Method for Parametric Evaluation of 3-D Surface Imaging Systems for Applications With Moving Objects

Mirko Kaiser, Tobia Brusa, Stefan Fiechter, Martin Bertsch, Marco Wyss, Saša Ćuković, *Senior Member, IEEE*, William R. Taylor, and Volker M. Koch, *Senior Member, IEEE*

***Abstract*— Medical applications in which patient movements are tracked with 3-D surface imaging systems are becoming increasingly popular. The 3-D imaging systems used for such applications must be able to cope with controlled and uncontrolled motions of the human body. The key factors for producing a high-quality 3-D representation of the moving human body are the spatial resolution, accuracy, and precision of the 3-D imaging system. To our knowledge, no international standard yet exists to assess these parameters. In this article, we propose a phantom model, a method, and parameters for the assessment of spatial resolution, accuracy, and precision to evaluate systems for the 3-D imaging of moving objects. The proposed phantom model is an extension of the standard 1951 USAF resolution test chart to 3-D and consists of two parallel staircases with varying step heights. The phantom model is actuated with an industrial robot. The application of our method to the Photoneo MotionCam-3-D showed that their multishot structured-light mode has a higher accuracy for static and slowly moving objects (accuracy of 0.1 mm) than their single-shot structured-light mode (accuracy of 0.5 mm). However, the single-shot mode can capture fast-moving objects without much loss of accuracy and precision. This provides practical quantifications for the Photoneo MotionCam-3-D. Furthermore, this confirms that the proposed phantom model, method, and parameters can be used as part of a standard to assess the spatial resolution, accuracy, and precision of systems for the 3-D imaging of moving objects.**

***Index Terms*— 3-D measurements, accuracy, moving objects, precision, quality assessment, spatial resolution.**

I. Introduction

THREE-DIMENSIONAL surface imaging is increasingly used in applications in the medical field [1], [2], [3], [4], [5], [6], [7], [8], [9], with approaches that track patient movements becoming ubiquitous. Some applications include captures of the human upper body, especially the human back in various postures and movements. The 3-D imaging systems used for such applications must be able to capture controlled and allow for uncontrolled movements. Controlled human movements include bending forward and bending sideways. Uncontrolled movements include breathing or small involuntary movements that occur even when capturing static postures. To cope with such movements, the spatial resolution, accuracy, and precision of the 3-D imaging system must fulfill certain conditions to capture a reproducible 3-D representation of the moving human body at sufficient quality for the intended use case. Many authors address this issue, including applications of moving objects outside the scope of medical applications [10], [11], [12], [13], [14], [15]. However, to our knowledge, no international standard exists to assess the spatial resolution, accuracy, and precision of 3-D imaging systems. Thus, people tend to evaluate multiple approaches and state various quality parameters to increase confidence in their results. Furthermore, a variety of phantom models are used to quantify the spatial resolution, accuracy, and precision of imaging systems. The phantom models used in various papers range from planar surfaces and spherical objects to complex objects such as step surfaces, 3-D printed letters, rotating propellers, statues, and human body parts [10], [11], [12], [13], [14], [15]. In the case of a planar surface, the precision error is found by comparing to a best-fitting plane [10], [11]. For spherical objects, the accuracy error is found by comparing to the diameter of a best-fitting sphere [11], [16]. For complex objects, a general reconstruction error incorporating both precision and accuracy can be determined by comparing to a best-fitting ground-truth model [11], [15]. The variety of phantom models makes it impossible to compare results between these systems.

In 2-D, the 1951 USAF resolution test chart is an accepted standard and is used in many applications [17], [18], [19], [20], [21]. In this article, a 3-D phantom model is proposed that extends the 1951 USAF resolution test chart to 3-D. Furthermore, a method and parameters for spatial resolution, accuracy, and precision are proposed to evaluate 3-D imaging systems for the application of moving objects. The proposed method is generic for various applications, including measurements of human motion.

Manuscript received 2 October 2023; revised 23 January 2024; accepted 21 February 2024. Date of publication 4 March 2024; date of current version 18 March 2024. This work was supported in part by the Innosuisse under Grant 47195.1 IP-LS. The Associate Editor coordinating the review process was Dr. Chao Tan. *(Mirko Kaiser, Tobia Brusa, and Stefan Fiechter are co-first authors.) (Corresponding author: Mirko Kaiser.)*

Mirko Kaiser and Martin Bertsch are with the Biomedical Engineering Laboratory (BME Lab), Bern University of Applied Sciences, 2502 Biel, Switzerland, and also with the Laboratory for Movement Biomechanics (LMB), ETH Zurich, 8092 Zurich, Switzerland (e-mail: mirko.kaiser@bfh.ch; martin.bertsch@bfh.ch).

Tobia Brusa, Stefan Fiechter, Marco Wyss, and Volker M. Koch are with the BME Laboratory, 2502 Biel, Switzerland (e-mail: tobia.brusa@bfh.ch; marco.wyss@bfh.ch; volker.koch@bfh.ch).

Saša Ćuković and William R. Taylor are with the Laboratory for Movement Biomechanics LMB, 8092 Zurich, Switzerland (e-mail: cukovics@ethz.ch; bt@ethz.ch).

Digital Object Identifier 10.1109/TIM.2024.3373077

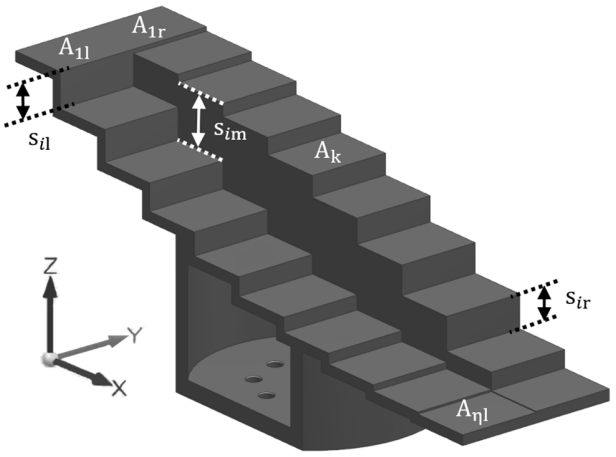

Fig. 1. CAD drawing of the proposed phantom model with coordinate system, tread surfaces (A_k, e.g., A_{1l}, A_{1r}, $A_{\eta l}$), and varying step heights for left (s_{il}), middle (s_{im}), last middle ($s_{\eta m}$), and right (s_{ir}) steps, where $i \in [1, \eta]$, and η is the number of steps.

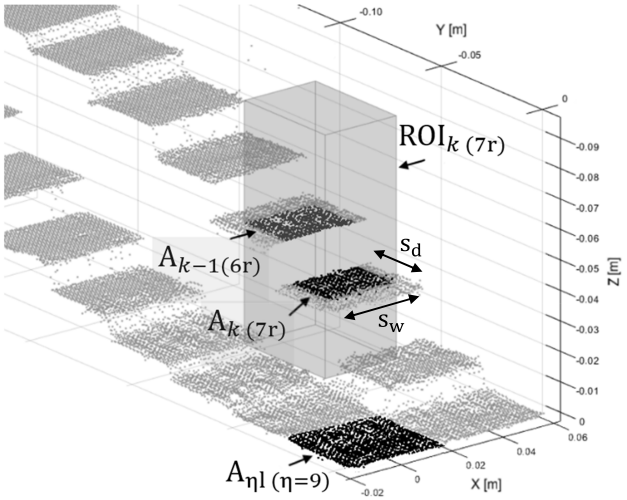

Fig. 2. Captured staircase model with visualization for ROI_{7r} covering the stair step between the stair tread surfaces A_{7r} and A_{6r}, where the tread surface A_{9l} is orthogonally aligned with the image axis. The ROI is defined by the height of the entire staircase model, the stair tread depth (s_d), and width (s_w).

II. Materials and Methods

Section A of this section describes the proposed phantom model. Section B explains the accuracy and precision analyses. Section C explains the steps that are necessary to set up the system and apply the proposed method correctly, and Section D outlines the data acquisition and processing steps in detail.

A. Phantom Model

The proposed phantom model (Fig. 1) allows an evaluation of spatial resolution, accuracy, and precision in a single and uniform manner. The design combines flat surfaces (A_k) and unique geometric structures such as those in the 1951 USAF resolution test chart. The structure consists of two parallel staircases with different step heights (s_{ir}, s_{il}, s_{im}) that replicate the step series in the spatial domain. This design allows steps in both directions, the left and right steps in the X-direction and the middle steps in the Y-direction. The step heights on the right (s_{ir}), left (s_{il}), and middle (s_{im}) steps, as well as the last middle ($s_{\eta m}$) step are determined by the following equation:

$$s_{ir} = i \cdot \zeta \tag{1}$$

$$s_{il} = (\eta - i + \frac{1}{2}) \cdot \zeta \tag{2}$$

$$s_{im} = s_{ir} - s_{il}, \quad s_{\eta m} = \frac{\zeta}{4} \tag{3}$$

where $i \in [1, \eta]$ is the counter variable, ζ the minimum desired accuracy, and η the number of steps. The stair treads (A_{1l}, A_{1r}) at the upper end are aligned to provide variable but defined step heights for the middle steps.

The model can be scaled to meet the requirements of the end application, including resolution capabilities, object size, and so forth. This includes the scaling of the overall model, the number of steps η, and the scaling of the smallest step height. To define a custom scaling, a minimum desired accuracy ζ has to be defined for the specific application. The smallest step height is chosen as a quarter of the size of the selected value for ζ to allow for sufficient margin.

Spatial resolution and accuracy are evaluated with the different step heights (s_{ir}, s_{il}, s_{im}), whereas precision is determined with the flat tread surfaces (A_k). All parameters are determined from the depth measurements in the Z-direction.

B. Accuracy and Precision Analysis

Absolute accuracy in this article states how close the measured distance between the stair tread and the camera is to the real distance. The relative accuracy states how close the measured step height is to the real step height. The focus is on the relative accuracy as this metric is more relevant to reconstructing the shape of a target. Precision states how dispersed the points of the measured stair tread surface are.

To analyze accuracy and precision, the captured staircase model is divided into $3{\cdot}\eta$ regions of interest (ROI). Each stair tread A_k is covered by three regions of interest: $\text{ROI}_{k,}$, $\text{ROI}_{k,m}$, and ROI_{k+1}. Each ROI (Fig. 2) is centered on a stair step such that only a single stair step is contained. The placement of each ROI is calculated from a previously placed bounding box around the entire staircase model (Section II-D). The width, length, and height of each ROI are $\text{ROI}_w = 0.5 \cdot s_w$, $\text{ROI}_l = 0.5 \cdot s_d$, and $\text{ROI}_h = 16 \cdot \text{bb}_h$, where s_w is the tread width, s_d the tread depth, and bb_h the bounding box height (height of the entire staircase model). The width and depth are smaller than the tread width and height to increase the robustness of our algorithm against small placement errors of the bounding box. The height of the ROI (ROI_h) is 16× larger than the height of the entire staircase model to include outliers in the Z-direction.

All points inside ROI_k (Fig. 2, A_{k-1} and A_k) are analyzed separately. For each ROI, a histogram (Fig. 3) is created over the distribution in the Z-direction of all contained points. Each histogram contains two distinct peaks (Fig. 3, Z_{p1} and Z_{p2}) corresponding to both tread surfaces with a distribution depending on the precision of the camera system. The two peaks are detected and separated by a convolution with a

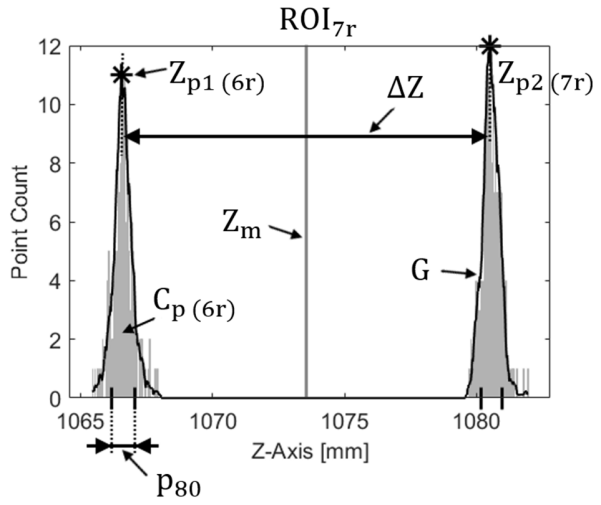

Fig. 3. Histogram of points contained in ROI_{7r} with point count of closer peak (C_{6r}), resulting graph of Gaussian convolution (G), detected peaks (Z_{6r}, Z_{7r}), midline of detected peaks (Z_m), distance between peaks (ΔZ), and 80th percentile of peak distribution (p_{80}).

normalized Gaussian kernel. The standard deviation σ of the Gaussian distribution is determined by the minimal desired accuracy $\sigma = \zeta/4$. The two highest peaks from the output of the convolution with the Gaussian kernel represent the mean position of the two stair treads.

In the following, we assume that we have captured from our staircase model in motion, called dynamic data (Section II-D), and captures from our staircase model at rest, called ground-truth reference (Section II-D).

The **relative accuracy** for each ROI is determined by comparing the distance (Fig. 3, ΔZ) between the two highest peaks from the captures of the staircase model in motion (dynamic data) to the distance between the two highest peaks from the captures of the staircase model at rest (ground-truth reference) for the same ROI

$$\text{acc}_{\text{rel}} = \text{abs}\left(Z_{p1,d} - Z_{p2,d}\right) - \text{abs}\left(Z_{p1,r} - Z_{p2,r}\right) \quad (4)$$

where acc_{rel} is the relative accuracy, $Z_{p1,d}$ the distance value of the first peak from the dynamic data, $Z_{p2,d}$ the distance value of the second peak from the dynamic data, $Z_{p1,r}$ the distance value of the first peak of the ground-truth reference, and $Z_{p2,r}$ the distance value of the second peak of the ground-truth reference.

The overall relative accuracy of the camera is defined by the median and interquartile range (IQR) of all relative accuracy outcomes of all regions of interest in which the two distributions of the tread surfaces can be separated. Consequently, the **minimal spatial resolution** of the camera is determined by the smallest step height at which the two distributions of the tread surfaces still can be separated.

The **precision** for each ROI is determined by separating the two peak distributions at their midline (Z_m). The 80th percentile is determined for both tread surface distributions. The 80th percentile is used to make the precision calculation more robust against outliers (Fig. 3). Especially for high-velocity motion, some outliers are always present. The precision is then determined by comparing the 80th percentile from the dynamic data to the 80th percentile from the ground-truth reference

$$\text{precision} = p_{80,d} - p_{80,r} \quad (5)$$

where $p_{80,d}$ is the 80th percentile from the dynamic data and $p_{80,r}$ the 80th percentile from the ground-truth reference.

The normalization with the 80th percentile from the ground-truth reference is used to cancel out any bias, e.g., due to a misalignment between the staircase model and the image plane of the camera. Even a small alignment error between the camera plane and staircase model leads to a nonzero distribution in the Z-direction because the stair treads are not perpendicular to the Z-axis.

The overall precision of the camera is defined by the median and IQR of all precision outcomes of all regions of interest in which the two distributions of the tread surfaces can be separated.

C. System Setup

The following steps are necessary to set up the system and apply our method correctly.

1) The first step is to define the minimum desired accuracy ζ in millimeters and the number of steps η. The staircase model is then scaled according to the number of steps such that the smallest step height is $\zeta/4$.
2) Next, the distance between the measurement system and the test object is defined to suit the specific application.
3) The direction of motion or path of motion is defined. We recommend starting with an isolated forward and backward motion, followed by an isolated motion parallel to the image plane, e.g., left and right, and an application-specific path of motion. We recommend using an industrial robot or precise linear actuator to follow the path with high repeatability. To compare and repeat the measurements, the repeatability error of the path should be magnitudes, e.g., 10×, smaller than the minimum desired accuracy ζ.
4) The velocities of interest are defined. We recommend starting with static and then with low velocities such that results can be compared between applications. The gradual increase of the velocity and upper limit is chosen depending on the specific application. The staircase model is moved in the direction of motion for each of these velocities, resulting in an evaluation at each velocity.
5) The settings for the measurement system are defined, including projector brightness, exposure time, calibration, and if applicable, other application-specific settings or environmental conditions. These values should not change for the entire measurement phase and should be recorded and stated with the results.
6) The image plane must be aligned with the tread surface of the staircase model (Fig. 4) and the image axis with the step tread $A_{\eta l}$ (Fig. 2). The stair steps are asymmetric and decrease from right to left and top to bottom. The alignment with the step tread $A_{\eta l}$ ensures that all steps are fully visible and not occluded by a higher tread surface.

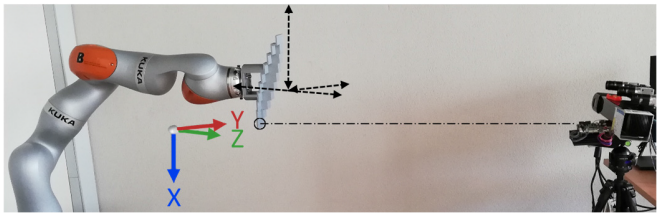

Fig. 4. Test setup with Photoneo MotionCam-3-D M+ and KUKA LBR iiwa 7 R800 with the image axis (dashed line) aligned with tread surface $A_{\eta l}$ of the proposed staircase phantom model.

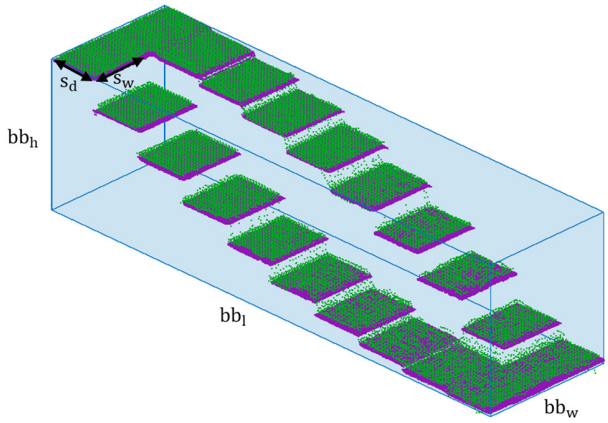

Fig. 5. Dynamic measurement data (green), static reference (violet) and bounding box (blue) with height (bb_h), length (bb_l) and width (bb_w), and stair tread depth (s_d) and width (s_w).

D. Data Acquisition and Processing

To analyze the captured data in a consistent and reproducible manner, all measurements must be made under the same conditions and the output saved in the same format.

Our method requires two types of recordings: a static ground-truth reference for the staircase model (Fig. 1) and dynamic measurement data (Fig. 5). The dynamic measurement data are compared to the ground-truth reference model. Using the ground-truth model as a reference eliminates any bias and delivers consistent and comparable results. The reference model can be either the exported 3-D CAD staircase model or captured with a high-quality 3-D scanner in static condition. If the reference model is captured with a high-quality 3-D scanner, the quality of the ground-truth reference must be verified according to the application requirements. This can be done by first applying the proposed method (Section II-B) to three consecutive static captures of the ground-truth reference. The resulting accuracy and precision must be magnitudes, i.e., 10×, higher than the minimum desired accuracy ζ. Using a static capture as ground-truth reference has the advantage that no initial alignment of position and orientation is needed with the dynamic recordings. If the 3-D CAD model is used as ground-truth reference for the dynamic data, this alignment need only be done once at the beginning. With the 3-D CAD model as reference, the quality verification step can be omitted.

The dynamic data must contain at least three recordings of the staircase model for any specific application parameter and each defined velocity (Section II-C). Three recordings are used to increase the robustness of the evaluation.

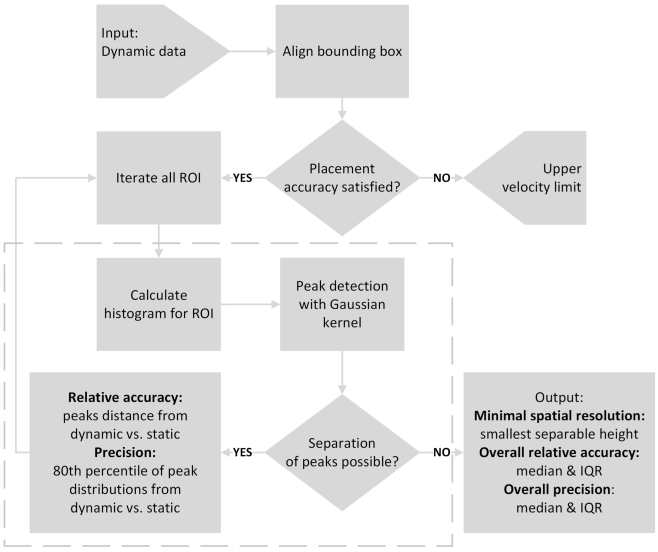

Fig. 6. Flowchart of the process to calculate minimal spatial resolution, relative accuracy, and precision. The analysis for a single ROI is indicated by the dashed border.

The next step involves placing a bounding box around the entire dynamic staircase model acquisitions and comparing all data inside the bounding box to the static ground-truth reference (Fig. 5). To fully leverage our pipeline automatically, the staircase motion path and image acquisition must be synchronized. The first option requires a hardware trigger between the camera and the actuator. Since this is not possible for all cameras, the second option is to record the actuator motion and trajectory beforehand and use this recording to predict the position of the staircase model during dynamic acquisition. If the motion path and image acquisition are not synchronized or the prediction is not accurate enough, the placement of the bounding box must be adjusted manually.

The bounding box is defined by

$$\begin{aligned} \mathrm{bb}_l &= (\eta + 1) \cdot s_d \\ \mathrm{bb}_w &= 2 \cdot s_w \\ \mathrm{bb}_h &= \sum_{i=1}^{\eta} s_{ir} \end{aligned} \tag{6}$$

where bb_l is the length of the bounding box, bb_w the width, bb_h the height, s_w the staircase tread width, s_d the staircase tread depth, s_{ir} the height of the ith right step, and η the number of steps.

The condition for the placement accuracy of the bounding box is described by

$$\mathrm{P}_{\mathrm{acc}} = \min\left(\frac{\mathrm{bb}_w}{8}, \frac{\mathrm{bb}_l}{4 \cdot (\eta + 1)}\right) \tag{7}$$

where bb_w is the width of the bounding box, bb_l the length of the bounding box, and η the number of steps. The condition includes a placement margin of 1/4 for each tread width and each tread depth. If this margin is too narrow, a slight misalignment may lead to border artifacts compromising the results. Therefore, the margin chosen is a compromise between robustness against placement errors and maximizing the region

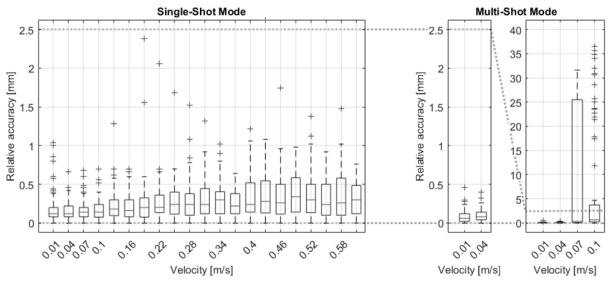

Fig. 7. Results for relative accuracy for single-shot (left) versus multishot (right) structured-light modes of the Photoneo MotionCam-3-D M+. Each boxplot for each velocity contains all relative accuracy outcomes of all regions of interest in which the two distributions of the tread surfaces could be separated.

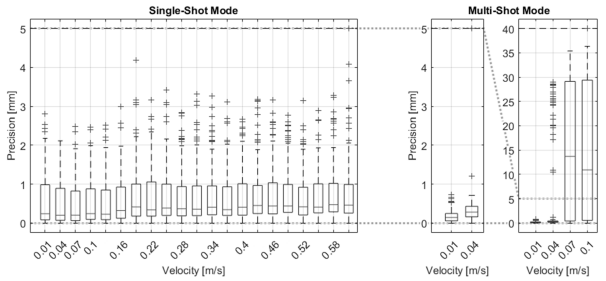

Fig. 8. Results for precision for single-shot (left) versus multishot (right) structured-light modes of the Photoneo MotionCam-3-D M+. Each boxplot for each velocity contains all precision outcomes of all regions of interest in which the two distributions of the tread surfaces could be separated.

used for the evaluation. There is no condition for the height, to include all outliers in the Z-direction.

In the presence of heavy artifacts or outliers, the placement accuracy condition will be violated, and the bounding box can no longer be placed with sufficient accuracy. The velocity at which the placement accuracy fails represents the upper limit for the 3-D imaging system, and the evaluation of higher velocities can be omitted.

The flowchart to calculate the minimal spatial resolution, accuracy, and precision is shown in Fig. 6.

III. Results

The specific application of interest in our case is to capture motion sequences of the human back with 3-D camera systems during medical examinations. During such medical examinations, patients perform various movements, e.g., bending forward or bending sideways. To capture the human back, a minimal spatial resolution of 2 mm is required [22], leading to a minimum desired accuracy ζ of 2 mm.

A. Phantom Model

With the minimum desired accuracy ζ of 2 mm and the number of steps $\eta = 9$, the step heights for the proposed phantom model are as follows [Section II-A, (1)–(3)].

$$s_{ir} = [2, 4, 6, 8, 10, 12, 14, 16, 8.5] \text{ mm}$$
$$s_{il} = [17, 15, 13, 11, 9, 7, 5, 3, 1] \text{ mm}$$
$$s_{im} = [15, 26, 33, 36, 35, 30, 21, 8, 0.5] \text{ mm}$$

where $i = 1, \ldots, 9$ and s_{ir} the ith right step, s_{il} the ith left step, and s_{im} the ith middle step.

The overall width and length scaling of the phantom model is determined by the choice of the tread width and depth. The values for the tread width and tread depth chosen for our application are 40 and 30 mm, respectively. This is a compromise that covers a considerable amount of a human back, has large enough tread surfaces, and remains compact.

B. Accuracy and Precision Analysis

The staircase model was divided into 27 ROIs (Section II-B) and the standard deviation of the Gaussian kernel used to detect the histogram peaks is $\sigma = 0.5$ mm ($\sigma = \zeta/4$).

The median and IQR of the relative accuracy for the single-shot mode of the Photoneo MotionCam is less than 0.3 and 0.5 mm for all velocities up to 0.6 m/s (Fig. 7), which is orders of magnitude less than the desired minimum accuracy ζ of 2 mm. The median and IQR of the relative accuracy for the multishot mode are both less than 0.1 mm for velocities up to 0.04 m/s but greater than the desired minimum accuracy for velocities above 0.04 m/s.

The median and IQR of the precision for the single-shot mode is less than 0.5 and 1.0 mm for all velocities up to 0.6 m/s (Fig. 8) and thus below the desired minimum accuracy ζ of 2 mm. The median and IQR of the precision for the multishot mode are both less than 0.3 mm for velocities up to 0.04 m/s but greater than the desired minimum accuracy for velocities above 0.04 m/s.

C. System Setup

According to the specific application of interest, the following parameters were defined for the system setup (Section II-C).

1) Minimum desired accuracy $\zeta = 2$ mm.
2) Distance between the measurement system and test object $d_{\text{mo}} = 1$ m. Our measurements showed that 1 m is a good compromise between optimal utilization of the camera field of view and sufficient distance from the human back to capture the border area as well.
3) Direction of motion: forward and backward motion along the image axis (Fig. 4, Z-axis) and a left and right motion parallel to the image plane (Fig. 4, Y-axis), both with an amplitude of ± 0.3 m from the initial position. The forward and backward motions cover the forward bending, the left and right motions cover the sideways bending, and our measurements showed that a typical range of motion of human back movements does not exceed ± 0.3 m. We used the KUKA LBR iiwa 7 R800 industrial robot to execute the motion with high repeatability.
4) Velocities of interest range from 0.01 to 0.6 m/s based on our own measurements of velocities during forward and sideways bending of the human back.
5) The 3-D camera under evaluation is the Photoneo MotionCam-3-D M+. The Photoneo MotionCam consists of a laser projector and a 2-D camera and has two modes, a multishot structured-light mode in which multiple stripe patterns are projected, and a single-shot

structured-light mode in which a single pattern is projected [23]. The laser power was reduced to 50% (value 2000) and the exposure times were set to 40.96 ms for the single-shot mode and 10.24 ms for the multishot mode. The Photoneo MotionCam is precalibrated; no additional calibration was necessary. The environmental conditions are usual daylight indoor conditions, without direct sunlight or artificial lighting.

6) The image plane and image axis alignment with the tread surface $A_{\eta l}$ was done directly with captures from the Photoneo MotionCam. The tread surface $A_{\eta l}$ was parallel to the image plane and at the origin of the X- and Y-axes.

D. Data Acquisition and Processing

All measurements were taken in the same indoor conditions (Section III-C), and the resulting output was saved in PLY format in millimeters. The ground-truth reference for the staircase model was created with the multishot mode of the Photoneo MotionCam. Photoneo reports an accuracy of < 0.3 mm [23] and thus the MotionCam is well suited to capture a static ground-truth reference for our application. For each defined velocity (Section III-C), three dynamic captures were taken for both the multishot and the single-shot modes.

We recorded the actuator motion and trajectory beforehand and used this recording to predict the position of the staircase model during the dynamic acquisition. No manual adjustment of the placement of the bounding box was necessary. The bounding box dimensions are as follows:

$$\begin{aligned} \mathrm{bb}_l &= 300 \text{ mm} \\ \mathrm{bb}_w &= 80 \text{ mm} \\ \mathrm{bb}_h &= 81 \text{ mm} \end{aligned}$$

resulting in a placement accuracy of

$$P_{\mathrm{acc}} = 7.5 \text{ mm}.$$

The placement accuracy condition (7) was fulfilled for the single-shot mode for all defined velocities, but was exceeded for the multishot mode at velocities higher than 0.04 m/s, and thus the measurement series was stopped at 0.1 m/s.

IV. Discussion

The literature describes several approaches to evaluating the spatial resolution, accuracy, and precision of 3-D surface imaging systems for applications assessing moving objects. In most cases, multiple measurements of different types of phantom models are performed. The models include flat surfaces, spheres, random and complex objects, and human body parts [10], [11], [13], [14], [15], [16], [24], [25], [26], [27], [28], [29]. Prior to the quality analysis, the acquired 3-D image is often registered to the model. In most cases, iterative closest-point algorithms (ICP) are used [10], [11], [12], [13], [14], [16], [24], [26], [27], [28]. The quality measure is often the root mean square error (RMSE) of a single or a few registered frames [10], [11], [12], [13], [14], [16], [28], but such small sample sizes make the analysis vulnerable. In addition, RMSE is not directly transferrable to accuracy and precision. Finally, most proposed methods use uncontrolled motion, e.g., performed a single time by hand, or only use static objects [10], [12], [13], [14], [15], [16], [25], [26], [28], [29]. The variety of phantom models and quality measures used by different research groups makes it impossible to compare results. Multiple evaluations of movement in a controlled manner are also hampered by the displacement of the object by hand. Finally, none of the referenced approaches provide quantification of spatial resolution, accuracy, and precision.

The method presented in this article allows spatial resolution, accuracy, and precision to be evaluated within a single approach. The proposed method reduces sources of error since it does not depend on registration or object fitting. The analysis is robust due to the large sample sizes (three frames with 27 steps each resulting in 81 values per measurement). In addition, our method allows for repeatable path and velocity because of the controlled motion of the robot. Subsequently, this allows direct comparison of different camera systems, even if evaluated at different locations and time points.

To ensure a proper application of our method, the following aspects (Section II-C) must be defined with care and stated together with the results: minimum desired accuracy ζ, distance between measurement system, and test object d_{mo}, direction of motion including used range, velocities of interest, measurement system settings including projector brightness and exposure time, and environmental conditions. The reference imaging system used must have an accuracy and precision magnitudes higher than the minimal desired accuracy. Using an unsuitable reference imaging system will influence the results. If a high-quality imaging system is not available, the CAD model can be used as a reference. However, this requires an initial registration of the CAD model with the first captured frame. As discussed above, registration influences the result. Yet another important aspect is the choice of the standard deviation (σ) used for the peak detection. If σ is too small, more than two distinct peaks will be detected. If σ is too large, the peaks detected will shift. We define σ according to the minimal desired accuracy. Depending on the 3-D imaging system used, σ may need to be adjusted. As a result, we advise visually inspecting the histogram with peaks to ensure a suitable σ.

The following limitations may be improved or overcome. The proposed staircase phantom focuses on the analysis in the Z-direction. This is the direction most often used in the evaluation of depth measurement systems. Expanding the phantom model to cover distinct landmarks in the X- and Y-directions would additionally allow a lateral spatial resolution to be evaluated, providing more information, and potentially extending the range of use cases. In addition, the step heights of the proposed phantom change linearly, whereas the 1951 USAF resolution test chart consists of a series of logarithmic steps covering a larger bandwidth. The suggested stair model could be rearranged accordingly. The idealization of the target objects as flat surfaces and step sizes may not cover all cases. Some applications may require lateral resolution, spherical surfaces, rough surfaces, or surfaces with different materials. This will be the focus of future work.

The alignment of the phantom model with the imaging system is very important. To correct a small misalignment of the phantom model with the image plane, the accuracy and precision are always calculated relative to the reference. This corrects small deviations from the image plane. However, this is not the case for the resulting occlusions from the higher stair treads onto the lower stair treads. Furthermore, moving the phantom model in the direction of the lower stair treads, e.g., in a negative Y-direction and a positive X-direction (Fig. 1), is only limitedly possible because the higher stair treads will occlude the lower ones. This could be overcome by adjusting the phantom model, e.g., by expanding the staircase with another row of steps in positive Y-direction and negative X-direction.

Furthermore, the evaluation algorithm could be improved to be more robust against partial occlusions. The evaluation of motion in the Z-direction (Fig. 1) is well-tested and validated. For evaluating motion in the image plane (X- and Y-directions), the orientation of the phantom model is important. We suggest moving the phantom model in a transverse direction. Moving the phantom model in a longitudinal direction might lead to outliers being shifted from neighboring stair treads, especially for high-velocity motion, and should be further investigated.

The placement accuracy condition of the bounding box is verified manually and inspected visually. This could be replaced with an automatic validation, which may lead to a more robust calculation of the upper-velocity limit; an automatic validation is more deterministic. However, such an automatization is not trivial due to outliers and motion artifacts for high-speed motions. Last, the proposed method has only been applied to structured light systems so far. Multishot structured light in particular produces very specific motion artifacts. Therefore, the proposed method should also be tested with other 3-D measurement modalities.

Capturing the motion sequences of the human back with a 3-D camera system is our specific application. A human back can be considered as a smooth surface and a rigid body. The proposed phantom model contains flat surfaces and steps. The dispersion measure of the flat surfaces defines how well a smooth surface can be captured (precision). The measured step heights define how well small deviations can be detected (accuracy). In this way, our phantom model enables the evaluation of smooth objects in motion. To confirm the requirements for individual applications, the authors recommend specific measurements in addition to our proposed method. These measurements should be conducted on the target object itself or on models that represent the corresponding target object.

V. Conclusion

In 2-D imaging, the 1951 USAF resolution test chart is an accepted standard and is still widely used. In 3-D, the most commonly used evaluation indicators in the literature are the RMSE of a planar surface, the error in diameter of a spherical object, and the RMSE of complex objects. The RMSE for a planar surface is calculated with its best-fitting plane. The error in the diameter of a spherical object is calculated with its best-fitting sphere. Respectively, a general reconstruction error of complex objects is compared to its ground-truth model, which is usually the RMSE. However, no international standard exists for evaluating the spatial resolution, accuracy, and precision of 3-D surface imaging systems. We propose a phantom model and a method to be used as part of such a standard. The phantom model is a compact extension of the 1951 USAF to 3-D. The different step heights of our staircase model reproduce a series of steps in the spatial domain. The overall precision of our proposed approach represents the RMSE evaluation indicator of a planar surface from the literature. With the overall accuracy, we propose an alternative to the error in diameter of a spherical object. The suggested alternative is not based on only a single value but on the evaluations of 27 different step heights for each frame. Subsequently, our suggested alternative covers a wider spectrum and is more robust. However, we did not investigate the difference between using step heights compared to spherical surfaces, which will be the focus of future work. Future investigations also include lateral resolution, roughness, and different materials. To confirm the requirements for individual applications, the authors recommend specific measurements, e.g., with complex objects, in addition to our proposed method.

The proposed method specifies how to set up the system and acquire, process, and analyze the data to obtain a statement about the spatial resolution, accuracy, and precision of 3-D imaging systems for capturing static and moving objects. We provide all details including the code[1] so that other research groups can easily apply our method to characterize and compare various other 3-D imaging systems. With the practical application of the Photoneo MotionCam-3-D, we show that the proposed phantom model and method are suitable for evaluating the spatial resolution, accuracy, and precision of 3-D imaging systems. The Photoneo MotionCam-3-D is a high-accuracy 3-D scanner with two structured-light modes: a multishot scanner mode with high accuracy for static objects and a single-shot camera mode with lower accuracy for moving objects. As expected, the application of our method confirmed that multishot structured light is very accurate and precise for static scenes and slowly moving objects. However, multishot structured light is not suitable for moving objects. This is because a single 3-D reconstruction consists of several successive 2-D images: in the case of the Photoneo MotionCam-3D, up to 24 images over 800 ms. The application also confirmed that single-shot structured light is less accurate and precise than multishot but can capture moving objects without much loss of accuracy or precision.

In summary, a standard method has been proposed for evaluating 3-D imaging systems' measurement quality with moving objects. The proposed method is independent of the 3-D measurement modality. The method has been tested with a structured light system, and it will be interesting to apply the method to other modalities, e.g., active stereoscopy and time of flight, and to compare the results of these different modalities.

[1]The data and MATLAB code presented in this article are openly available on GitHub at https://github.com/mkaisereth/InfluenceOfMovingObjects.

REFERENCES

[1] T. Bassani, E. Stucovitz, F. Galbusera, and M. Brayda-Bruno, "Is rasterstereography a valid noninvasive method for the screening of juvenile and adolescent idiopathic scoliosis?" *Eur. Spine J.*, vol. 28, no. 3, pp. 526–535, Mar. 2019, doi: 10.1007/s00586-018-05876-0.

[2] L. Marin et al., "Acute effects of self-correction on spine deviation and balance in adolescent girls with idiopathic scoliosis," *Sensors*, vol. 22, no. 5, p. 1883, Feb. 2022, doi: 10.3390/s22051883.

[3] S. Paśko and W. Glinkowski, "Combining 3D structured light imaging and spine X-ray data improves visualization of the spinous lines in the scoliotic spine," *Appl. Sci.*, vol. 11, no. 1, p. 301, Dec. 2020, doi: 10.3390/app11010301.

[4] D. Ledwoń et al., "Real-time back surface landmark determination using a time-of-flight camera," *Sensors*, vol. 21, no. 19, p. 6425, Sep. 2021, doi: 10.3390/s21196425.

[5] J. A. Albert, V. Owolabi, A. Gebel, C. M. Brahms, U. Granacher, and B. Arnrich, "Evaluation of the pose tracking performance of the Azure Kinect and Kinect v2 for gait analysis in comparison with a gold standard: A pilot study," *Sensors*, vol. 20, no. 18, p. 5104, Sep. 2020, doi: 10.3390/s20185104.

[6] Z. Xu et al., "Back shape measurement and three-dimensional reconstruction of spinal shape using one Kinect sensor," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2020, pp. 1–5, doi: 10.1109/ISBI45749.2020.9098481.

[7] H. Rehouma, R. Noumeir, S. Essouri, and P. Jouvet, "Advancements in methods and camera-based sensors for the quantification of respiration," *Sensors*, vol. 20, no. 24, p. 7252, Dec. 2020, doi: 10.3390/s20247252.

[8] T. Kokabu et al., "Three-dimensional depth sensor imaging to identify adolescent idiopathic scoliosis: A prospective multicenter cohort study," *Sci. Rep.*, vol. 9, no. 1, pp. 1–8, Jul. 2019, doi: 10.1038/s41598-019-46246-0.

[9] K. W. Nam, J. Park, I. Y. Kim, and K. G. Kim, "Application of stereo-imaging technology to medical field," *Healthcare Informat. Res.*, vol. 18, no. 3, p. 158, 2012, doi: 10.4258/hir.2012.18.3.158.

[10] M. Duan, Y. Jin, H. Chen, J. Zheng, C. Zhu, and E. Chen, "Automatic 3-D measurement method for nonuniform moving objects," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021, doi: 10.1109/TIM.2021.3106119.

[11] C. Sui, K. He, C. Lyu, and Y.-H. Liu, "Accurate 3D reconstruction of dynamic objects by spatial–temporal multiplexing and motion-induced error elimination," *IEEE Trans. Image Process.*, vol. 31, pp. 2106–2121, 2022, doi: 10.1109/TIP.2022.3150297.

[12] R. Sagawa, R. Furukawa, and H. Kawasaki, "Dense 3D reconstruction from high frame-rate video using a static grid pattern," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 9, pp. 1733–1747, Sep. 2014, doi: 10.1109/TPAMI.2014.2300490.

[13] T. Weise, B. Leibe, and L. Van Gool, "Fast 3D scanning with automatic motion compensation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8, doi: 10.1109/CVPR.2007.383291.

[14] P. Wissmann, R. Schmitt, and F. Forster, "Fast and accurate 3D scanning using coded phase shifting and high speed pattern projection," in *Proc. Int. Conf. 3D Imag., Model., Process., Visualizat. Transmiss.*, May 2011, pp. 108–115, doi: 10.1109/3DIMPVT.2011.21.

[15] M. Schmidt and B. Jähne, "Efficient and robust reduction of motion artifacts for 3D time-of-flight cameras," in *Proc. Int. Conf. 3D Imag. (IC3D)*, Dec. 2011, pp. 1–8, doi: 10.1109/IC3D.2011.6584391.

[16] S. Giancola, M. Valenti, and R. Sala, *A Survey on 3D Cameras: Metrological Comparison of Time-of-Flight, Structured-Light and Active Stereoscopy Technologies* (SpringerBriefs in Computer Science). Cham, Switzerland: Springer, 2018, pp. 89–90. [Online]. Available: https://link.springer.com/book/10.1007/978-3-319-91761-0, doi: 10.1007/978-3-319-91761-0.

[17] P. L. Reu, W. Sweatt, T. Miller, and D. Fleming, "Camera system resolution and its influence on digital image correlation," *Exp. Mech.*, vol. 55, no. 1, pp. 9–25, Jan. 2015, doi: 10.1007/S11340-014-9886-Y.

[18] S. D. Alaruri, "Practical methods for characterizing the optical performance of digital camera-based imaging systems," *Int. J. Electron., Commun., Meas. Eng.*, vol. 9, no. 1, pp. 31–45, Jan. 2020, doi: 10.4018/ijecme.2020010103.

[19] A. K. Singh, G. Pedrini, M. Takeda, and W. Osten, "Scatter-plate microscope for lensless microscopy with diffraction limited resolution," *Sci. Rep.*, vol. 7, no. 1, pp. 1–8, Sep. 2017, doi: 10.1038/s41598-017-10767-3.

[20] B. Xu et al., "Metalens-integrated compact imaging devices for wide-field microscopy," *Adv. Photon.*, vol. 2, no. 6, Nov. 2020, Art. no. 066004, doi: 10.1117/1.ap.2.6.066004.

[21] Z. Shen, A. Nacev, A. Sarwar, R. Lee, D. Depireux, and B. Shapiro, "Automated fluorescence and reflectance coregistered 3-D tissue imaging system," *IEEE Trans. Magn.*, vol. 49, no. 1, pp. 279–284, Jan. 2013, doi: 10.1109/TMAG.2012.2222360.

[22] M. Kaiser et al., "Minimal required resolution to capture the 3D shape of the human back—A practical approach," *Sensors*, vol. 23, no. 18, p. 7808, Sep. 2023, doi: 10.3390/s23187808.

[23] *MotionCam-3D M+–Photoneo*. Accessed: Mar. 6, 2023. [Online]. Available: https://www.photoneo.com/de/products/motioncam-3d-m-plus/

[24] C. Sui, K. He, C. Lyu, Z. Wang, and Y.-H. Liu, "Active stereo 3-D surface reconstruction using multistep matching," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 4, pp. 2130–2144, Oct. 2020, doi: 10.1109/TASE.2020.2991803.

[25] R. Sagawa, Y. Matsumoto, H. Kawasaki, and R. Furukawa, "Parallel processing of grid-based one-shot structured-light system for online 3D reconstruction of moving objects," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2014, pp. 1976–1982, doi: 10.1109/ROBIO.2014.7090626.

[26] Y. Li and Z. Wang, "RGB line pattern-based stereo vision matching for single-shot 3-D measurement," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021, doi: 10.1109/TIM.2020.3041086.

[27] C. Duan, J. Tong, L. Lu, J. Xi, Y. Yu, and Q. Guo, "Improving the performance of 3D shape measurement of moving objects by fringe projection and data fusion," *IEEE Access*, vol. 9, pp. 34682–34691, 2021, doi: 10.1109/ACCESS.2021.3061415.

[28] M. Maruyama, S. Tabata, Y. Watanabe, and M. Ishikawa, "Multi-pattern embedded phase shifting using a high-speed projector for fast and accurate dynamic 3D measurement," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 921–929, doi: 10.1109/WACV.2018.00106.

[29] H. Kayaba and Y. Kokumai, "Non-contact full field vibration measurement based on phase-shifting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2548–2556, doi: 10.1109/CVPR.2017.273.