overview of the door's inside features selected for training and evaluating the proposed GP models

- Action (exposure time): Based on the observed image and lighting conditions, the operator adapts the exposure time.
- Assessment (point cloud): The operator evaluates the measurement based on the resulting point cloud. The previous steps are repeated if the required quality has not been reached.

Towards replicating the cognitive visual and spatial capabilities and expertise of a human operator, our study proposes a data-driven approach using GPs trained on 2D images, point clouds and spatial data to predict the expected measurement quality and estimate an adequate exposure time for a given feature. To this purpose, our study considers two GP models, one for assessing the quality of the point cloud ( ${}^q\mathcal{GP}$ ) and one for estimating the exposure time ( ${}^e\mathcal{GP}$ ) given a local image intensity and the spatial relationships between the sensor and a given feature. Since the proposed approach does not globally optimize measurements, a local image intensity value is used to assess the quality of the point cloud or predict an exposure time matching it. The local image intensity represents the illumination around a feature and serves as an input for the point cloud assessment or nominal value for prediction of the exposure time.

- ${}^q\mathcal{GP}$ : The present research assumes that the quality of the point cloud can be predicted (operator assessment) using a local image light intensity and the spatial relationships between the sensor and the features (operator observation). Hence, in a first step, this study considers a GP model to predict the local point cloud quality corresponding to a feature.

- ${}^e\mathcal{GP}$ : On the assumption that the required local image exposure matching an expected point cloud quality can be found, a second GP model is used to estimate (operator action) the exposure time also considering the spatial relationships between the sensor and a feature.

A simplified overview of the proposed models, inputs, and outputs is shown in Fig. 2. In addition, the proposed approach assumes that the overall image exposure can be well controlled solely by the camera exposure time if the spatial relationships between the sensor and measurement object are known, and a constant scene illumination can be guaranteed during each measurement. This characteristic is mostly given for active sensors, where the projector can be considered as the dominant light source of the scene. Therefore, the influence of external light sources can usually be neglected [15].

### 1.3 Contributions

Our work presents a novel and data-based framework, which proposes a systematic and incremental design of two GP regression models for assessing the point cloud quality and estimating the exposure time for 3D active sensors. The main contributions of the study are the following:

- Synthesization of generic input and output variables for characterizing the point cloud quality, imaging, and spatial system state.
- Four-dimensional (4D) GP kernel to predict the point cloud quality (number of points) around a feature based on a given local image intensity and its relative position (3D) to the sensor.
- Six-dimensional (6D) GP kernel to estimate the required exposure time for a feature's nominal local image intensity considering its relative position (3D) and rotation (2D) to the sensor.

### 1.4 Outline

This paper first introduces in Sect. 2 the experimental setup for training the proposed regression models and presents the fundamentals of GPs. Then, the core contribution of this paper, i.e., the development and evaluation of both regression models as well as the design of the model inputs and outputs, is discussed in detail in Sect. 3. A thorough validation based on experiments with different lighting conditions, surface finishes and the generalization of the framework to other sensors is presented in Sect. 4.

## 2 Experimental setup and Gaussian process models

### 2.1 Experimental setup for data collection

The integrated industrial robot vision system *AIBox* from *ZEISS* was utilized to acquire all training and evaluation datasets of the present study. The *AIBox* is an industrial robot-based measurement cell comprising an SLS (*ZEISS Comet PRO AE*) used for different vision tasks, e.g., dimensional metrology, digitization, and reverse engineering. The SLS is attached to a six-axis industrial robot (*Fanuc M20ia*) to enable a free positioning of the sensor; hence, the full automation of the measurement process. Figure 3 shows the core elements of the *AIBox*, which are more detailed described in the following subsections.

#### 2.1.1 Object

A sheet metal car door with different surface finishes as shown in Fig. 4 was used for acquiring training and evaluation datasets. Car doors are well-known reference objects in industry for the evaluation of measurement systems, since they exhibit a high feature density and variability (e.g., edges, pockets, holes, slots, and spheres) and different surface characteristics (e.g., finishes, topological complexity).

#### 2.1.2 Features

The car door contains up to 500 features. Our study focuses exclusively on the acquisition and evaluation of hole features. In coordinate metrology, holes are challenging to evaluate since they require an assessment of the surface around

the perforation. Given their inherently hollow topology, considering this property within the data processing pipeline facilitates a seamless transferability of this approach to the assessment of similar features, such as pockets and slots. The pipeline should also be easily applicable to other geometric features, such as points, spheres, and squares. Additionally, the car door has a high variability (sizes, finishes, and locations) of hole features, which is useful for a comprehensive evaluation of the models. Fig. 4 visualizes some exemplary circular features used within this study. It is assumed that the 3D position and 3D orientation of all features are known with respect to a reference coordinate system.

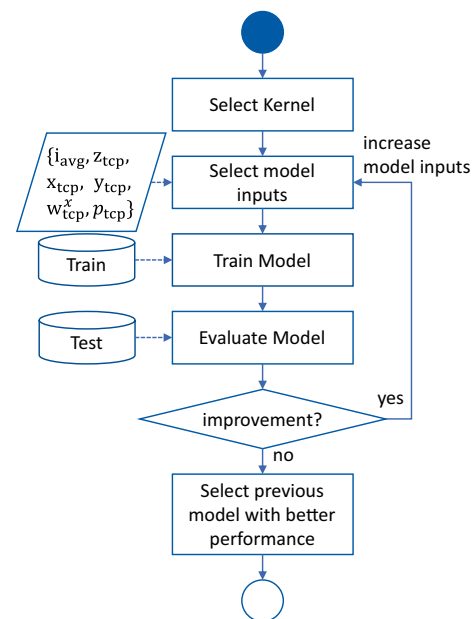


Fig. 6 Generic incremental model development workflow

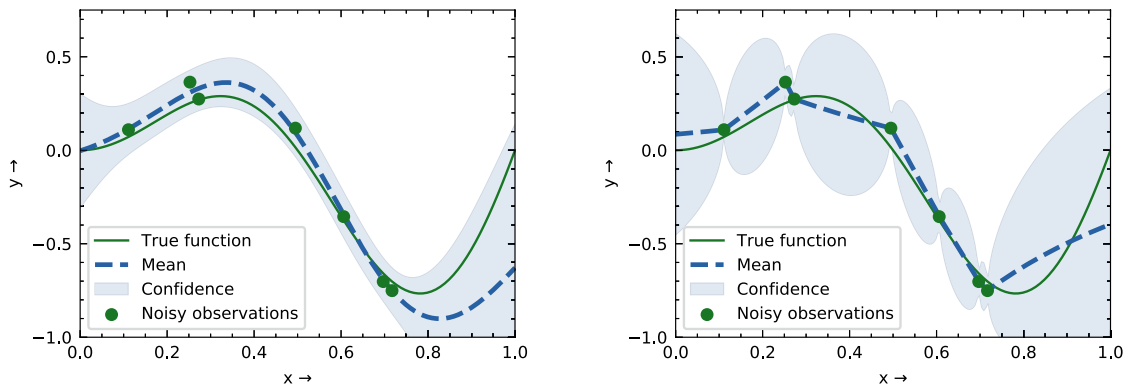


Fig. 5 Exemplary GP regression models; the true function  $f(x) = x \sin(2\pi x)$  is reconstructed by two different GP regression models with a SE (left) and a RQ (right) kernel. Both models are trained based on a set of 7 noisy observations

### 2.1.3 Sensors

The SLS ZEISS Comet PRO AE consists of a monochrome camera and a blue-light digital fringe projector. The camera integrates a band-pass filter adjusted to the digital projector. Additionally, the SLS has a photogrammetric camera for high-accuracy image registration.

The sensor is controlled by the proprietary software ZEISS colin3D, which allows the setting of the camera exposure time. Furthermore, the software has an automatic exposure function and a proprietary HDR mode, which were deactivated for the acquisition phase. A parameterization of further camera settings, such as light sensitivity, lens aperture, or any light projector configuration is not allowed.

To evaluate the transferability of our approach with other sensors and acquisition principles (cf. Sect. 4.2), a stereo sensor (*rc\_visard 65*) and a dot projector from *Roboception* were attached to the SLS (see Fig. 3). The imaging parameters of both sensors are given in Table 7.

### 2.1.4 Positioning devices

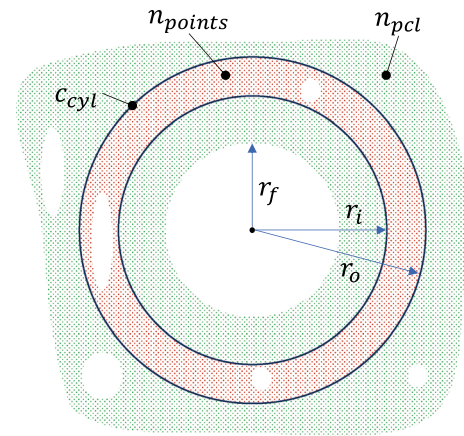
The *AIBox* integrates a six-axis industrial robot to position the SLS in six degrees of freedom. The rotary table enables the acquisition of the car door inside and outside.

### 2.1.5 Measuring environment

The whole system is enclosed by a cell box with tinted windows to minimize the influence of external lighting sources. The measurement cell is mounted in our laboratory, which resembles a real production hall and corresponding lighting conditions.

## 2.2 Gaussian processes

The present study proposes two GP regression models for predicting the point cloud quality and the exposure time. This subsection provides a brief overview of the GPs mathematical foundations and their general characteristics.



**Fig. 7** Simplified representation of the output variable  $n_{points}$  for assessing the quality of the point cloud for a specific circular feature.  $n_{points}$  represent the subset of points of the acquired point cloud  $n_{pcl}$  that are inside the hollow cylinder  $c_{cyl}$

### 2.2.1 Fundamentals

Formally, a GP is a stochastic process, i.e., a collection of random variables over a time or space domain, in which any finite subset of it has a multivariate Gaussian distribution. According to [18], a GP can also be seen as the probability distribution of an infinite number of functions over the output  $g(\mathbf{x})$  for a particular input value  $\mathbf{x}$ . A GP is fully specified by its mean function  $m(\mathbf{x})$  and its kernel  $k(\mathbf{x}, \mathbf{x}')$  representing the covariance function between its inputs:

$$m(\mathbf{x}) = \mathbb{E}[g(\mathbf{x})], \tag{1}$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(g(\mathbf{x}) - m(\mathbf{x}))(g(\mathbf{x}') - m(\mathbf{x}'))]. \tag{2}$$

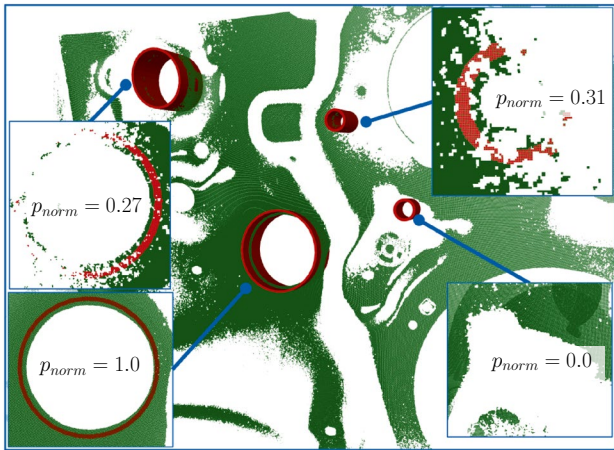
If a function  $g(\mathbf{x})$  follows a GP, this can be denoted as

$$g(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')). \tag{3}$$

Usually, the mean function corresponds to a zero mean function  $m(\mathbf{x}) = \mathbf{0}$  and the kernel  $k(\mathbf{x}, \mathbf{x}')$  remains as the key element to determine the values of  $g(\mathbf{x})$ .

**Table 1** Overview of the synthesized target variables and inputs with their respective units and descriptions

Notation	Unit	Description
<i>Target variables</i>		
$p_{norm}$	–	Normalized number of 3D points from a defined area around a feature
$t_{exp}$	ms	Camera exposure time in milliseconds
<i>Input variables</i>		
$x_{tcp}, y_{tcp}, z_{tcp}$	mm	Feature’s relative position to sensor’s TCP
$w_{tcp}^x, p_{tcp}^y, r_{tcp}^z$	°	Feature’s relative orientation to sensor’s TCP
$i_{avg}$	–	Average light intensity of 2D image in greyscale [0–255] from a defined area around a feature

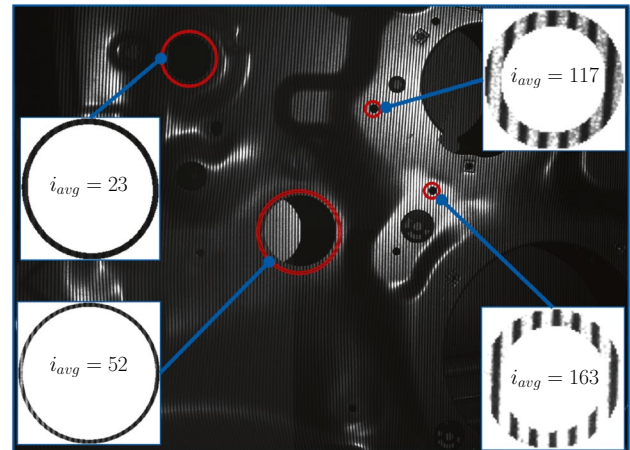


**Fig. 8** Extraction of 3D points (red)  $n_{points}$  using the acquired point cloud (green) and a 3D Mask (red synthetic cylinder) (color figure online)

### 2.2.2 Kernel design

The kernel function  $k(\mathbf{x}, \mathbf{x}')$  defines the covariance between a pair of function values of  $f(x)$ . This characteristic is considered one of the principal strengths of kernel methods, such as GPs, since the kernel function can be designed using explicit prior knowledge and assumptions about the unknown function  $f(x)$ . Therefore, selecting or designing an appropriate kernel is crucial for effective training, as it influences how random variables and observations correlate and ultimately impact the model’s performance. The mathematical definition of some exemplary kernels is given in Appendix 1.

Moreover, Fig. 5 presents a graphical representation of the impact of the kernel through the utilization of two exemplary GP regression models trained on the basis of a dataset  $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$  with  $N = 7$  observations. The models aim to predict the function  $f(x) = x \sin(2\pi x)$  using a squared exponential (SE) kernel (left of Fig. 5) and a rational quadratic (RQ) kernel (right of Fig. 5). The trends emphasize the significance of kernel design. Although both models manifest effective performance on observations, the SE kernel provides better predictions within interpolation ranges. In addition, Fig. 5 depicts a further characteristic of GPs, which shows that model predictions are intrinsically associated to an uncertainty. This feature is illustrated in Fig. 5, which exhibits a 10% confidence interval which increases beyond the training regions in both scenarios. This characteristic is of great advantage in industrial applications, as it allows also to evaluate the reliability of the prediction and to act accordingly.



**Fig. 9** Computation of the local light intensity  $i_{avg}$  for each feature using one camera’s fringe image and a 2D Mask

## 3 Model development

This section provides first an overview of the acquired dataset used for the training and evaluation of the GP models. First Sect. 3.1 introduces the characterization of the models inputs and outputs. Then, an overview of the used trained data is given in Sect. 3.2.2. Section 3.3 outlines the development of the GP model for predicting the point cloud quality and Sect. 3.4 presents the development of the GP model for the estimation of the exposure time. Following an incremental and systematic model design, the models are evaluated and compared with each other to find the best balance between complexity and performance. Fig. 6 provides an overview of the generic workflow followed in this research to develop, train, and evaluate the models.

### 3.1 Model input and target variables

The selection and synthesis of adequate model input and target variables are considered fundamental steps within the design and employment of data-based models. Given the proposed goals of our study, the synthesized variables should be able to describe the correlations between the local image exposure of a specific area corresponding to a feature and its spatial relationships to the sensor for predicting the resulting point cloud quality and the corresponding exposure time.

This section introduces the design of the target and input variables using linear algebra, geometric analysis, and machine vision algorithms. The characterization of the

inputs aims at a high generalization and an automated dataset generation. An overview of the model target variables and the synthesized input variables is given in Table 1.

### 3.1.1 Target variables

This subsection introduces the synthesization of the target variables, i.e., the point cloud quality and the exposure time, for the quality and the exposure time prediction models.

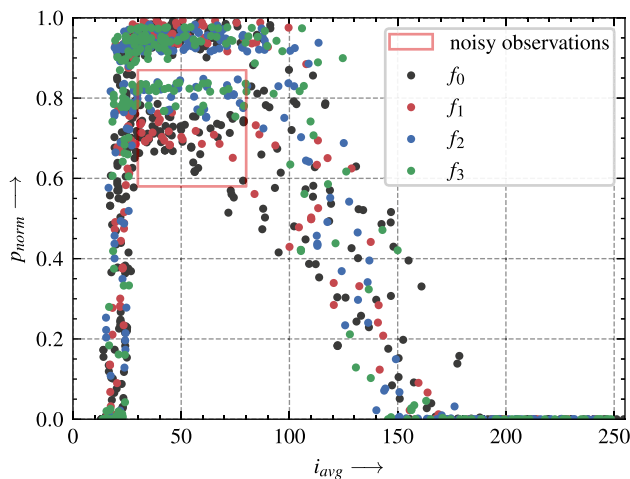
#### Point Cloud Quality

Our study introduces a quantitative metric to express the acquisition quality of a feature. This metric is expressed by the absolute number of points  $n_{points}$  in an acquired point cloud that correspond to a given area of a feature.

This area is calculated by intersecting the point cloud with a synthetic hollow cylinder. The hollow cylinder is built based on the geometry, position and orientation of the features. Given a circular feature with a radius  $r_f$ , a hollow cylinder  $c_{cyl}$ , with an arbitrary height  $h_{cyl}$ , an inner radius  $r_i = r_f + r_{tol,i}$  and outer radius  $r_o = r_f + r_{tol,o}$  is defined. The parameters  $h_{cyl}$ ,  $r_{tol,i}$  and  $r_{tol,o}$  define tolerance variables that must be selected based on the specific application, system or standards.  $n_{points}$  represent then the subset of points of the point cloud  $n_{pcl}$  that are within the hollow cylinder.  $n_{points}$  is given as the intersection of these elements:

$$n_{points} = n_{pcl} \cap c_{cyl}. \quad (4)$$

The position of the cylinder corresponds to the feature's origin and its orientation to the feature's normal vector. Figure 7 shows a simplified 2D graphical representation of



**Fig. 10** Trend of the normalized number of points and local light intensities for different features

the circular feature, the hollow cylinder, and the  $n_{points}$ . The hollow cylinders and the resulting intersected points of some exemplary features of an exemplary real measurement are shown in Fig. 8.

#### Normalized Point Cloud Quality

Since the number of points depends theoretically on the feature's radius and acquisition plane (the point density depends on the measurement's depth), a further post-processing step must be considered to allow a better generalization of the local point cloud quality. Consequently, it can be assumed that an optimal measurement is given when the maximum possible number of 3D points denoted as  $n_{max,3D}$  is obtained. The normalized number of points given as  $p_{norm}$  is calculated as follows:

$$p_{norm} = \frac{n_{points}}{n_{max,3D}}. \quad (5)$$

The maximum number of 3D points that can be acquired is equivalent to the total number of pixels of the camera image  $n_{max,2D}$ , hence:

$$n_{max,3D} \cong n_{max,2D}. \quad (6)$$

Assuming that an acquired 3D point  $t_{3D} \in \mathbb{R}^3$  can be associated with one pixel of the camera image, the 2D projection of  $t_{3D}$  can be approximated using a pinhole-camera model and the projection matrix  $M_s$ :

$$\begin{pmatrix} x_{pixel} \\ y_{pixel} \end{pmatrix} \sim M_s \cdot t_{3D}. \quad (7)$$

The projection matrix comprising the camera's intrinsic parameters can be obtained using a camera calibration procedure based on a calibration pattern [19].

Using Eq. 7 the 2D projection of the hollow cylinder at the feature's plane can be computed and the total number of pixels  $n_{max,2D}$ , which corresponds to  $n_{max,3D}$ , can be estimated. The 2D projection represents a 2D mask, which can be directly applied to the camera's image. Figure 9 shows the corresponding masks of the hollow cylinders from Fig. 8.

#### Exposure time

The exposure time  $t_{exp}$  is used as the target variable of the second GP model. Since most active sensors, such as the examples shown in Sect. 2.1.3, allow for the specific setting of an exposure time, additional pre-processing of this variable is unnecessary.

### 3.1.2 Input variables

In the following subsection, the input variables and their pre-processing steps for their synthesization are described.



**Local light intensity**

Like other works [3, 20], this study assumes that the external lighting conditions remain constant, the projector provides the dominant lighting source of the scene, and the projector’s configuration remains constant. Hence, a single fringe image is considered to quantify the light intensity around each feature at each measurement.

Similar to [3, 20], this work proposes to utilize the camera image of a stripe pattern to determine the light intensity as accurately as possible. In contrast to the related works, this research follows a local approach to quantify the image intensity around a feature using the 2D mask of the hollow cylinder (cf. Sect. 3.1.1). The average light intensity of the masked section denoted as  $i_{avg}$  characterizes the local illumination provided by the projector, which is given in an 8-Bit grayscale range  $i_{avg} \in [0, \dots, 255]$  (see Fig. 9).

**Relative position and orientation**

It is assumed that at least one reference frame of the sensor, e.g. at the tool center point (TCP) or the lens, is known. Hence, the rigid transformation between this sensor’s reference frame and a feature frame is utilized to characterize the spatial relationships between sensor and features for each measurement. The rigid transformation can be calculated using different methods, e.g., robot kinematics, feature extraction, surface-based registration, or photogrammetry. In the present study, the relative position is given in the sensor’s TCP by the following translation vector  $t_{tcp} = (x_{tcp}, y_{tcp}, z_{tcp})^T$  and the orientation by the Euler angles  $r_{tcp} = (w_{tcp}^x, p_{tcp}^y, r_{tcp}^z)^T$ . In addition, it is assumed that the sensor’s TCP lies in the middle plane of the sensor’s working volume. For example, Fig. 19 depicts the translation vectors between the sensor TCP and two features.

**3.1.3 Implementation**

The input and target variables are generated using a self developed processing pipeline considering following characteristics:

- Normalized number of points: The height of the cylinder and the radii tolerances were defined based on an internal defined norm as follows:  $h = 20$  mm,  $r_{tol,i} = 3$  mm, and  $r_{tol,o} = 5$  mm. The generation of the synthetic cylinder and the estimation of  $n_{points}$  was computed using the trimesh Library [21].
- Exposure time: The exposure time  $t_{exp}$  does not require any special pre-processing. Hence, the raw value in milliseconds (ms) was stored for each measurement.
- Position and orientation: The position and orientation of the door features are known and given in the door’s coordinate system. The rigid transformations between the sensor’s TCP and all features were calculated using the aligned transformation of each measurement. To increase

the accuracy of the alignment, most point clouds were aligned using the system’s photogrammetry capabilities. Otherwise, the robot’s kinematic model was utilized.

- Local light intensity: For the computation of the light intensity, the initial stripe pattern image was used for computing the image intensity. The initial stripe pattern is permanently projected and helps the manual operator to estimate the current camera exposure. The computation of the 2D masks, grayscale values, and projection matrix of the camera calibration was performed using the Python API of the OpenCV library [22].

An example dataset can be found in the supplementary files of this publication.

**3.2 Data acquisition**

The data acquisition phase represents a critical and resource-demanding step during the design of data-based models. Therefore, this section presents a thoughtful design of experiments (DoE) that utilizes a uniform sampling method of the sensor’s workspace.

**3.2.1 Design of experiments**

In the first step, a measurement strategy was developed to estimate the positions at which the Comet PRO AE sensor could acquire data within its workspace. This selection of the measurement positions followed the space-filling design strategy as suggested by [23], which advocates using an orderly grid pattern to evenly distribute the input values across the experimental range. Hence, the data collection positions were uniformly discretized using a grid along the

**Table 2** Performance overview of the QualityPrediction Model  ${}^qG^P$  considering different kernels and input variables

Kernel	Data set	Evaluation			
		Data set	R2 in %	MSE	RMSE in %
${}^qk_1$	$X_1^{train}$	$X_1^{val}$	79.53	0.030	17.32
		$X_1^{test}$	77.57	0.034	18.44
${}^qk_2$	$X_2^{train}$	$X_2^{val}$	96.03	0.006	7.75
		$X_2^{test}$	29.91	0.105	32.40
${}^qk_3$	$X_3^{train}$	$X_3^{val}$	94.78	0.008	8.94
		$X_3^{test}$	11.06	0.134	36.61
${}^qk_4$	$X_4^{train}$	$X_4^{val}$	<b>96.96</b>	<b>0.005</b>	<b>7.07</b>
		$X_4^{test}$	92.74	0.011	10.49
${}^qk_6$	$X_6^{train}$	$X_6^{val}$	97.45	0.004	6.32
		$X_6^{test}$	89.51	0.016	12.65

The bold values indicate the top performance of the trained kernel model on the test data

$x$  and  $y$  planes for three distinct depths ( $z$ -plane) to guarantee the collection of training data throughout the entire sensor's 3D workspace (see Table 7). The resulting measuring sensor positions were generated using the feature  $f_1$  as reference. Figure 21 depicts the discretized workspace for the measuring positions. Moreover, to minimize the data collection effort and for storage resource efficiency reasons, the exposure time acquisition ranges were defined based on a logarithmic scale enabling a finer discretization at low exposure time range. Moreover, the minimum and maximum exposure time range values were selected empirically for three different depth planes (near, middle, and far). Table 9 summarizes the considered spatial and exposure time ranges comprising 480 measurements at 48 different positions. The DoE regards a constant sensor orientation for all measurements. However, a minimal orientation variability was achieved by considering features with different normal surface orientations. The generated dataset is denoted as  $X$ .

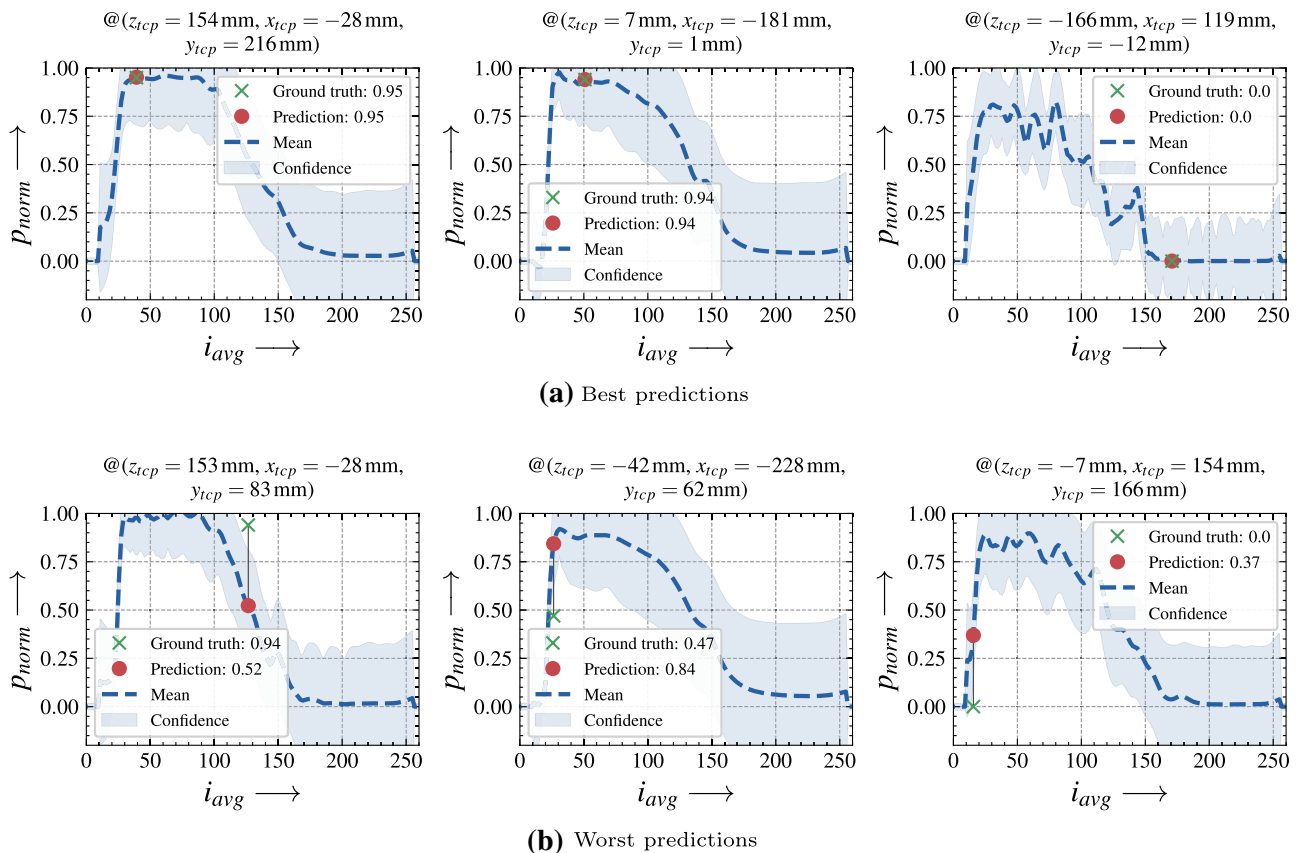
In the final step, the planned 480 measurements were acquired automatically using the integrated robot. To ease and automate the generation of the robot's path plan, the viewpoint planning framework introduced in our previous works [24, 25] was used. Moreover, the whole acquisition

process was automated using the C# and Python API provided by the proprietary software of ZEISS, i.e., Visio7 and colin3D. The acquisition of all measurements required approximately 105 min. The complete dataset size comprises 480 point clouds and 2D images with a resolution of 4 megapixels.

### 3.2.2 Training, validation and test datasets

The input and target variables were synthesized from the acquired dataset  $X$  for a set of four features denoted as  $\{f_1, \dots, f_4\}$  using the parameters presented in Sect. 3.1.3. The resulting 6D spatial range spanned by the training features is given in Table 9 in the Appendix. Figure 21 visualizes the spatial distribution of the training data in the Euclidean space.

All models are trained with 80% of all observations; let this training dataset be given as  $X^{train} \subseteq X$ . Moreover, the model metrics are computed using the remaining 20% ( $X^{val}$ ). Additionally, the trained models are evaluated against a test dataset  $X^{test} \subseteq X$  with unknown observations from different features denoted as  $\{f_1^{test}, \dots, f_4^{test}\}$  to evaluate the models'



**Fig. 11** Selection of the best (a) and worst (b) predictions of the 4D  $\mathcal{GCP}_4$  at different positions

performance. The total size of observations of each subset corresponds to  $N_{train} = 912$ ,  $N_{val} = 228$ , and  $N_{test} = 779$ . Figure 4 depicts all training and test features, whose radii lie between 3 and 30mm.

The dataset  $X^{test}$  includes individual observations (see Table 9) lying outside the 6D space range spanned by the training dataset. Hence, the models' interpolation and extrapolation capabilities can be individually investigated.

**Note:** Some of the acquired measurements in the front plane could not be photogrammetrically aligned. These measurements were aligned using a less accurate method based on the robot and measurement cell kinematic model. The summed alignment error of some of these *noisy* observations corresponded up to 5 mm. These alignment inaccuracies led to a miscalculation of some 2D and 3D masks and ultimately to a miscalculation of  $p_{norm}$ . This effect can be observed in some observations within the red square of Fig. 10 showing a considerable drop of  $p_{norm}$  under 80%. However, these misaligned measurements are included in the dataset to consider a more realistic and biased acquisition scenario. Additionally, it is important to note that the present study does not aim to evaluate the sensor's accuracy and that neither the DoE nor the input extraction pipeline is aligned to a standard for sensor qualification.

### 3.2.3 Preliminary data analysis

After synthesizing a dataset that can be used for the models' design, the next step regards selecting the input variables with the highest correlation to the model outputs. Hence, a preliminary prioritization of the inputs is performed based on the maximum information coefficient (MIC) of the acquired data to evaluate these correlations. The MIC matrix of the training dataset  $X^{train}$  is depicted in Fig. 20 and is used for supporting the incremental kernel design of each model. The MIC is often used as a measure of the correlation between features that exhibit partially nonlinear relationships, rather than relying on the more commonly used Pearson correlation coefficient. The interpretation of these values is individually addressed in the following subsections.

## 3.3 Quality prediction model

Given that light intensity is a characteristic used to estimate surface point depth utilizing image processing techniques, appropriate image illumination is vital to ensure accurate depth calculations. Moreover, measurement uncertainty in some sensors can lead to varying point cloud quality within the measurement volume, as noted by [26]. For these reasons, to predict the local quality of the point cloud, expressed by the number of points belonging to a specific feature, this

study employs a GP regression model that takes into account the local light intensity and spatial relation to the sensor. This subsection outlines an incremental and systematic approach for designing the kernel of a prediction model, which considers the local light intensity of the image and the spatial relationships between the sensor and a feature.

### 3.3.1 Formulation

According to Eq. 3, a simplified representation of the  $n$  dimensional GP quality regression model is given as follows:

$$g(p_{norm}) \sim {}^q\mathcal{GP}_n, \tag{8}$$

with

$${}^q\mathcal{GP}_n := GP(0, {}^qk_n(X^{train}(\mathbf{x}_n))),$$

and a zero mean GP (i.e.,  $m(X^{train}) = 0$ ).  ${}^qk_n$  denotes the model's kernel and  $\mathbf{x}_n$  the input vector.

### 3.3.2 Inputs selection

The MIC correlation values between all inputs and the model's output  $p_{norm}$  (see Fig. 20) are used to support the incremental kernel design of the quality regression model. As expected, the local light intensity  $i_{avg}$  has the highest correlation with  $p_{norm}$  followed by the spatial inputs.

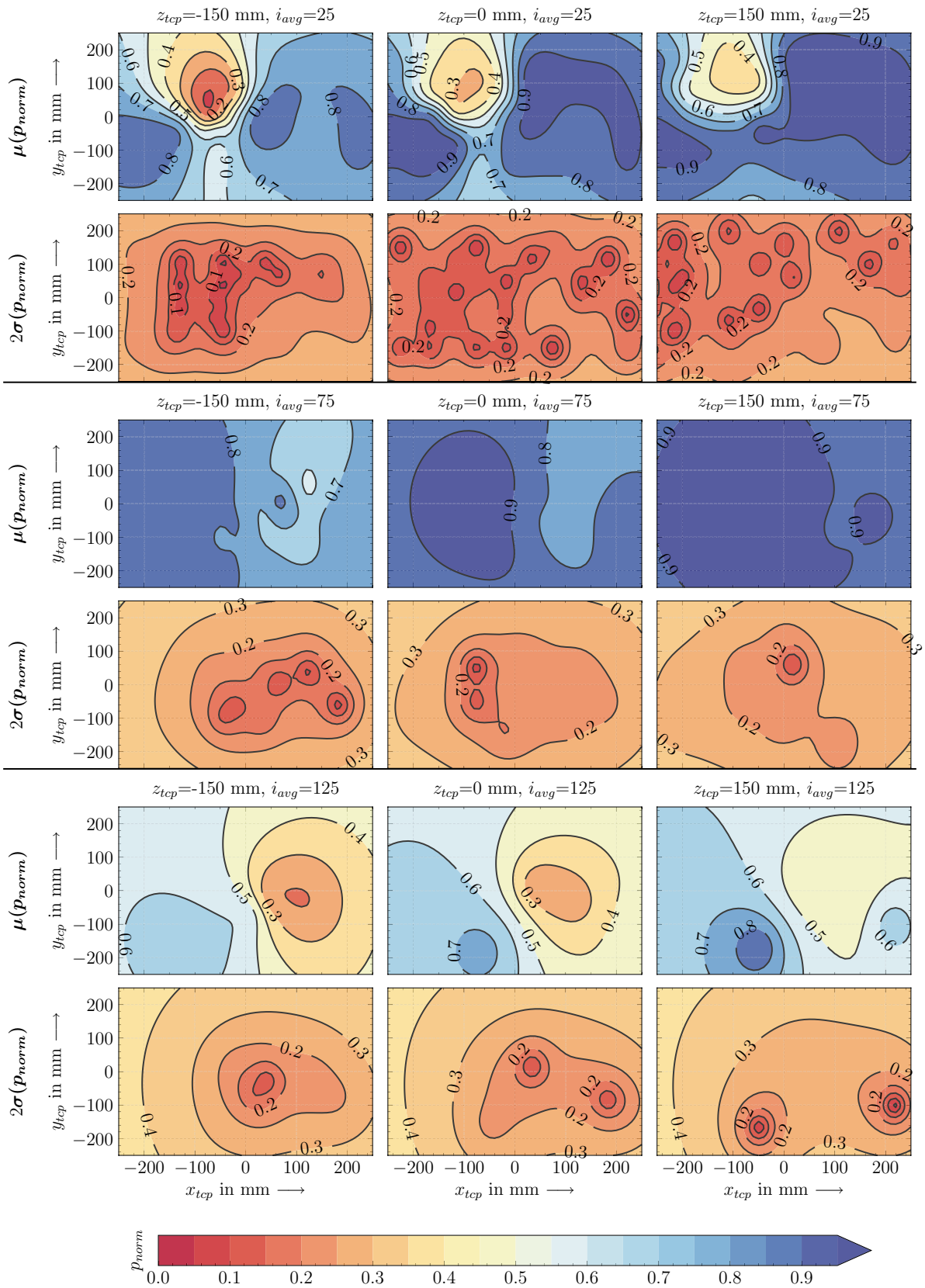
Moreover, the normalized number of points shows a considerable correlation to the features' radii  $r_f$  and positional inputs, i.e.,  $x_{icp}$ ,  $y_{icp}$ , and  $z_{icp}$ , suggesting a remaining dependency between  $p_{norm}$  and the individual features. A similar behavior can be observed with the inputs modeling the features' orientation ( $w_{icp}^x$ ,  $p_{icp}^y$ , and  $r_{icp}^z$ ), which also exhibit a high correlation to  $p_{norm}$ , between themselves and the positional inputs despite the fact that the sensor's orientation remained constant for all observations. For these reasons, our work considers primarily the positional inputs for the kernel design. Furthermore, the exposure time  $t_{exp}$  is disregarded because the information regarding the camera exposure is implicitly represented within the average light intensity  $i_{avg}$ .

### 3.3.3 Base kernel

Based on a comprehensive evaluation of different kernel combinations and a visual interpretation of the data between  $p_{norm}$  and  $i_{avg}$  (see Fig. 10), the following kernel was selected to be the most suitable for modeling the required correlations:

$${}^qk_n(X_n) := k_{RQ}(X_n) \cdot k_{CP,min}(i_{avg}) \cdot k_{CP,max}(i_{avg}) \tag{9}$$

where  $k_{RQ}(X_n)$  represents a RQ kernel.



**Fig. 12** Discretized visualization of the mean prediction  $\mu(p_{norm})$  and corresponding double standard deviations  $2\sigma$  of the 4D GP quality regression model  ${}^q\mathcal{GP}_4$  at three different depths  $z_{tcp} = [-150, 0, 150]$ mm (left, middle, right) for three different light intensities  $i_{avg} = [25, 75, 125]$  (top, middle, bottom) for the SLS Zeiss Comet Pro AE. The graphs demonstrate that the point cloud around a feature depends on the image intensity but also on the 3D relative position between a feature and the sensor. The trends confirm the nonlinear correlation between the point cloud quality, the image light intensity, and the spatial variables

$$k_{RQ}(\mathbf{x}, \mathbf{x}') = \sigma^2 \left( 1 + \frac{(\mathbf{x} - \mathbf{x}')^2}{2\alpha\ell^2} \right)^{-\alpha}.$$

First, the rational quadratic kernel  $k_{RQ,n}(X_n)$  (see Eq. 23) is used to model the general relationships between the normalized points and the local light intensity and positional inputs. Furthermore, an improvement of the model’s performance and smoother progress was observed using a change point kernel for a targeted kernel activation. The change point kernel  $k_{CP,min}(i_{avg})$  behaves like a step function, which changes from 0 to 1 at a minimal average brightness of  $i_{avg} \geq 10$  (cf. Fig. 10). Consequently, an intensity value below 10 automatically yields  $p_{norm} = 0$ . The reverse behavior applies for observations with an  $i_{avg}$  value over 255 with the change point kernel  $k_{CP,max}(i_{avg})$ .

### 3.3.4 1D kernel

In the first step a kernel including just the feature’s light intensity was designed. Hence, the model was trained considering the following dataset:

$$X_1^{train} := X^{train}(\mathbf{x}_1(i_{avg})) \tag{10}$$

and, consequently, the corresponding kernel (see Eq. 9):

$${}^qk_1 := k_{RQ}(X_1^{train}) \cdot k_{CP,min}(i_{avg}) \cdot k_{CP,max}(i_{avg}). \tag{11}$$

The GP denoted as  ${}^q\mathcal{GP}_1$  is given as follows:

$$g(p_{norm}) \sim {}^q\mathcal{GP}_1 \tag{12}$$

$${}^q\mathcal{GP}_1 := \mathcal{GP}(0, {}^qk_1). \tag{13}$$

The notation is analogously used for the following kernels consisting of more dimensions.

The 1D GP regression model given by Eq. 12 and all following GPs were trained and evaluated using the GPy library [27]. The model hyperparameters were calculated using an L-BFGS optimizer and a triple restart optimization. Table 8 provides an overview of the optimized hyperparameters.

The results of the evaluation using the corresponding datasets  $(X_1^{test}, X_1^{val})$  are summarized in Table 2. The  ${}^q\mathcal{GP}_1$

shows a solid performance and generalization for unseen features in view of the 1D kernel simplicity with an R2 score over 78%. However, the model’s predictions also demonstrate that the image exposure is insufficient for predicting the point cloud quality for all validating and testing features with a root mean square error (RMSE) up to 18%. This effect can also be observed within various observations of Fig. 10, suggesting that the same local image intensity can lead to different values of  $p_{norm}$ .

### 3.3.5 2D and 3D kernel

The second model considers a 2D feature vector integrating  $z_{tcp}$  due to its next highest correlation (cf. Fig. 20):

$$X_2^{train} := X^{train}(\mathbf{x}_2(i_{avg}, z_{tcp})). \tag{14}$$

The regression model is trained similarly to the 1D model, but the 2D rational quadratic kernel ( ${}^qk_{RQ,2}$ ) uses the automatic relevance determination (ARD) variant (see Eq. 24), which allows an individual optimization of the lengthscale for each input. Table 2 presents the performance results, which suggest at first glance an outstanding performance for the validation data with a model accuracy of R2 = 96% and an RMSE of 8%. However, to judge by the generalization accuracy on the unknown observations (RMSE = 32%), it can be assumed that the model was overfitted.

Furthermore, a 3D kernel integrating the spatial feature with the following highest correlation, i.e.  $x_{tcp}$ , was considered. However, the model showed a similarly poor performance on the test features (see Table 2).

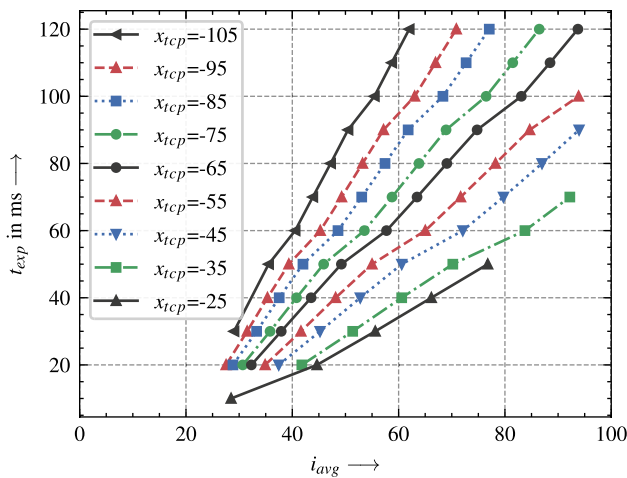
### 3.3.6 4D kernel

In the last step a 4D kernel integrating all spatial inputs was considered. The input matrix is given as follows:

$$X_4^{train} := X^{train}(\mathbf{x}_4(i_{avg}, z_{tcp}, x_{tcp}, y_{tcp})). \tag{15}$$

The 4D kernel model yielded remarkable results with an R2 of 92% and an RMSE of 10.5% for the test features confirming the model’s transferability and performance (see Table 2).

The graphs from Fig. 11 depict the 2D model’s trend of some of the best and worst predictions at different positions. The curves show the model’s mean prediction and the corresponding confidence interval ( $2\sigma$ , shaded region) for the whole light intensity range. On the one hand, Fig. 11b illustrates a selection of three of the worst predictions at the rising and falling slopes exhibiting that the model still has some difficulties making accurate predictions within these



**Fig. 13** Observations of the local light intensity  $i_{avg}$  and exposure time  $t_{exp}$  of the training dataset  $X^{train}$  with a  $p_{norm} > 75\%$  at different feature positions along the x-axis for  $y_{tcp} = 0$  mm and  $z_{tcp} = 0$  mm

areas. On the other hand, the best predictions in Fig. 11a confirm the overall performance and generalizability of the model. A constant and overall acceptable prediction accuracy can be observed mainly at the curve's plateau, where the highest number of points is expected.

A more comprehensive visualization of the model's mean prediction and corresponding double standard deviation in the x-y plane and light intensities at different sensor-depth planes in the z direction is given in the form of 3D surface plots in Fig. 12. The 3D model slices also visualize the nonlinear characteristic of the problem and the existence of local maximum values depending on the image intensity and the spatial variables. Additionally, these trends further confirm the present study's motivation showing the existence of local minima of the point cloud quality for different positions and light intensities.

### 3.3.7 6D Kernel

For the sake of completeness, a 6D kernel comprising the sensor orientation represented by the  $p_{tcp}^y$  and  $w_{tcp}^x$  was considered. The results (see Table 2) showed a non-statistical performance improvement and even a poorer performance than the 4D kernel on the unknown observations. Thus, a detailed analysis is not further considered. However, it should be noted that the training dataset considered feature surfaces with a relative sensor orientation up to  $13^\circ$ . The present study does not rule out that higher incidence angles may have a higher correlation with the point cloud quality.

### 3.3.8 Discussion

This section outlined and evaluated different multi-dimensional  ${}^q\mathcal{GP}$ s to predict the number of expected 3D points around a feature considering an increasing number of inputs. The best performance was delivered by a 4D model, whose characteristics and limitations are discussed more extensively below.

- Performance** The considered 1D model (see Sect. 3.3.4) demonstrated that the number of 3D points assigned to a feature could be estimated with an RMSE of 18% using just the local light intensity of a 2D image. These initial results are consistent with the findings of [16] and suggest that the quality of the point cloud could be optimized using only the image exposure, for example, by adjusting the sensor exposure time or the projector illumination parameters. However, the performed evaluation also revealed that the image exposure alone would be insufficient to predict accurately the number of expected surface points. Hence, the present study considered spatial variables and the design of a 4D kernel to learn the missing correlations. The evaluation of the 4D GP regression model showed that the prediction confidence could be increased to an RMSE of 10.5% ( $R^2 = 93\%$ ), if the relative position between features and sensor are considered. Moreover, the  ${}^q\mathcal{GP}_4$  demonstrated a high computational efficiency, e.g., 100 predictions were calculated under 19 ms.
- Inputs** The target variable  $p_{norm}$  and input variable  $i_{avg}$  showed a significant validity for modeling the point cloud quality and image intensity independent of the feature dimensions. However, Fig. 20 shows that  $p_{norm}$  and  $i_{avg}$  have a correlation to the features' radii. Hence, it can be hypothesized that the model's performance and generalizability could be improved by considering further input variables, e.g., a discretization of the 2D mask, standard deviation or median of the light intensity.
- Dataset Size:** It is well known that the performance of data-driven approaches depends on the available training dataset size. For this reason, the  ${}^q\mathcal{GP}_4$  was retrained with different training dataset sizes for analyzing the influence on the model's performance. Figure 22 shows the results of this analysis based on the test dataset  $X_4^{test}$ . The trend shows that the model's accuracy ( $R^2 = 92\%$ ,  $RMSE = 11\%$ ) starts converging with a dataset size of  $\approx 450$  observations. Therefore, it is reasonable to assume that collecting more data from the same training features would not improve the performance of the model. Furthermore, this analysis also demonstrates that an acceptable prediction accuracy can be obtained

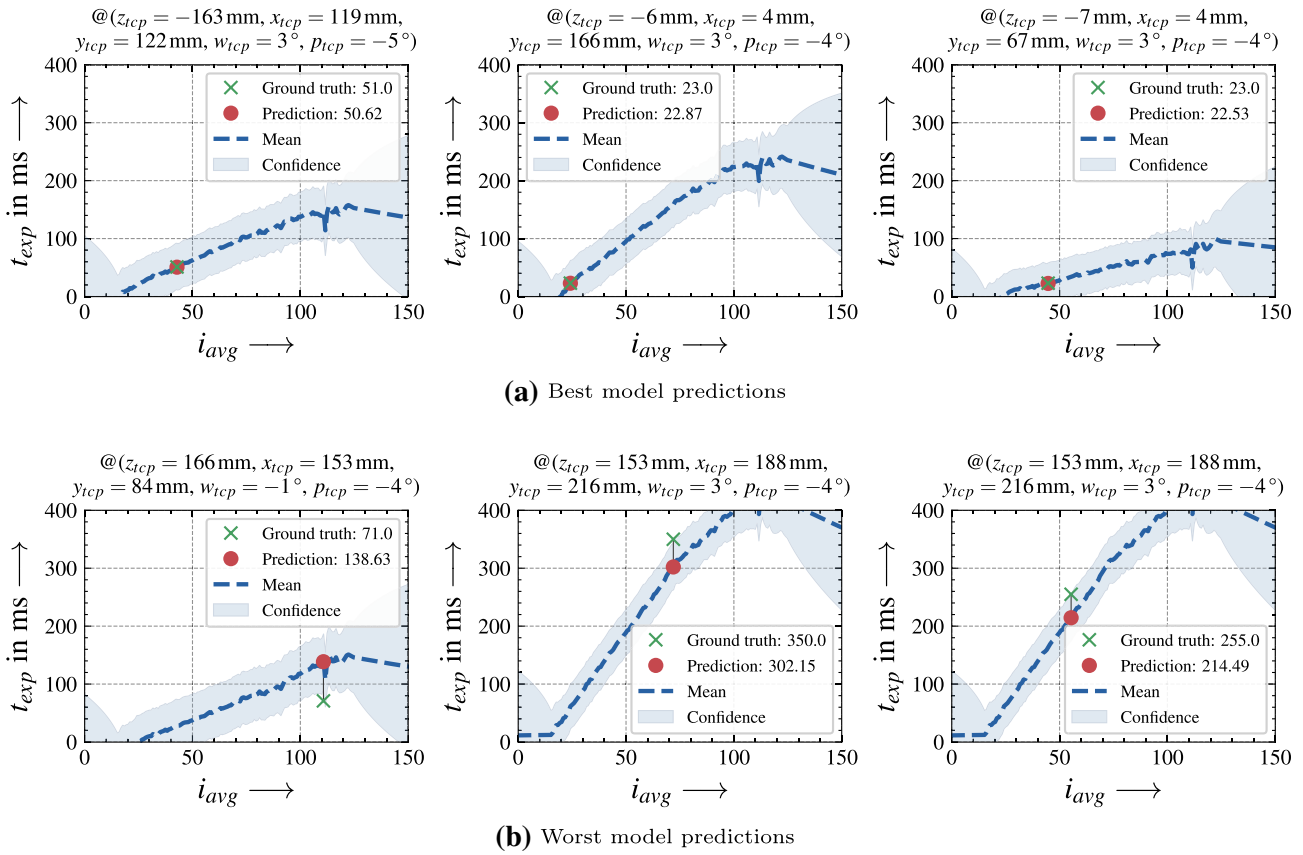


Fig. 14 Selection of the best (top) and worst (bottom) predictions of the 6D  $e\mathcal{GP}_6$  at different positions

with a reduced number of observations, e.g., an RMSE of 13.8% with 114 observations. Nevertheless, it should be noted that the spatial distribution should still be maintained to obtain similar results.

- **Kernel design** The training loops of the models revealed showed that the kernel and its hyperparameters might considerably influence the model’s performance. In consideration of the findings, it can be concluded that the proposed kernel and hyperparameters are capable of adequately modeling the necessary system correlations. However, it is possible that a different set of kernel hyperparameters or an alternate kernel could yield a superior performance.

### 3.4 Exposure time prediction model

This section outlines the design of a GP to estimate the camera’s exposure time for a given local light intensity around a feature and its spatial relationships to the sensor.

#### 3.4.1 Preliminary dataset analysis

Contrary to the  ${}^q\mathcal{GP}$ , the performance and training efficiency of the  $e\mathcal{GP}$  can be increased by considering exclusively observations leading to a satisfactory point cloud quality. Hence, taking advantage of the measurement quality quantification expressed by  $p_{norm}$ , the proposed models are trained and evaluated considering observations with  $p_{norm} \geq 75\%$ . The datasets’ sizes are then reduced:  $N_e^{train} = 421$ ,  $N_e^{val} = 106$ , and  $N_e^{test} = 508$ . The train, validation, and test datasets are denoted as follows  $eX^{train}$ ,  $eX^{val}$ , and  $eX^{test}$ .

Figure 13 shows the trend between the exposure time and local light intensity of feature  $f_1$  at different positions considering a translation in the x-axis alone. At first glance, it can be assumed that a linear model would be sufficient to model the correlation between  $i_{avg}$  and  $t_{exp}$  confirming the validity of the approximation of [3, 15]. However, this reduced set of observations, which considers the shift of a feature in only one direction, shows that

exposure time has a considerably high correlation with spatial inputs. This insight confirms the motivation of our work that spatial relationships must be considered for selecting an appropriate exposure time. Recalling the qualitative distribution of the projector illumination from Fig. 1, a nonlinear correlation is to be expected, suggesting the necessity of a more complex model.

### 3.4.2 Formulation

Aligned to the generic definition of GPs (see Eq. 3), the regression model for estimating the exposure time is given:

$$g(t_{exp}) \sim {}^e\mathcal{GP}_n \tag{16}$$

with

$${}^e\mathcal{GP}_n := \mathcal{GP}(0, {}^e k_n({}^e X^{train}(\mathbf{x}_n))). \tag{17}$$

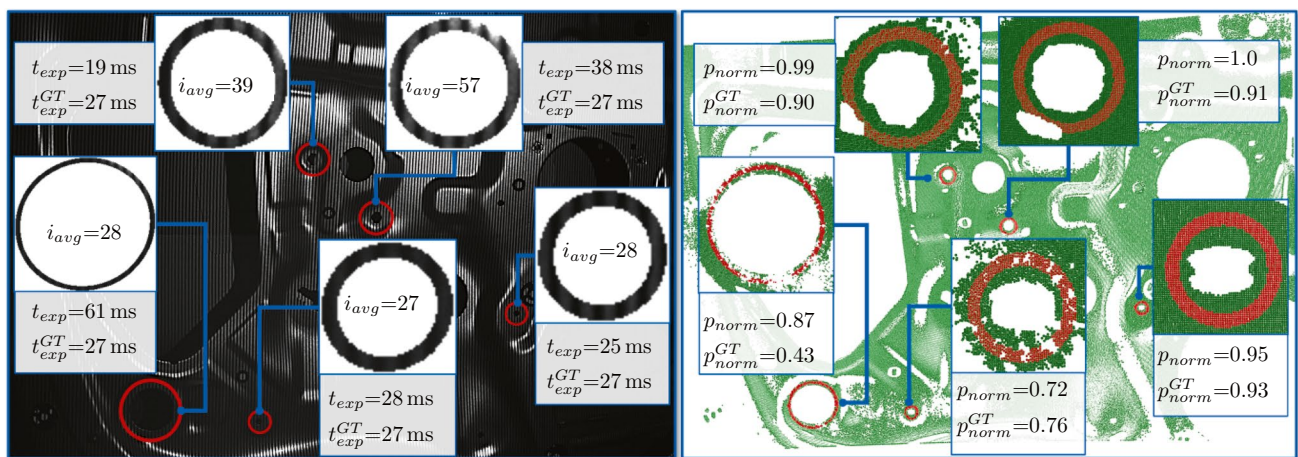
### 3.4.3 Input selection

Based on the MIC values from Fig. 20, the kernel design is supported by the correlation magnitude between the exposure time  $t_{exp}$  and the rest of the inputs. Moreover, it should be noted that the input  $z_{tcp}$  has an overestimated correlation with the  $t_{exp}$  due to the different exposure time ranges considered at each acquisition depth plane (cf. Table 9). Once again the rotational variables are not considered in

**Table 3** Performance overview of the Exposure Time Prediction Model  ${}^e\mathcal{GP}$  considering different kernels and input variables

Kernel	Data set	Evaluation			
		Data set	R2 in %	MSE	RMSE in ms
${}^e k_1$	${}^e X_1^{train}$	${}^e X_1^{val}$	15.87	6586.35	81.16
		${}^e X_1^{test}$	7.41	5644.60	75.13
${}^e k_2$	${}^e X_2^{train}$	${}^e X_2^{val}$	31.47	5365.17	73.25
		${}^e X_2^{test}$	31.66	4166.19	64.55
${}^e k_3$	${}^e X_3^{train}$	${}^e X_3^{val}$	63.96	2821.29	53.12
		${}^e X_3^{test}$	25.00	4572.26	67.62
${}^e k_4$	${}^e X_4^{train}$	${}^e X_4^{val}$	88.38	909.33	30.16
		${}^e X_4^{test}$	17.76	5013.19	70.80
		${}^e X_4^{test,i}$	49.55	2478.58	49.79
${}^e k_4^+$	${}^e X_4^{train+}$	${}^e X_4^{val+}$	95.32	255.07	15.97
		${}^e X_4^{test,i}$	80.22	1059.81	32.55
${}^e k_6^+$	${}^e X_6^{train+}$	${}^e X_6^{val+}$	95.72	233.38	15.28
		${}^e X_6^{test,i}$	84.01	856.40	29.26
${}^e k_6^{++}$	${}^e X_6^{train++}$	${}^e X_6^{val++}$	97.55	85.86	9.27
		${}^e X_6^{test,i}$	86.25	736.42	27.14
${}^e k_6^*$	${}^e X_6^{train++}$	${}^e X_6^{val++}$	98.50	56.51	7.52
		${}^e X_6^{test,i}$	95.88	220.63	14.85
		${}^e X_6^{val++}$	<b>98.91</b>	<b>38.32</b>	<b>6.19</b>
${}^e k_6$	${}^e X_6^{train++}$	${}^e X_6^{val++}$	96.80	171.25	13.09
		${}^e X_6^{test,i}$	49.47	1897.95	43.57
${}^e k_6^{lin}$		${}^e X_6^{val++}$	49.47	1897.95	43.57

The bold values indicate the top performance of the trained kernel model on the test data



**Fig. 15** Performance overview of the GP models compared to the ground truth (GT) values for predicting the exposure time using the  ${}^e\mathcal{GP}_6$  (left) and for predicting the point cloud quality using the  ${}^a\mathcal{GP}_4$  (right) with five different features



a first step due to their high correlation to the translational features.

### 3.4.4 Base kernel

Based on an exhaustive benchmarking of various kernel combinations, the exponential kernel  $k_{exp}$  (see Eq. 22) was discovered to perform slightly best than the well-known SE kernel and was best suited for learning the necessary correlations to predict the exposure time:

$${}^e k_n({}^e X_n) = k_{exp}({}^e X_n). \tag{18}$$

with

$$k_{exp}(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{(\mathbf{x} - \mathbf{x}')^2}{\ell^2}\right).$$

### 3.4.5 1D kernel

In a first step, a 1D kernel considering just the light intensity  $i_{avg}$  was evaluated. The input matrix is given as follows:

$${}^e X_1^{train} := {}^e X^{train}(\mathbf{x}_1(i_{avg})). \tag{19}$$

The hyperparameters of all models are also found in Table 8. Table 3 shows the model’s evaluation results considering the validation and test datasets. The model’s poor performance on the test data (R2 = 16%, RMSE = 81ms) confirms the visual interpretation of Fig. 13 corroborating that the local light intensity is ineffective for predicting the corresponding exposure time and that a linear model alone would be ineffective. Consequently, the model’s accuracy on the test dataset confirms this thesis and suggests that a more complex model including spatial inputs must be applied.

**Table 4** Overview of the evaluation of the  ${}^q \mathcal{GP}_4$  and  ${}^e \mathcal{GP}_6$  considering different lighting scenarios

Model	Evaluation			
	Dataset	R2 in %	MSE	RMSE
${}^q \mathcal{GP}_4$	$X_4^{I,I}$	97.11	0.006	7.75 %
	$X_4^{I,II}$	97.37	0.005	7.07 %
	$X_4^{I,III}$	97.33	0.005	7.07 %
${}^e \mathcal{GP}_6$	$X_6^{I,I}$	87.11	144.69	12.03 ms
	$X_6^{I,II}$	82.93	191.62	13.84 ms
	$X_6^{I,III}$	80.41	219.90	14.83 ms

### 3.4.6 2D kernel

Based on the previous observations, in the next step, a model with a 2D and a 3D kernel (without ARD) comprising the spatial inputs  $z_{tcp}$  and  $x_{tcp}$  is evaluated. Although a slight performance improvement for both models was achieved, the prediction accuracy remains unsatisfactory, cf. Table 3.

### 3.4.7 4D kernel

In the next step, a 4D kernel  ${}^e k_4({}^e X_4)$  composed analogously to the 2D kernel and considering all positional inputs, i.e.,  $x_{tcp}$ ,  $y_{tcp}$ , and  $z_{tcp}$  was evaluated. The evaluation results from Table 3 exhibit an overall improvement of the model’s performance compared to the 2D counterpart. However, the model’s accuracy between the validation (R2 = 88%, RMSE = 30.1ms) and test (R2 = 18%, RMSE = 70.8ms) dataset continues exhibiting major discrepancies requiring a detailed analysis of the data.

Given the qualitative and nonlinear trend of the light distribution and the door’s surface topology (see Fig. 1), it seems logical that the training dataset comprising four features fails to represent the high variability and complexity of the car door surface. Hence, in a further step the 4D model was reevaluated considering exclusively the interpolating test dataset denoted as  ${}^e X^{val,i}$ . The results confirmed the proposed thesis and showed a considerable improvement in the model’s performance (R2 = 50%, RMSE = 49.8ms) within the interpolated range. Note that hereafter the evaluation of the following models considers only observations within the interpolated spatial range of the training dataset.

Moreover, these findings confirm the high correlation between the exposure time and the spatial correlations and suggest that the model’s accuracy and generalization should be improved by increasing the spatial variability within the training observations. Hence, the training dataset was extended by considering five additional features ( $f_{5-9}$ ). The extended dataset denoted as  ${}^e X^{train+}(f_{1-9})$  considers an increased number of observations with  $N_e^{train+} = 758$  and  $N_e^{val+} = 190$ . As expected, the extended dataset contributed to a considerable improvement in the model’s accuracy (R2 = 80%, RMSE = 32.7ms). Nevertheless, the model’s performance difference between test and validation data is still significant suggesting that a more complex model needs to be considered.

### 3.4.8 6D kernel

In the last step, a 6D kernel ( ${}^e k_6^+$ ) considering the local light intensity and three positional plus two rotational inputs is

investigated. The training input matrix for the 6D kernel is given as follows:

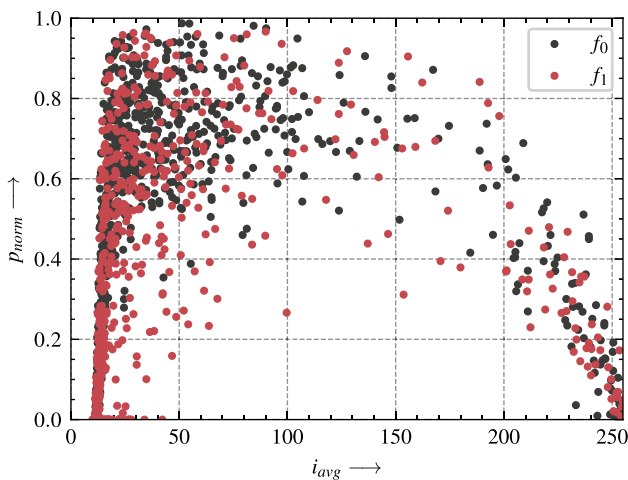
$${}^e X_6^{train+} := {}^e X^{train+}(\mathbf{x}_6(i_{avg}, x_{icp}, y_{icp}, z_{icp}, \dots, w_{icp}^x, p_{icp}^y)). \quad (20)$$

The results of the 6D model exhibit a slightly better performance ( $R2 = 84\%$ ,  $RMSE = 29\text{ms}$ ) on the test dataset. Given the high correlation between exposure time and spatial inputs, an additional dataset with a finer discretization of the exposure time and displacement in the  $x$  and  $y$  axes was considered in a next step for training the model. Table 9 summarizes the considered DoE. The extended dataset is denoted as  ${}^e X_6^{train++}$  with  $N_e^{train++} = 1936$  and  $N_e^{val++} = 485$  observations. Once again, the model's accuracy and, in particular, the generalization performance could be improved to an RMSE below 27ms.

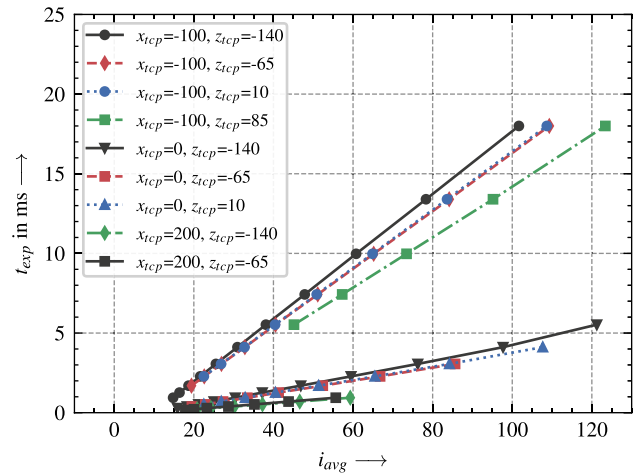
In a next step, an exponential kernel ( ${}^e k_6^*$ ) with ARD was considered. The kernel showed a significant improvement on the generalization performance yielding an RMSE of 15ms. In the last step, a modification of the base kernel (18) was considered by using the multiplication of two exponential kernels, one for the light intensity and one for the spatial inputs:

$${}^e k_6 := {}^e k_6({}^e X_6^{++}) = k_{exp}({}^e X_1^{++}(i_{avg})) \cdot k_{exp}({}^e X_5^{++}(x_{icp}, y_{icp}, z_{icp}, \dots, w_{icp}^x, p_{icp}^y)). \quad (21)$$

The modified kernel (21) yielded the best performance, i.e.,  $R2 = 97\%$  and an RMSE of 13.1ms, for unknown observations.



**Fig. 16** Trend of the normalized number of points and local light intensities for different features



**Fig. 17** Observations of the local light intensity  $i_{avg}$  and the exposure time  $t_{exp}$  of the training dataset  ${}^e X^{train}$  with a  $p_{norm} > 75\%$  and  $y_{icp} = 0$  mm

A graphical evaluation of the model's performance is provided in Fig. 14, which depicts three of the best and worst predictions including a confidence interval of 95% ( $2\sigma$ ). The best predictions visualized in Fig. 14a demonstrate that the model can accurately predict the exposure time at various positions for different light intensities. On the contrary, the worst trends shown in Fig. 14b are

**Table 5** Overview of the validation and test results of the  ${}^q \mathcal{GP}^{s_2}$  and  ${}^e \mathcal{GP}^{s_2}$  of the 3D active stereo sensor

Training	Evaluation			
	Dataset	R2 in %	MSE	RMSE in
Model: ${}^q \mathcal{GP}^{s_2}$				
${}^q k_1^{s_2} (X_1^{s_2,train})$	$X_1^{s_2,val}$	69.37	0.039	19.75%
	$X_1^{s_2,test}$	68.80	0.042	20.49%
${}^q k_4^{s_2} (X_4^{s_2,train})$	$X_4^{s_2,val}$	<b>89.16</b>	<b>0.014</b>	<b>11.83%</b>
	$X_4^{s_2,test}$	63.96	0.059	24.29 %
	$X_4^{s_2,test,i}$	84.72	0.022	14.83%
${}^q k_6^{s_2} (X_6^{s_2,train})$	$X_6^{s_2,val}$	91.27	0.011	10.49%
	$X_6^{s_2,test,i}$	79.48	0.028	16.73%
Model: ${}^e \mathcal{GP}^{s_2}$				
${}^e k_1^{s_2} ({}^e X_1^{s_2,train})$	${}^e X_1^{s_2,val}$	46.73	12.80	3.58 ms
	${}^e X_1^{s_2,test}$	46.90	13.37	3.66 ms
${}^e k_4^{s_2} ({}^e X_4^{s_2,train})$	${}^e X_4^{s_2,val}$	96.08	0.94	0.97 ms
	${}^e X_4^{s_2,test}$	54.94	11.35	3.37 ms
${}^e k_6^{s_2} ({}^e X_6^{s_2,train})$	${}^e X_6^{s_2,val}$	<b>97.54</b>	<b>0.59</b>	<b>0.77 ms</b>
	${}^e X_6^{s_2,test}$	65.72	8.63	2.94 ms
	${}^e X_6^{s_2,test,i}$	97.44	0.62	0.79 ms

The bold values indicate the top performance of the trained kernel model on the test data

mainly in the observations of the more distant acquisition plane. This behavior was to be expected, considering that in the more distant plane, the illumination provided by the projector loses its dominance, and other lighting sources might more easily influence the image exposure. Nevertheless, the model generally shows a high prediction confidence and demonstrates that even the worst predictions lie within the  $2\sigma$  interval.

A more comprehensive visualization of the model’s trend and corresponding double standard deviation is provided by Fig. 24, which shows 3D surface plots in the  $x$ – $y$  plane for a constant sensor orientation with three different light intensities at different depths. Similar to the trend of  ${}^c\mathcal{GP}_4$ , these graphs visualize and quantify the nonlinearity behavior of the problem. Furthermore, the graphs also agree with the qualitative light distribution of the SLS shown in Fig. 1, confirming that the right side (positive  $x$ -axis) is more overexposed than the left side.

### 3.4.9 Discussion

This section outlined the design of a Gaussian Process regression model for predicting the exposure time of a 3D measurement using a feature’s local image exposure and its relative displacement and orientation to the sensor. Moreover, the incremental and systematic design of the model’s kernel contributed to a better general understanding of the correlations between the exposure time, input variables, dataset size, and data entropy. A more detailed discussion regarding these correlations follows below.

- **Performance** Our 1D model analysis showed that the light intensity alone was insufficient to predict an adequate exposure time ( $R^2 = 7\%$ ,  $RMSE = 75$  ms). Hence, by increasing the model’s complexity and integrating spatial inputs the 6D model (cf. Sect. 3.4.8) demonstrated that the prediction accuracy ( $R^2 = 97\%$ ) could be considerably improved up to an  $RMSE$  of 13 ms for unseen observations. In addition, a model with a 6D linear kernel ( ${}^e k_6^{lin}$ ) was trained for performance comparison with previous research proposing linear models. The results showed that the performance of the linear model ( $R^2 \approx 50\%$  and an  $RMSE = 44$  ms) was insufficient to compensate the inhomogeneous and nonlinear surface illumination.
- **Inputs** Our study revealed that the exposure time exhibited a high correlation to the spatial inputs. Based on our incremental kernel design, this observation could be corroborated, confirming that the model requires a dataset comprising a high variability of exposure times combined with a good spatial distribution of the observations. Our experiments showed that the model’s

accuracy held well for observations within the training space (interpolation), which covered the sensor’s workspace and considered a relative sensor orientation up to  $10^\circ$  between sensor and feature’s normal plane. A more exhaustive acquisition and comprehensive analysis to evaluate the model’s performance and correlation for higher sensor inclinations are required to be further investigated.

- **Dataset size** Furthermore, the correlation between dataset size and model performance was analyzed for efficiency purposes. The evaluation results of the  ${}^c\mathcal{GP}_6$  trained with a different number of test observations are displayed in Fig. 23. In contrast to the quality prediction model, the exposure time model requires more observations throughout the sensor’s workspace. Hence, for replication purposes and efficient and effective data acquisition, we recommend a thoughtful design of the probing object considering a heterogeneous spatial distribution of the features with different surface orientations.

### 3.5 Summary

This section outlined and investigated the design of two multi-dimensional GPs trained on the basis of a local image exposure and spatial relationships to predict the point cloud quality around a feature and to estimate the corresponding exposure time. Both models showed promising performance

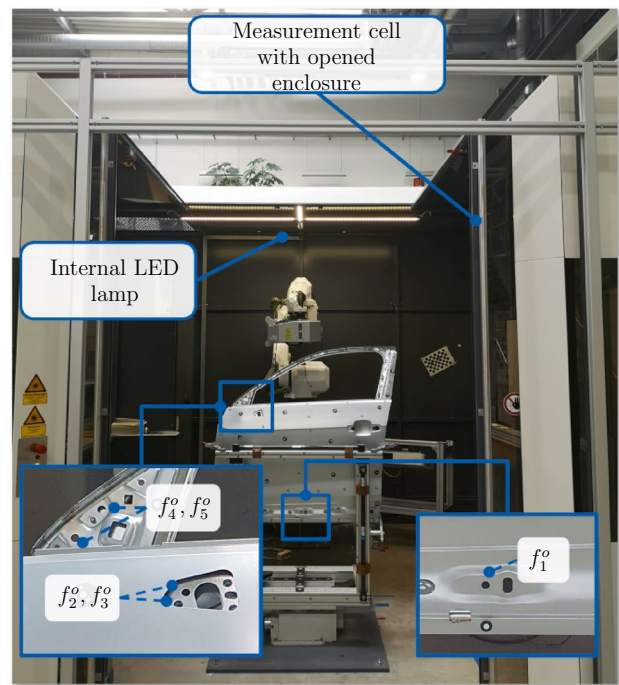


Fig. 18 Overview of the lighting conditions and outdoor features

and demonstrated that spatial variables are necessary for increasing the models' prediction accuracy. A visual representation of the performance of both models using five different test features is depicted in Fig. 15. The image on the left displays a 2D sensor image taken with an exposure time of 27ms, while the image on the right shows the corresponding point cloud. The selected exposure duration and resulting point cloud quality set the ground truth (GT) values as the benchmark to qualitatively evaluate model performance for this empirical measurement. In this scenario, the GT values for the point cloud quality model are the individual normalized points for each feature, while for the exposure regression model, the chosen exposure time  $t_{exp}^{GT}$  denotes the GT value. The exposure time predictions on the left demonstrate that the GP  ${}^e\mathcal{GP}_6$  can accurately predict the exposure time for different features and image intensities. The performance of the GP  ${}^q\mathcal{GP}_4$  is confirmed on the right side for estimating the expected point cloud quality at different locations. Moreover, the left bottom feature can be considered as an outlier showing a poorer prediction for both models. However, it must be noted that this feature lies outside the training range due to its surface orientation. Hence, a model's performance improvement should be expected by extending the dataset with different sensor orientations.

Note that some of the results, including the proposed kernels, are to some extent limited to the considered experimental setup. Hence, to evaluate the transferability of our

approach to other systems, Sect. 4 presents a further evaluation of both models regarding their transferability to different surfaces, sensitivity to external conditions, and generalizability of kernel design for an alternative active sensor.

## 4 Extended evaluation

This section considers a comprehensive evaluation of the 4D  ${}^q\mathcal{GP}$  and the 6D  ${}^e\mathcal{GP}$  outlined in Sect. 3 considering the models sensitivity to external lighting conditions, transferability to other active sensors, and performance on alternative surface finishes.

### 4.1 External lighting conditions

Within our work it is assumed that the stripe projector of an SLS can be considered the predominant lighting source within the acquisition environment. This requirement can be well fulfilled by the employed measurement setup which comprised a closed measurement cell with tinted windows and the band-pass filter of the sensor (cf. Sect. 2.1). However, the performance of the models may be affected under different lighting conditions. Hence, the models' sensitivity to external lighting sources is evaluated in this subsection.

**Table 6** Evaluation of the 4D quality prediction model ( ${}^q\mathcal{GP}$ ) and the 6D exposure time prediction model ( ${}^e\mathcal{GP}$ ) for the SLS and active stereo sensor ( $s_2$ ) for different surface finishes

Model	Evaluation				
	Dataset	Features	R2 in %	MSE	RMSE
Model ${}^q\mathcal{GP}$					
${}^q\mathcal{GP}_4$	$X_4^{out}$	$f_{1-5}^o$	88.12	0.018	13.42%
		$f_1^o$	93.69	0.010	10.00%
		$f_{2-3}^o$	91.26	0.015	12.25%
		$f_{4-5}^o$	77.84	0.026	16.12%
${}^q\mathcal{GP}_4^{s_2}$	$X_4^{s_2,out}$	$f_{1-5}^o$	70.59	0.040	20.00%
		$f_1^o$	84.48	0.017	13.04%
		$f_{2-3}^o$	84.31	0.018	13.42%
		$f_{4-5}^o$	61.73	0.058	24.08%
Model ${}^e\mathcal{GP}$					
${}^e\mathcal{GP}_6$	${}^eX_6^{out}$	$f_{1-5}^o$	75.66	286.28	16.92 ms
		$f_1^o$	63.67	436.48	20.89 ms
		$f_{2-3}^o$	67.07	385.49	19.63 ms
		$f_{4-5}^o$	90.93	94.35	9.71 ms
${}^e\mathcal{GP}_6^{s_2}$	${}^eX_6^{s_2,out}$	$f_{1-5}^o$	89.81	2.60	1.61 ms
		$f_1^o$	95.01	0.90	0.95 ms
		$f_{2-3}^o$	85.56	3.56	1.89 ms
		$f_{4-5}^o$	84.54	4.14	2.03 ms

#### 4.1.1 Dataset

The DoE considered three different use-cases and the acquisition of their corresponding datasets with following lighting conditions: I) equal lighting as training, i.e., closed measurement cell without external lighting ( $X^{I,I}$ ), II) closed measurement cell with internal LED light of the AI-Box ( $X^{I,II}$ ), and III) open measurement cell with the typical external lighting of a production hall ( $X^{I,III}$ ). Figure 18 depicts the internal LED light and the opened measurement cell as considered for the third scenario. The three evaluation datasets considered just one single sensor position and five exposure times for each lighting scenario (see Table 9). The models were evaluated only with the features  $f_{1-3}^{test}$ , because  $f_4^{test}$  could not be captured from the selected sensor position.

#### 4.1.2 Evaluation and discussion

The sensitivity of the  ${}^q\mathcal{GP}_4$  (see Sect. 3.3.6) and  ${}^e\mathcal{GP}_6$  (see Sect. 3.4.8) to external light sources was evaluated with the acquired datasets of the previously mentioned lighting scenarios. The results are summarized in Table 4.

On the one hand, the  ${}^q\mathcal{GP}_4$  showed an encouraging and constant performance with statistically insignificant differences for all lighting scenarios. The results are also in line with the previous validating results from Table 2 and confirm the robustness and usability of the  ${}^q\mathcal{GP}_4$  to some measure for different lighting conditions.

On the other hand, the exposure estimation model  ${}^e\mathcal{GP}_6$  showed a more significant variance for the different lighting scenarios. As could have been anticipated, the best performance is observed in the scenario I, which corresponds to the same lighting conditions as for the training dataset. On the contrary, the scenarios II and III consider a more illuminated environment compared to the first scenario and exhibit an acceptable overestimation of the exposure time around 2 ms to 3 ms. From these results, it can be concluded that the image exposure is not significantly affected by external light sources and the projector can be assumed to be the dominant lighting source. Consequently, it can be assumed that the prediction accuracy of the  ${}^e\mathcal{GP}_6$  should hold for different lighting conditions as long as the projector can be considered the dominant light source.

Even though these experiments demonstrate the usability of the proposed models under different lighting conditions, a more exhaustive design of experiments comprising more observations and different light sources is necessary to assess the models' light sensitivity and prediction limitations properly.

## 4.2 Transferability to an active stereo sensor

In the following subsection, the transferability of the kernel design of the  ${}^q\mathcal{GP}_4$  and  ${}^e\mathcal{GP}_6$  using a different active system is evaluated. The combination of a dot projector and the *rc\_visard 65* can be classified as a 3D active stereo sensor system (see Sect. 2.1). Such configurations are not rare in industrial applications and even encouraged to improve the point cloud density for acquiring homogeneous surfaces. Similar to the SLS, it is assumed that the dot projector is the constant and predominant active lighting source. Hence, our study evaluates the transferability of the proposed inputs and kernels with a 3D active stereo system.

The GP models of the stereo sensor are denoted as  ${}^q\mathcal{GP}^{s_2}$  and  ${}^e\mathcal{GP}^{s_2}$ . Note that the stereo system GP models were trained and evaluated independently of the SLS models.

### 4.2.1 Dataset

Similar to the proposed DoE from Sect. 3.2, the uniformly discretization of the measurement volume was also applied for this dataset acquisition. Taking into account the storage size of one measurement and the acquisition speed of the stereo sensor, a finer discretization of the acquisition space was used (see Table 9).

The models were trained and validated considering the same features as for the SLS. Due to the high number of acquired observations and based on the findings associated to the influence of the dataset size (cf. Figs. 22 and 23), just 20% (2693) of the observations were used for training and the rest for validation. The datasets are denoted  $X^{s_2,train}$ ,  $X^{s_2,val}$ , and  $X^{s_2,test}$ .

The models' input and target variables were generated analogously as suggested in Sect. 3.1.3. In view of the sensor's short baseline (65 mm) and based on empirical experiments, it was assumed that the image light intensity is similar for both cameras. Hence, the  $i_{avg}$  was calculated just for the left camera of the *rc\_visard 65*.

### 4.2.2 Quality prediction model

Figure 16 illustrates the relation between the number of normalized points ( $p_{norm}$ ) and the local light intensity ( $i_{avg}$ ) for two features. The curve shows, similar to the SLS, a nonlinear decrease after a certain number of points has been reached. Unlike the SLS, the stereo sensor can still capture some surface points for higher exposures. Moreover, these observations demonstrate that the same light intensity can also lead to a different number of points. Hence, it can be assumed that a more accurate prediction should be expected if spatial inputs are considered.

In the first step, a 1D model was trained using the same kernel (9) to evaluate the influence of the image exposure represented by  $i_{avg}$ . The evaluation results (see Table 5) yielding an RMSE approx. of 21% for the unseen observations indicate a similar correlation between the number of points and the local light intensity compared to the SLS. Hence, the model's performance could be increased to an RMSE of 15% by considering a 4D model integrating positional inputs and observations within the training spatial range ( $X_4^{s_2,test,i}$ ). Note that if observations outside the spatial range (extrapolation) are considered, the model accuracy falls to an RMSE of 24%. In general these insights further confirm the high correlation of the spatial inputs to the quality of the point cloud. A more detailed visualization of the model trend and double standard deviation at different light intensities and depths is given in Fig 25. The graphs show, similar to the SLS model (see. Fig. 12), a nonlinear trend with multiple local optimum values depending on the spatial inputs and light intensity.

Furthermore, a detailed analysis showed that the model performed significantly better (up to an RMSE of 8%) on features with bigger radii agreeing with the average performance of the SLS model. The poorer accuracy on the smaller features can be easily justified given that the stereo sensor has a lower resolution and point cloud density. Since the camera's resolution directly affects the accuracy of the 2D

and 3D masks, the computation of  $p_{norm}$  and  $i_{avg}$  is consequently less accurate for features with small radii.

### 4.2.3 Exposure time model

Similar to the SLS, the observations on the graph displayed in Fig. 17 show mostly a linear trend between the light intensity and the exposure time for different positions. Moreover, it can also be appreciated that the curves have a high correlation with the feature's position. For this reason, it is assumed that a  ${}^e\mathcal{GP}$  could also be used for the active stereo system to predict more effectively the exposure time based on the feature's local light intensity and spatial relationships.

Analogously to the SLS, a training dataset  ${}^eX^{s_2}$  was considered with a threshold of  $p_{norm} > 75\%$ . The same incremental kernel design as proposed in Sect. 3.4 was followed. In the first step, a 1D kernel was considered to analyze the correlation magnitude between  $t_{exp}$  and  $i_{avg}$ . The results ( $R2 = 46.7\%$ ,  $RMSE = 3.6ms$ ) of the 1D kernel (shown in Table 5) confirm that the light intensity is also ineffective for setting an adequate exposure time. Hence, the model was extended first to a 4D kernel considering just positional inputs and then to a 6D kernel considering the sensor inclination. The 6D model denoted as  ${}^e\mathcal{GP}_6^{s_2}$  with the kernel  ${}^e k_6^{s_2}({}^eX_6^{s_2,train})$  (see Eq. 21) showed a considerable improvement on the test data as well as on the unseen observations with an  $R2 = 65.7\%$  and an  $RMSE$  of 2.9ms.

Moreover, a more detailed analysis showed that a performance equivalent to the test dataset could then be obtained by considering features ( ${}^eX_6^{s_2,test,i}$ ) within the training range (interpolation), i.e.,  $R2 = 97.4\%$  and  $RMSE = 0.79ms$ . These findings align well to our previous observations from Sect. 3.4.8 confirming that the the DoE for the training dataset should consider an heterogeneous spatial distribution of the features. Furthermore, Fig. 26 confirms the nonlinear trend of the problem and the existence of local maxima and minima at different positions with different light intensities.

### 4.2.4 Discussion

The results of this subsection confirmed that spatial relationships must also be considered when adjusting the camera exposure of active 3D stereo systems. Moreover, the model evaluations showed that the quality of the point cloud and the exposure time can be predicted analogously to the SLS following the proposed data-driven approach of the present study.

The GP models of the active stereo system showed a comparable performance to the SLS models demonstrating the high transferability and generalizability of the designed model inputs and GP kernels. However, it must be noted

that the kernels' validity is limited to the camera-projector configuration of the present study (cf. Sect. 2.1). For replication purposes, the suggested kernels can be used as a solid initial base. However, to optimize the accuracy of the models, a thorough and systematic kernel design should always be considered.

## 4.3 Different surface finishes

To evaluate the models' performance with different surface finishes, exemplary features from the car's exterior were used. Figure 18 provides an overview of the selected features.

### 4.3.1 Dataset

To properly evaluate the GP models of both sensors, additional datasets with random sensor positions were collected for the SLS  $X^{out}$  and the active stereo sensor  $X^{s_2,out}$ . The features are categorized into three groups depending on their surface finishes. The dataset characteristics for both sensors are given in Table 9. The evaluation of all models comprises exclusively observations lying within the spatial range of the training dataset.

### 4.3.2 Quality prediction model

The results of the evaluation of the quality prediction models from both sensors are summarized in Table 6. For both sensors, it must be mentioned that the models show a performance similar to that of the test dataset (see. Tables 2 and 5) for the features  $f_1^o-f_3^o$ . These results confirm the models' accuracy and transferability to similar surfaces.

However, both models show a higher uncertainty for the features ( $f_4^o, f_5^o$ ), limiting to some extent the models' usability for this surface finish. Moreover, it should be noted that most predictions exhibited an average overestimation of  $p_{norm}$ . This observation agrees with a common understanding that the model's accuracy could be improved by considering an additional input variable representing the surface reflectance properties.

### 4.3.3 Exposure time model

The accuracy of the exposure time models for both sensors showed a performance similar to that of the test features from the inside of the door (compare Table 3 and Table 5 with Table 6). While for the SLS model,  ${}^e\mathcal{GP}_6$ , the predictions of the first two feature groups yielded an acceptable  $RMSE$  of 20ms, the evaluation of the features  $f_{4-5}^o$  showed an outstanding performance with an  $RMSE$  below 10ms. On the contrary, the performance of the stereo sensor prediction model  ${}^e\mathcal{GP}_6^{s_2}$  revealed an even better prediction accuracy compared to the test features from the door's inside

with an average RMSE under 2ms. These results also seem plausible, given that the radii of the outside features were bigger than those from the inside.

On the one hand, our experiments showed that the SLS is more sensitive to the surface reflectance properties. For this reason, it can be expected that the model performance could be improved by integrating an input representing the surface reflectivity. On the other hand, the active stereo sensor seemed less sensitive to the surface reflectivity showing a better model's generalizability for similar surfaces. Given the results from Sect. 3.4.5, it is also reasonable to assume that a 6D model trained with the interior door surface would outperform a 1D model trained on each specific surface from the outside.

#### 4.3.4 Discussion

The evaluation results in this subsection demonstrate that the performance of the respective sensors'  ${}^q\mathcal{GP}$  and  ${}^e\mathcal{GP}$  apply to similar surface surfaces to some extent. While the surface reflectance properties seem to have little influence on the  ${}^q\mathcal{GP}$ s' performance, the material reflectivity showed a higher correlation to the exposure time affecting the  ${}^e\mathcal{GP}$ s' prediction accuracy more significantly. Hence, our results suggest that the models' performance should improve by considering an input variable characterizing the object's reflectance properties. Such investigations fall outside the scope of this paper. Therefore, future studies should consider a more exhaustive evaluation including different surface finishes with more observations.

## 5 Conclusion

A well-known challenge when using active optical 3D sensors (e.g., SLSs, active stereo systems) for scanning an object's surface is to adjust the image brightness using the camera's exposure time to produce a dense point cloud without holes. This task becomes even more complicated in dealing with highly reflective materials with a complex surface topology that produce nonlinear and inhomogeneous illumination for the camera image. Therefore, the optimal parameterization of the sensor for such use cases is still done manually, as it requires an advanced multisensory perception (spatial and visual) of the acquisition environment.

### 5.1 Summary

To tackle this challenge and automate finding an adequate exposure time while ensuring a successful acquisition, the present study introduced a data-driven approach using two GP models. In the first step, seven input and two output

variables were synthesized to model the imaging and spatial model of a generic measurement setup. For example, the local image intensity around a defined feature (region of interest) was used to model the projector illumination, and the rigid transformation between features and sensor were used to characterize the spatial correlations. In the next step, the present publication outlined a 4D GP to predict the measurement quality and a 6D GP for estimating the exposure time. Both models were trained and evaluated on an exhaustive dataset that was collected using an SLS and a car door as probing object.

On the one hand, the 4D quality regression model achieved a promising performance on unknown observations with an RMSE of 10% of the number of expected scanning points belonging to a feature's surface. On the other hand, the 6D exposure time regression model was able to predict the required exposure time for a local camera image brightness with an outstanding accuracy of an RMSE = 13ms (exposure time range 1–350 ms). The results from both models indicate considering spatial inputs is critical to making accurate predictions. Moreover, within a more comprehensive evaluation, the transferability of the models was assessed for different lighting conditions and surface finishes. Additionally, the overall usability of our framework (data pipeline, model inputs and outputs, and GP kernels) was evaluated satisfactorily considering a 3D active stereo sensor.

## 5.2 Outlook

Although the proposed models and designed kernels delivered promising results, it is worth noting that their validity and performance are limited to the measurement setup of this study. Therefore, we do not dispute that alternative data-based models could perform better with other systems. Furthermore, the validation of our models is limited to the finish of the car door sheet. Future work should consider a broader range of materials with different reflectance properties and lighting conditions to investigate the robustness and validity of the model and proposed inputs. Further fundamental research should explore transfer learning techniques to alleviate the effort required to collect data and retrain models. In addition, future work should comprehensively compare the proposed solutions with different approaches. This analysis is beyond this study's scope since most of the related strategies consider static environments requiring an individual model calibration for each operating point. However, in Sect. 3.4.9, the evaluation of a linear kernel (which could be to some extent compared to a linear model as suggested by the related research) demonstrated that such an approximation would be insufficient ( $R^2 = 10\%$ , see Table 3) to accurately predict a valid exposure time. Also, due to the configuration limitations

of the used sensors, a comparison with other approaches that propose hardware modifications would not be feasible.

Finally, a strategy that delivers an end-to-end solution, which combines both GP models to calculate the optimal exposure time for a nominal, optimal image brightness taking into account the relative location between a feature and the sensor, was beyond the scope of this study. In particular, designing a multi-objective optimization strategy that guarantees the acquisition of multiple features while minimizing the number of needed exposure times remains an open issue for future research. Following this idea, future studies should also explore integrating and combining the proposed approach with HDR techniques to optimize multiple exposure times. The design of such a strategy will also allow a thorough comparison of the proposed models with other approaches.

As far as we know, the present study outlines a first, effective, data-driven approach to predict the local point cloud quality of a feature and corresponding exposure time for 3D active optical sensors systems based on a local light intensity value and spatial relationships. We are convinced that this publication will provide a new stimulus for researchers and the industry to consider data-based approaches to an automated parameterization of 3D active sensors.

## Appendix 1. Kernel definition

Popular kernels which are also used in this study are the so-called SE kernel and the RQ kernel. They are defined as [28]:

$$k_{SE}(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{(\mathbf{x} - \mathbf{x}')^2}{2\ell^2}\right), \quad (22)$$

$$k_{RQ}(\mathbf{x}, \mathbf{x}') = \sigma^2 \left(1 + \frac{(\mathbf{x} - \mathbf{x}')^2}{2\alpha\ell^2}\right)^{-\alpha}, \quad (23)$$

where  $\ell$ ,  $\alpha$  and  $\sigma$  are hyperparameters to be learned during the training phase of the model.

In Eqs. 22 and 23, a single so-called *length scale* hyperparameter  $\ell$  is learned. If this length scale hyperparameter

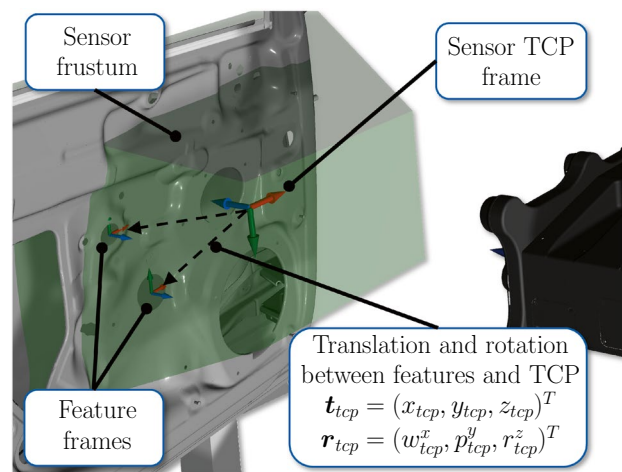
is to be learned separately for every dimension of  $\mathbf{x}$ , this can be achieved by multiplying kernels defined on each dimension  $x_d$ . Such kernels are called ARD kernels. For example, the ARD variant of the SE kernel is given as follows [28]:

$$\begin{aligned} k_{SE-ARD}(\mathbf{x}, \mathbf{x}') &= \prod_{d=1}^D \sigma_d^2 \exp\left(-\frac{1}{2} \frac{(x_d - x'_d)^2}{\ell_d^2}\right) \\ &= \sigma^2 \exp\left(-\frac{1}{2} \sum_{d=1}^D \frac{(x_d - x'_d)^2}{\ell_d^2}\right). \end{aligned} \quad (24)$$

Valid kernels can be recombined to form new valid ones using algebraic operations such as addition or multiplication. A list of valid kernel combinations can be found in [29]. In the next subsection, the use of GPs in regression tasks will be briefly introduced.

## Appendix 2. Tables and figures

See Figs. 19, 20, 21, 22, 23, 24, 25, 26 and Tables 7, 8, 9



**Fig. 19** Overview of the kinematic model for extracting the relative position and rotation between sensor and features



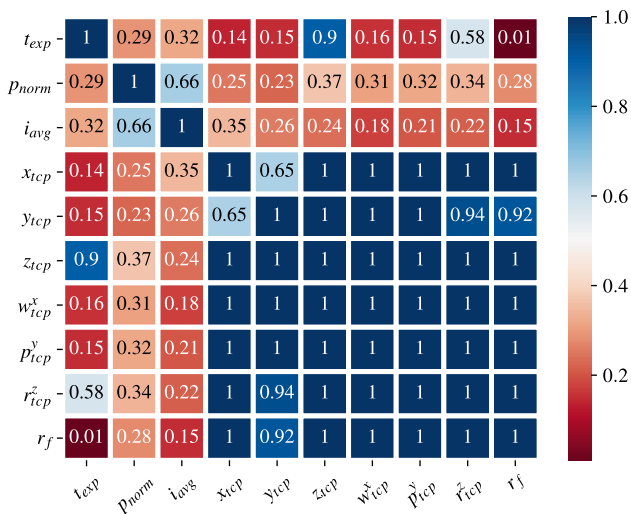


Fig. 20 Graphical representation of the data correlation using the MICs and the dataset  $X^{train}$

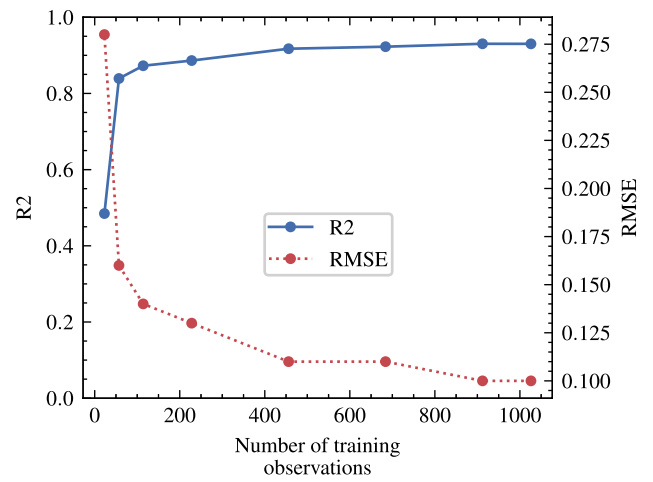


Fig. 22 Performance overview of the 4D Quality Prediction Model  ${}^q\mathcal{GP}_4$  for an increasing number of training observations based on the test dataset  $X_4^{test}$

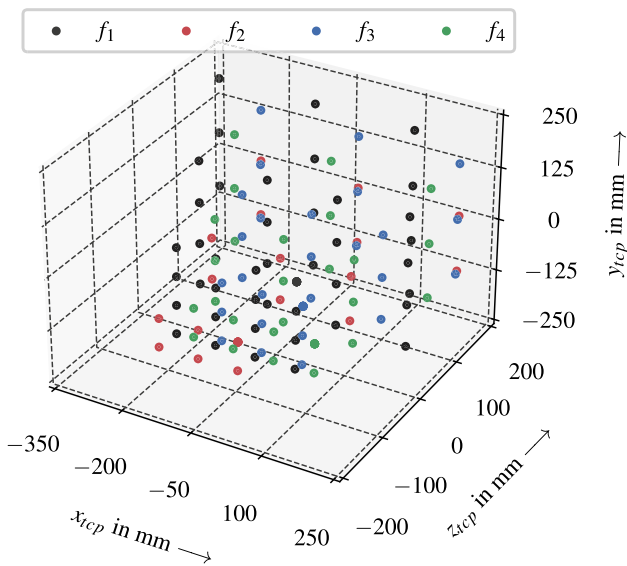


Fig. 21 Positional distribution of the four training features ( $f_1$ – $f_4$ ) of the dataset  $X$

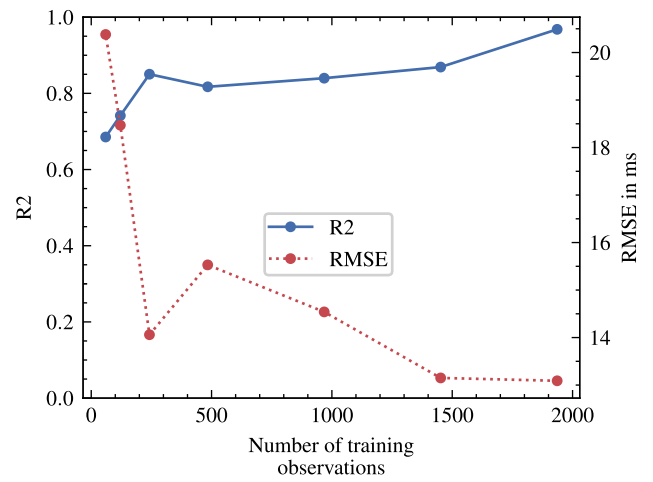
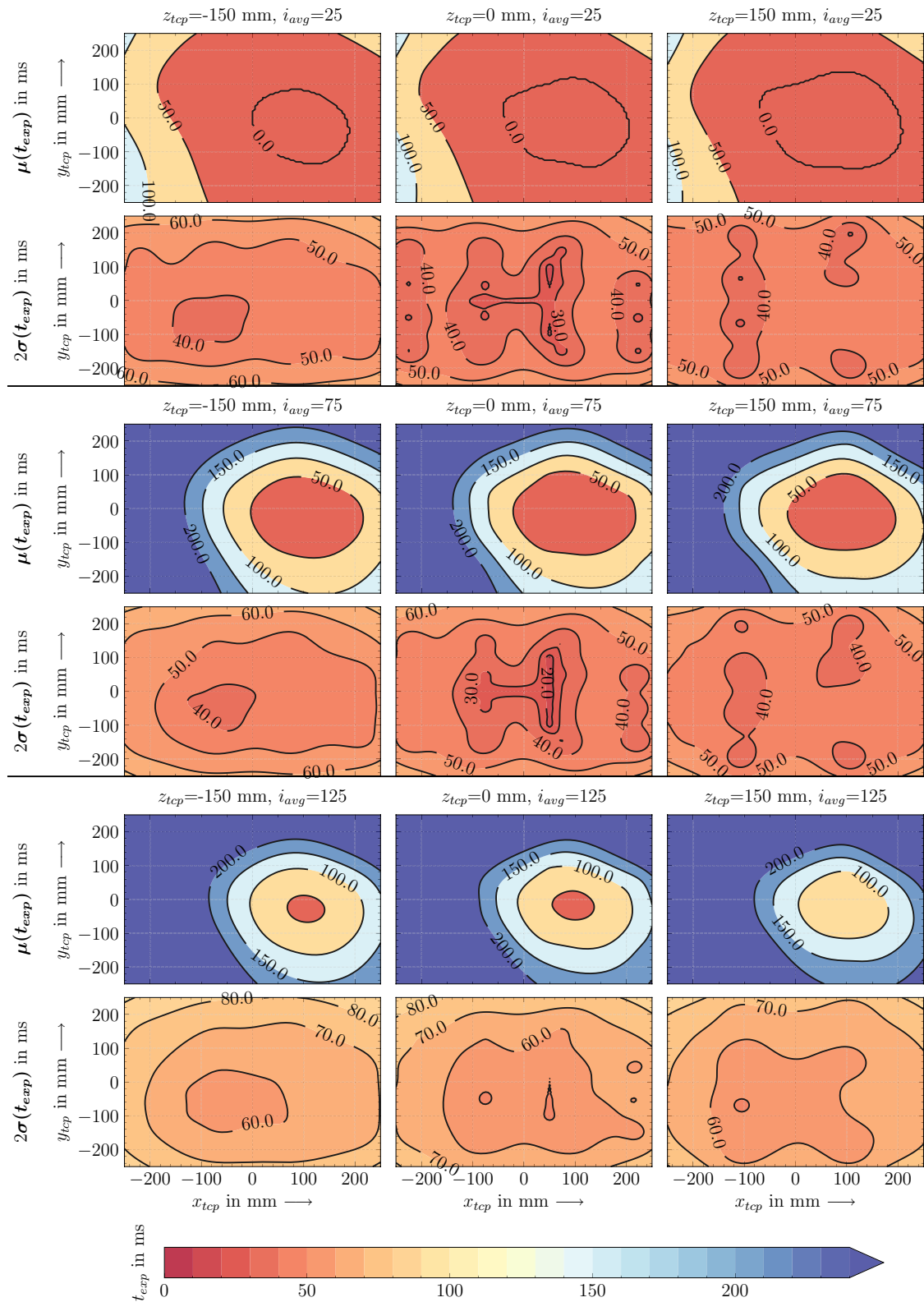
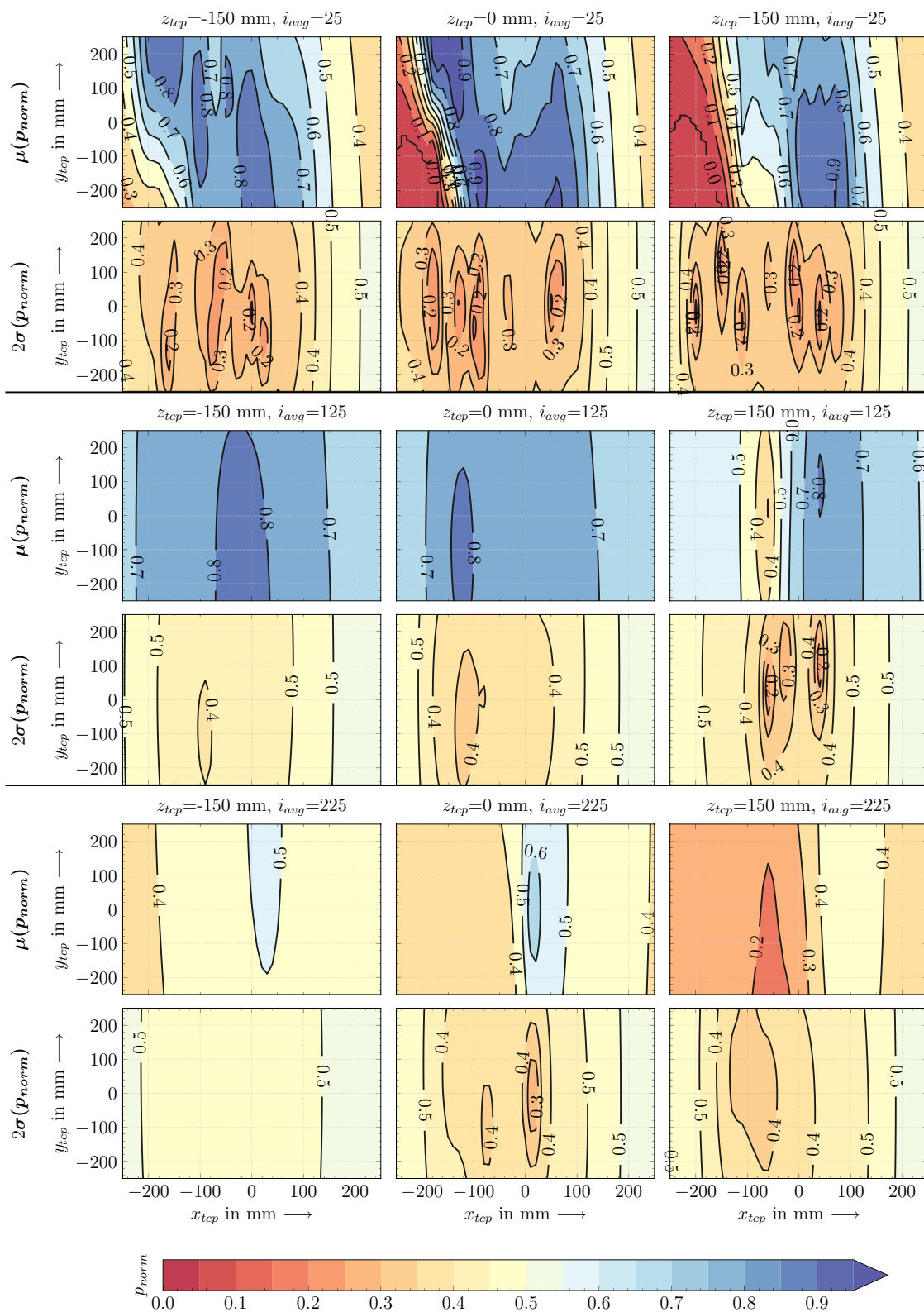


Fig. 23 Performance overview of the 6D Exposure Time Prediction Model  ${}^e\mathcal{GP}_6$  for different numbers of training observations using the test dataset  $X_6^{test,i}$



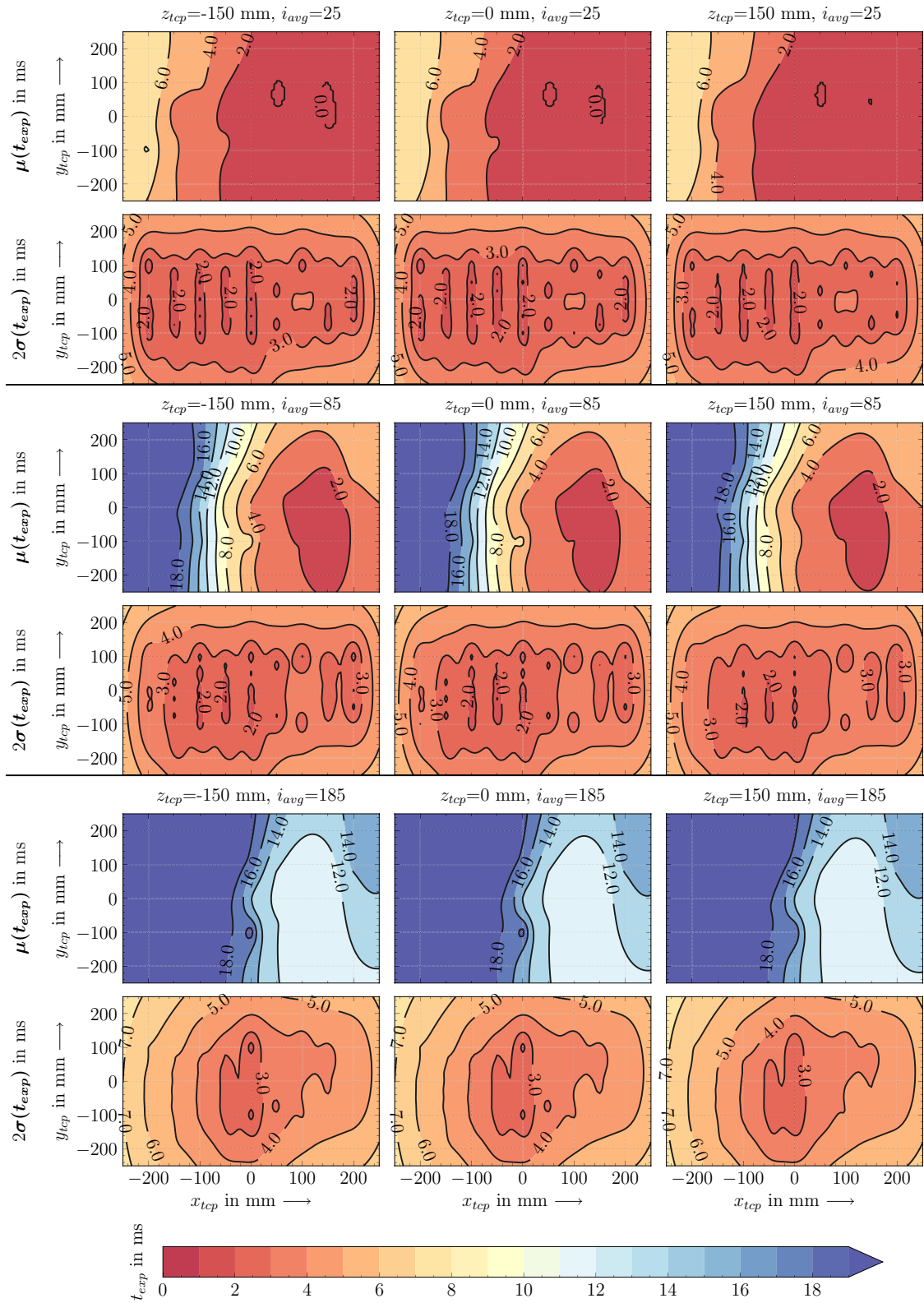
**Fig. 24** Discretized visualization of the mean prediction  $\mu(t_{exp})$  and corresponding double standard deviations  $2\sigma$  of the 6D GP exposure time regression model  $e\mathcal{GP}_6$  at three different depths  $z_{tcp} = [-150, 0, 150]$  mm

(left, middle, right) for three different light intensities  $i_{avg} = [25, 75, 125]$  (top, middle, bottom) with a sensor orientation of  $w_{tcp}^x = p_{tcp}^y = 0^\circ$  for the SLS Zeiss Comet Pro AE



**Fig. 25** Discretized visualization of the mean prediction  $\mu(p_{norm})$  and corresponding double standard deviations  $2\sigma$  of the 4D GP quality regression model  $rc\_visard\ 65$  at three different depths  $z_{tcp} = [-150, 0, 150]$

mm (left, middle, right) for three different light intensities  $i_{avg} = [25, 75, 125]$  (top, middle, bottom) for the active stereo sensor regression model  $rc\_visard\ 65$



**Fig. 26** Discretized visualization of the mean prediction  $\mu(t_{exp})$  and corresponding double standard deviations  $2\sigma$  of the 6D GP exposure time regression model  $eGP_6^{D2}$  at three different depths

$z_{tcp} = [-150, 0, 150]$  mm (left, middle, right) for three different light intensities  $i_{avg} = [25, 75, 125]$  (top, middle, bottom) with a sensor orientation of  $w_{tcp}^x = p_{tcp}^y = 0^\circ$  for the active stereo sensor *rc\_visard 65*

**Table 7** Overview of the imaging parameters of the used active optical sensors

	Sensors	
Model	COMET Pro AE	rc_visard 65 monochrome and RandomDot projector
Manufacturer	Carl Zeiss Optotechnik GmbH	Roboception GmbH
3D acquisition method	Digital fringe projection	Stereo vision and active projector
Devices	Monochrome camera and blue light LED fringe projector	Two monochrome cameras and dot projector
Measurement planes at working distances	@0.40 m: 0.40 m × 0.27 m @0.60 m: 0.59 m × 0.39 m @0.80 m: 0.78 m × 0.52 m	@0.10 m: 0.17 m × 0.18 m @0.50 m: 0.54 m × 0.45 m @1.00 m: 1.14 m × 0.90 m
Transformation between camera and projector	Position in m: (0.22, 0.0, 0.08) Rotation in °: (0.0, − 20.0, 0.0)	Position in m: (0.18, − 0.05, 0.21) Rotation in °: (0.0, − 20.0, 0.0)
Exposure time setting range	1–1000 ms	0.07–18 ms

**Table 8** Overview of the optimized kernel hyperparameters

Kernel	Hyperparameters: variance ( $\sigma^2$ ), lengthscale ( $l$ ), power( $\alpha$ ) and Gaussian noise ( $\epsilon$ )
${}^qk_1$	$\sigma^2 = 0.60, l = 17.00, \alpha = 0.39, \epsilon = 0.04$
${}^qk_2$	$\sigma^2 = 0.94, l = [0.94, 0.03], \alpha = 0.004, \text{ARD}=1, \epsilon = 0.00$
${}^qk_3$	$\sigma^2 = 1.41, l = [1.33, 0.71, 0.08], \alpha = 0.04, \text{ARD}=1, \epsilon = 0.00$
${}^qk_4$	$\sigma^2 = 1.45, l = [1.06, 0.52e2, 0.18e2, 0.33e2], \alpha = 0.010, \text{ARD}=1, \epsilon = 0.00$
${}^qk_6$	$\sigma^2 = 0.98, l = [1.50, 1.63e2, 0.37e2, 0.51e2, 7.2, 0.1], \alpha = 0.01, \text{ARD}=1, \epsilon = 0.037$
${}^ek_1$	$\sigma^2 = 1.63e4, l = 1.26e2, \epsilon = 4.67e3$
${}^ek_2$	$\sigma^2 = 1.71e4, l = 1.07e3, \epsilon = 3.48e3$
${}^ek_3$	$\sigma^2 = 1.25e5, l = 3.41e3, \epsilon = 9.92e2$
${}^ek_4$	$\sigma^2 = 3.74e5, l = 5.61e3, \epsilon = 0.00$
${}^ek_4^+$	$\sigma^2 = 6.19e5, l = 2.13e4, \epsilon = 6.07$
${}^ek_6^+$	$\sigma^2 = 6.94e5, l = 2.49e4, \epsilon = 6.05$
${}^ek_6^{++}$	$\sigma^2 = 7.34e5, l = 4.63e4, \epsilon = 0.58$
${}^ek_6^*$	$\sigma^2 = 1.14e6, l = [1.20e5, 5.45e5, 4.63e5, 5.51e5, 2.59e4, 1.86e4], \text{ARD}=1, \epsilon = 0.58$
${}^ek_6$	$k_{\text{exp}}(i_{\text{avg}}): \{ \sigma^2 = 0.66e2, l = 2.20e2 \}, k_{\text{exp}}(x_{\text{tcp}}, y_{\text{tcp}}, z_{\text{tcp}}, w_{\text{tcp}}^x, p_{\text{tcp}}^y): \{ \sigma^2 = 3.31e2, l = [9.97e3, 2.97e3, 3.77e3, 2.47e2, 1.73e2], \text{ARD}=1 \}, \epsilon = 0.04$
${}^ek_6^{\text{lin}}$	$\sigma^2 = [8.69e-1, 6.7e-2, 2.6e-2, 2.4e-2, 5.08, 5.73], \epsilon = 1543.51$
${}^qk_1^{\text{s}2}$	$\sigma^2 = 1.16, l = 1.94, \alpha = 0.02, \epsilon = 0.04$
${}^qk_4^{\text{s}2}$	$\sigma^2 = 1.17, l = [0.69, 8.42, 3.73, 54.4], \alpha = 0.01, \text{ARD}=1, \epsilon = 0.00$
${}^qk_6^{\text{s}2}$	$\sigma^2 = 1.13, l = [1.08, 87.01, 30.05, 36.4, 2.00, 0.05], \alpha = 0.01, \text{ARD}=1, \epsilon = 0.00$
${}^ek_1^{\text{s}2}$	$\sigma^2 = 0.88e2, l = 7.50e2, \epsilon = 11.2$
${}^ek_4^{\text{s}2}$	$k_{\text{exp}}(i_{\text{avg}}): \{ \sigma^2 = 0.17e2, l = 2.93e2 \} k_{\text{exp}}(x_{\text{tcp}}, y_{\text{tcp}}, z_{\text{tcp}}): \{ \sigma^2 = 2.22, l = [3.11e4, 2.79e2, 2.20e2], \text{ARD}=1 \}, \epsilon = 0.00$
${}^ek_6^{\text{s}2}$	$k_{\text{exp}}(i_{\text{avg}}): \{ \sigma^2 = 5.10, l = 2.62e2 \} k_{\text{exp}}(x_{\text{tcp}}, y_{\text{tcp}}, z_{\text{tcp}}, w_{\text{tcp}}^x, p_{\text{tcp}}^y): \{ \sigma^2 = 5.1, l = [1.35e4, 3.47e2, 1.09e3, 2.61e3, 9.22], \text{ARD}=1 \}, \epsilon = 0.00$

**Table 9** Overview of the training range of all collected datasets

Dataset	Features	Nr. of observations	Dataset Range		$t_s$ in ms	$i_{avg}$	$x_{tcp}$ in mm	$y_{tcp}$ in mm	$z_{tcp}$ in mm	$w_{tcp}^x$ in °	$p_{tcp}^y$ in °	$r_{tcp}^z$ in °
			$p_{norm}$									
X		390	0.0 to 0.9	{5, 7, 10, 14, 19, 26, 37, 51, 72, 100}	15 to 247	- 135 to 182	- 101 to 105	- 188 to - 149	1 to 11	- 3 to 0	0	
	$f_{1-4}^I$	400	0.0 to 1.0	{5, 7, 11, 16, 23, 33, 48, 70, 103, 150}	13 to 244	- 226 to 224	- 151 to 152	- 28 to 12	1 to 11	- 2 to 1	0	
		350	0.0 to 1.0	{20, 27, 38, 52, 71, 98, 135, 185, 225, 350}	14 to 255	- 325 to 233	- 200 to 200	132 to 169	1 to 11	- 2 to 1	0	
$eX$	$f_{1-4}^{II}$	880	0.0 to 1.0	5 to 350	15 to 254	- 304 to 269	- 183 to 218	- 177 to 167	- 1 to 5	- 13 to - 3	0	
	$f_{1-4}^{III}$	527	0.75 to 1.0	5 to 350	17 to 125	- 325 to 233	- 200 to 200	- 179 to 169	1 to 11	- 3 to 1	0	
	$f_{1-4}^{IV}$	508	0.75 to 1.0	5 to 350	17 to 127	- 304 to 269	- 183 to 218	- 177 to 167	- 1 to 5	- 13 to - 3	0	
$X^+$	$f_{1-9}^V$	948	0.75 to 1.0	5 to 350	15 to 133	- 325 to 304	- 200 to 223	- 189 to 170	1 to 11	- 6 to 1	0	
	$f_{1-9}^{VI}$	2421	0.75 to 1.0	0 to 350, 5 to 120	15 to 133	- 325 to 304	- 200 to 223	- 189 to 170	1 to 11	- 6 to 1	0	
$X^{I,I}$				(in 5 ms steps) {15, 40, 65, 90, 115}	20 to 210							
$X^{I,II}$	$f_{1-3}^{VII}$	15	0.0 to 1.0		21 to 210	- 189 to 86	18 to 152	- 14 to 5	- 1 to 3	- 10 to - 4	0	
					21 to 211							
$X^{II}$	$f_{1-4}^{VIII}$	13466	0.0 to 1.0	0.07 to 18	10 to 255	- 293 to 309	- 234 to 100	- 172 to 160	0 to 10	- 2 to 1	0	
	$f_{1-4}^{IX}$	6966	0.0 to 1.0	0.07 to 18	10 to 255	- 393 to 321	- 197 to 250	- 161 to 156	- 5 to 2	- 9 to 8	0	
	$f_{1-4}^{X}$	3173	0.75 to 1.0	0.09 to 18	12 to 241	- 200 to 209	- 200 to 100	- 166 to 142	0 to 10	- 2 to 1	0	
$eX^{II}$	$f_{1-4}^{XI}$	4201	0.75 to 1.0	0.12 to 18	13 to 255	- 343 to 229	- 140 to 250	- 161 to 156	- 5 to 2	- 9 to 8	0	
	$f_{1-5}^{XII}$	240	0.0 to 1.0	{15, 40, 65, 90, 115}	14 to 229	- 300 to 258	- 190 to 133	- 101 to 134	- 3 to 4	- 2 to 0	0	
$X^{S_2,out}$	$f_{1-5}^{XIII}$	498	0.0 to 0.9	0.07 to 18	10 to 255	- 287 to 26	- 146 to 81	33 to 146	- 2 to 6	- 2 to 1	0	

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11740-023-01240-4>.

**Acknowledgements** The framework presented in this paper was developed and thoroughly evaluated within the scope of the “CyMePro” (Cyber-physical measurement technology for 3D digitization in the networked production) project. We thank our research partners AUDI AG and Carl Zeiss Optotechnik GmbH for the fruitful discussions and their cooperation.

**Author Contributions** All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by AM and LS. The first draft of the manuscript was written by AM and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** Open Access funding enabled and organized by Projekt DEAL. This research was funded by the Bavarian Ministry of Economic Affairs, Regional Development and Energy (funding code ESB036/001).

**Data availability** A sample training set for two features can be found in the attachment of this paper. The set includes a CSV file, 2D camera measurements, and 2D and 3D masks utilized for calculating input and output variables.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Leach RK (2020) Advances in optical form and coordinate metrology ISSN. IOP, Bristol
2. Chen F, Brown Gordon M, Song M (2000) Overview of three-dimensional shape measurement using optical methods. *Opt Eng* 39(1):10. <https://doi.org/10.1117/1.602438>
3. Ekstrand L, Zhang S (2011) Autoexposure for three-dimensional shape measurement using a digital-light-processing projector. *Opt Eng* 50(12):123603. <https://doi.org/10.1117/1.3662387>
4. Nayar SK, Ikeuchi K, Kanade T (1991) Surface reflection: physical and geometrical perspectives. *IEEE Trans Pattern Anal Mach Intell* 13(7):611–634. <https://doi.org/10.1109/34.85654>
5. Palousek D et al (2015) Effect of matte coating on 3D optical measurement accuracy. *Opt Mater* 40:1–9. <https://doi.org/10.1016/j.optmat.2014.11.020>
6. Shi Q, Xi N, Sheng W, Chen Y (2006) Development of dynamic inspection methods for dimensional measurement of automotive body parts. *IEEE Int Conf Robot Autom*. <https://doi.org/10.1109/ROBOT.2006.1641730>
7. Zhang S, Yau S-T, Huang PS, Yoshizawa T, Harding KG (eds) (2008) High dynamic range scanning technique. In: Huang PS, Yoshizawa T, Harding KG (eds) Two- and three-dimensional methods for inspection and metrology VI, SPIE proceedings, pp A1–A9 (SPIE, 2008)
8. Lin H et al (2017) Three-dimensional shape measurement technique for shiny surfaces by adaptive pixel-wise projection intensity adjustment. *Opt Lasers Eng* 91:206–215. <https://doi.org/10.1016/j.optlaseng.2016.11.015>
9. Feng S et al (2014) General solution for high dynamic range three-dimensional shape measurement using the fringe projection technique. *Opt Lasers Eng* 59:56–71. <https://doi.org/10.1016/j.optlaseng.2014.03.003>
10. Feng S, Chen Q, Zuo C, Asundi A (2017) Fast three-dimensional measurements for dynamic scenes with shiny surfaces. *Opt Commun* 382:18–27. <https://doi.org/10.1016/j.optcom.2016.07.057>
11. Gribben J, Boate AR, Douglass MR, Lee BL, Ehmke J (eds) (2019) Structured light 3D measurement of reflective objects using multiple DMD projectors. In: Douglass MR, Lee BL, Ehmke J (eds) Emerging digital micromirror device based systems and applications XI, 20 (SPIE, 02.02.2019–07.02.2019)
12. Jiang C, Bell T, Zhang S (2016) High dynamic range real-time 3D shape measurement. *Opt Express* 24(7):7337–7346. <https://doi.org/10.1364/OE.24.007337>
13. Lin H et al (2017) Review and comparison of high-dynamic range three-dimensional shape measurement techniques. *J Sens* 2017:1–11. <https://doi.org/10.1155/2017/9576850>
14. Feng S et al (2018) High dynamic range 3D measurements with fringe projection profilometry: a review. *Meas Sci Technol* 29(12):122001. <https://doi.org/10.1088/1361-6501/aae4fb>
15. Zhang S (2020) Rapid and automatic optimal exposure control for digital fringe projection technique. *Opt Lasers Eng* 128:106029. <https://doi.org/10.1016/j.optlaseng.2020.106029>
16. Koutecký T, Paloušek D, Brandejs J (2016) Sensor planning system for fringe projection scanning of sheet metal parts. *Measurement* 94:60–70. <https://doi.org/10.1016/j.measurement.2016.07.067>
17. Catalucci S et al (2020) Measurement of complex freeform additively manufactured parts by structured light and photogrammetry. *Measurement* 164:108081. <https://doi.org/10.1016/j.measurement.2020.108081>
18. Rasmussen CE, Williams CKI (2006) Gaussian processes for machine learning Adaptive computation and machine learning. MIT, Cambridge
19. Giancola S, Valenti M, Sala R (2018) A survey on 3D cameras: metrological comparison of time-of-flight, structured-light and active stereoscopy technologies. Springer, Cham
20. Zheng Y, Wang Y, Suresh V, Li B (2019) Real-time high-dynamic-range fringe acquisition for 3D shape measurement with a RGB camera. *Meas Sci Technol* 30(7):075202. <https://doi.org/10.1088/1361-6501/ab0ced>
21. Dawson-Haggerty et al. trimesh (2022) <https://trimsh.org/>
22. Bradski G (2000) The OpenCV library. *Dr. Dobb’s Journal of Software Tools*
23. Santner TJ, Williams BJ, Notz WI (2003) Space-filling designs for computer experiments. In: Santner TJ, Williams BJ, Notz WI (eds) The design and analysis of computer experiments springer series in statistics. Springer, New York, pp 121–161

24. Magaña A, Dirr J, Bauer P, Reinhart G (2023) Viewpoint generation using feature-based constrained spaces for robot vision systems. *Robotics* 12(4):108. <https://doi.org/10.3390/robotics12040108>
25. Magaña A et al (2023) Viewpoint planning for range sensors using feature cluster constrained spaces for robot vision systems. *Sensors* (Basel, Switzerland). <https://doi.org/10.3390/s23187964>
26. Scott WR (2009) Model-based view planning. *Mach Vis Appl* 20(1):47–69. <https://doi.org/10.1007/s00138-007-0110-2>
27. GPY (2012) GPY: a Gaussian process framework in python (since 2012). <http://github.com/SheffieldML/GPY>
28. Duvenaud DK (2014) Automatic model construction with Gaussian processes. Dissertation, University of Cambridge
29. Bishop CM (2006) *Pattern recognition and machine learning* Corrected at 8th printing. Information science and statistics, 2009th edn. Springer, New York

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.