ASSESSING THE ALIGNMENT BETWEEN GEOMETRY AND COLORS IN TLS COLORED POINT CLOUDS

Zhaoyi Wang, Matej Varga, Tomislav Medić, Andreas Wieser

Institute of Geodesy and Photogrammetry, ETH Zurich, Zurich, Switzerland
(zhaoyi.wang, matej.varga, tomislav.medic, andreas.wieser)@geod.baug.ethz.ch

KEY WORDS: Terrestrial laser scanning (TLS), scanner calibration, point cloud colorization, built-in RGB camera

ABSTRACT:

The integration of the color information from RGB cameras with the point cloud geometry is used in numerous applications. However, little attention has been paid on errors that occur when aligning colors to points in terrestrial laser scanning (TLS) point clouds. Such errors may impact the performance of algorithms that utilize colored point clouds. Herein, we propose a procedure for assessing the alignment between the TLS point cloud geometry and colors. The procedure is based upon identifying artificial targets observed in both LiDAR-based point cloud intensity data and camera-based RGB data, and quantifying the quality of the alignment using differences between the target center coordinates estimated separately from these two data sources. Experimental results with eight scanners show that the quality of the alignment depends on the scanner, the software used for colorizing the point clouds, and may change with changing environmental conditions. While we found the effects of misalignment to be negligible for some scanners, they exhibited clearly systematic patterns exceeding the beam divergence, image and scan resolution for four of the scanners. The maximum deviations were about 2 mrad perpendicular to the line-of-sight when colorizing the point clouds with the respective manufacturer's software or scanner in-built functions, while they were up to about 5 mrad when using a different software. Testing the alignment quality, e.g., using the approach presented herein, is thus important for applications requiring accurate alignment of the RGB colors with the point cloud geometry. [1]

1. INTRODUCTION

Point clouds and RGB images can provide complementary information due to their different sensing modalities and resolutions. Colored point clouds, as products of incorporating point cloud geometry and RGB colors, recently have shown their remarkable benefits in numerous applications compared to pure point cloud geometry. These applications include high-level tasks, such as 3D object detection (Liang et al., 2018) and semantic segmentation (Kweon and Yoon, 2022), and also low-level tasks, e.g., point cloud registration (Zhang et al., 2022) and point cloud completion (Aiello et al., 2022).

While there are numerous studies that address the joint processing of color information and point clouds using data from RGB-D cameras, very few have focused on terrestrial laser scanning (TLS) in this context. RGB-D cameras generate RGB colors associated with planar image coordinates using an RGB camera and depth coordinates utilizing a dedicated depth sensor or a photogrammetric solution based on multiple cameras within a single housing. It is necessary to align the depth sensor to the RGB camera(s), and the alignment details about either parameter values or solutions are often openly available (Basso et al., 2018). However, compared to RGB-D cameras, TLS scanners can generate more accurate point clouds with a wider field of view (FoV). TLS scanners obtain point cloud geometry and RGB colors using separate sensors, typically the LiDAR scanning unit and one or multiple built-in RGB cameras. The alignment of these separate sensors is neither perfectly known nor can it be assumed stable over time, particularly with temperature changes, transport, and scanner use. Moreover, TLS scanner manufacturers enable users to assign RGB colors to point clouds using their software, details regarding the colorization procedure and the accuracy of transformation parameter values are not openly available to users. Hence, it remains unclear whether an extra up-to-date user-based alignment is necessary.

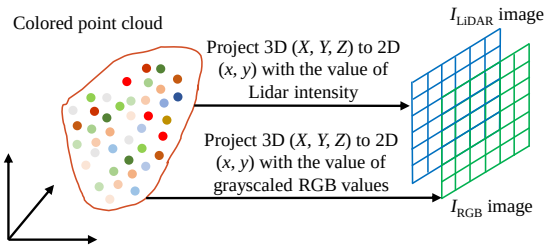


Figure 1. The concept of assessing the alignment between the TLS point cloud geometry and colors. A TLS-colored point cloud of an artificial target is projected independently to 2D (X, Y) images with the value of LiDAR intensity (I_{LiDAR}) and the value of grayscaled RGB values (I_{RGB}).

On the other hand, RGB colors in TLS-colored point clouds are often used for visualization purposes (Balado et al., 2023), with limited research focusing on their accuracy and quality. Pleskacz and Rzonca (2016) investigate the correctness of RGB colors and interpolation methods for colored point clouds. Similarly, Julin et al. (2020) evaluate the quality of RGB colors generated from built-in cameras of TLS scanners. Herein, our focus lies instead on assessing the alignment between point cloud geometry and colors, which is essential for multi-modal fusion-based research (Feng et al., 2019; Pham et al., 2020).

We attempt to clarify how to effectively and efficiently quantify the alignment quality between TLS point cloud geometry and colors. To address this problem, we develop a procedure that

[1] Our code is available at: https://github.com/zhaoyiww/AssessAlignment

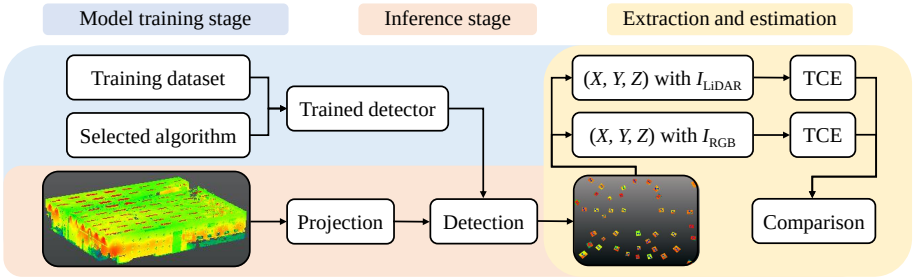


Figure 2. The proposed procedure and its building blocks for assessing the alignment quality. During the model training stage, we train target detectors using the training dataset and object detection algorithms (only one-time effort). The proposed procedure contains the inference stage, target extraction and target center estimation. During the inference stage, the point cloud is used as input and is projected onto a panoramic image. Targets are then detected in the panoramic image through using one of the trained target detectors.

comprises point cloud projection, target detection, extraction, center estimation and analysis of the resulting discrepancies. The core idea is that we can identify artificial targets in the scene independently using the RGB data and LiDAR intensity data (or received signal power). LiDAR intensity data and distance are measured simultaneously using the same laser beam, which enables us to interpret the intensity data as perfectly aligned with the point cloud geometry. On the other hand, the RGB data generated by separate camera(s) are projected onto the point cloud such that their alignment with the geometry depends on the quality of the projection. By comparing the corresponding target center coordinates estimated using (i) LiDAR intensity and (ii) RGB colors, we can analyze the alignment quality between the TLS point cloud geometry and colors (see Figure 1). Our contributions are summarized as follows:

- We develop a simple but efficient procedure for assessing the alignment between the TLS point cloud geometry and colors by comparing the discrepancy between the target center coordinates estimated independently using (i) LiDAR intensity and (ii) RGB color data.
- We evaluate this alignment quality for eight TLS scanners, and analyze the magnitudes and spatial distribution of the identified deviations. We also investigate the alignment stability with temperature changes for four of the scanners.

2. THEORETICAL BACKGROUND

This section begins with an explanation of how RGB colors can be integrated with TLS point clouds. Then, we briefly describe how we utilize current state-of-the-art object detection algorithms to detect artificial targets.

2.1 Principle of TLS point cloud colorization

The details of how specific TLS scanners or software packages assign RGB colors to point clouds are often unrevealed. However, the main principles are publicly available in some software manuals and scanner patents, e.g., Steffey et al. (2018) and Vollrath and Ossig (2018). For example, TLS scanners that have integrated or externally attached RGB cameras capture point clouds and RGB images separately and sequentially. To project the images onto the point cloud, they can either be combined into a full panoramic image and transformed or directly projected using the collinearity equations with intrinsic and extrinsic parameters of the cameras within the coordinate system of the scanner. The manufacturer typically determines the numeric values of these parameters, which are stored in the scanner or its own software but not made accessible to users. Interpolation techniques, such as bilinear or cubic methods, are necessary to determine the RGB values for each point in the TLS point clouds, given the difference in resolutions and parallax resulting from the camera mounting, which may not be perfectly collinear.

2.2 Object detection

Object detection aims to recognize and locate objects of interest within an image. This is achieved through using deep learning algorithms that return bounding boxes indicating the image coordinates and predicted class of each detected object. The primary objective of using object detection algorithms in this study is to automatically detect artificial targets and define bounding boxes for further target extraction and center estimation. To achieve this objective, we employ two common groups of object detection algorithms: two-stage and one-stage object detection algorithms (Zaidi et al., 2022), as each has its own unique advantages (e.g., high accuracy or high efficiency) and is still widely used.

The two-stage detection algorithm first generates candidate anchor boxes in the image. These boxes are then fed into the network. Examples of commonly used two-stage detection algorithms include Faster R-CNN (Ren et al., 2015) and its upgraded version, Mask R-CNN (He et al., 2017). In contrast, the one-stage detection algorithm such as YOLO (Redmon et al., 2016) divides the image into small regions and directly carry out object detection in each of them.

3. THE PROPOSED PROCEDURE AND METRICS USED FOR ASSESSING THE ALIGNMENT

Our procedure employs artificial, planar black-and-white targets. The rationale for this choice is based on the high accuracy achievable for target center estimation (TCE) compared to the scanning resolution (Omidalizarandi et al., 2019b; Janßen et al., 2019). This enables TCE utilizing both LiDAR intensity data and RGB data, without sacrificing the alignment quality assessment. It is worth noting that extending the procedure to use natural features for the assessment remains a direction for future investigations.

Moreover, we assume that the scans for the assessment are conducted in an environment with a relatively uniform distribution of targets throughout the FoV of the scanner. Without loss of generality, we will use targets distributed in a single, large room herein. We illustrate the proposed procedure and its building

blocks in Figure 2. The procedure consists of (1) a projection of the 3D point cloud onto a 2D LiDAR intensity image, (2) target detection within this image utilizing a trained detector, (3) target extraction from the 3D point cloud, (4) target center estimation independently using the LiDAR intensities and camera-based intensities (derived from the RGB data) for each detected target, and finally (5) calculation and analysis of the differences between the target center coordinates of the same target extracted from the two different intensity data sets.

3.1 Point cloud projection

Let $\mathcal{G} \in \mathbb{R}^{m\times 7}$ denotes a point cloud with m points x 7 dimensions (i.e., X, Y, Z, LiDAR intensity, R, G, B). To enhance the processing speed while preserving the target quality that only used for target detection, we downsample $\mathcal{G}$ to $\mathcal{G}_{\mathrm{D}}$. We transform the Cartesian coordinates of the points in $\mathcal{G}_{\mathrm{D}}$ to polar ones:

$$\begin{cases} S = \sqrt{X^2 + Y^2 + Z^2} \\ H = \arctan(Y/X) \\ V = \arccos(Z/S) \end{cases} \tag{1}$$

where S = the distance from local coordinate origin
H, V = the horizontal and vertical angle, respectively.

We then project the point cloud $\mathcal{G}_{\mathrm{D}}$ onto a panoramic image with dimensions (H, V) where the LiDAR intensities serve as the pixel values for the image. This panoramic image is subsequently employed for the target detection.

3.2 Target detection and extraction

Target detection involves utilizing a trained target detector (see Figure 2). It is worth noting that the image sizes may vary, which can lead to a situation where the detectors that are trained on a certain size of images and be used to infer results on images with unequal sizes. Fortunately, Feature Pyramid Network (FPN) (Lin et al., 2017) integrated in the detection algorithms enable us to overcome this. Thus, during the inference stage, we detect targets directly on the projected panoramic image. Once a target is detected, its region in the image is determined by its upper-left (H_0, V_0) and lower-right image coordinates $(H_0 + h, V_0 + w)$ (see Figure 3).

For target extraction, we compute the Cartesian coordinates (S, H, V) of the point cloud $\mathcal{G}$ using Equation 1, then derive the indexes of (H, V) that within the image region. These indexes also correspond to the target region in the point cloud $\mathcal{G}$, as (H, V) and (X, Y, Z) are one-to-one corresponding. We therefore extract the point cloud $\mathcal{T} \in \mathbb{R}^{n\times 7}$ for each target. Then, we save $\mathcal{P} \in \mathbb{R}^{n\times 4}$ and $\mathcal{Q} \in \mathbb{R}^{n\times 6}$ as subsets of $\mathcal{T}$, which correspond to $(X, Y, Z,$ LiDAR Intensity) and (X, Y, Z, R, G, B), respectively. Both $\mathcal{P}$ and $\mathcal{Q}$ are used to estimate the target center. Note that (R, G, B) values in $\mathcal{Q}$ will then be converted to grayscale values (i.e., I_{RGB}).

3.3 Target center estimation

The estimation of target center coordinates is performed by following the steps mentioned in Janßen et al. (2019). Initially, we apply RANSAC (Fischler and Bolles, 1981) with a small threshold (e.g., 0.05 m) to detect and remove outliers while estimating the initial plane parameter values based on (X, Y, Z). The initial values are then refined through Gauss-Helmert model-based plane estimation. Further, the target point cloud is orthogonally projected to this plane. To determine the target center coordinates, we adopt a sub-pixel image registration method (Guizar-Sicairos et al., 2008). This method computes the 2D transformation information (i.e., rotation and translation) between the target image and its corresponding template image, with center coordinates (0, 0). Using this transformation information and the template center coordinates, we determine 2D target center coordinates in the target plane. We further obtain 3D target center coordinates using the target plane parameters.

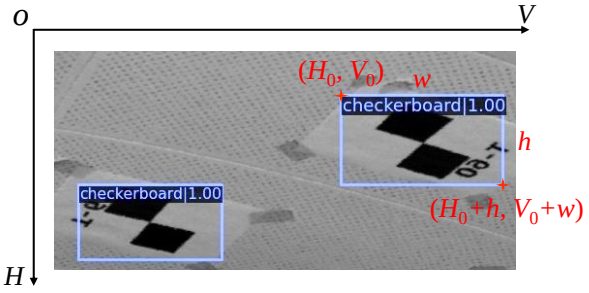


Figure 3. Detected targets with bounding boxes. w and h denote the width and height of the detected target, respectively.

We estimate the 3D target center coordinates from I_{LiDAR} (i.e., LiDAR intensity) and I_{RGB} target images independently. This enables us to analyze the discrepancy between point cloud geometry and colors.

3.4 Metrics for assessing the alignment

In order to assess the quality of the alignment in later experiments, we define two metrics, namely homogenized horizontal and vertical angle deviation, and angular deviation.

3.4.1 homogenized horizontal and vertical angle deviation

Let $(H_{\mathrm{LiDAR}}, V_{\mathrm{LiDAR}})$ and $(H_{\mathrm{RGB}}, V_{\mathrm{RGB}})$ denote the polar coordinates of the target center estimated from I_{LiDAR} (i.e., LiDAR intensity) and I_{RGB} target images, respectively. To visualize deviations on vertical-horizontal angle deviation maps, we compute $(H_{\mathrm{LiDAR}}, V_{\mathrm{LiDAR}})$ and $(H_{\mathrm{RGB}}, V_{\mathrm{RGB}})$ for each target center using Equation 1. Let $(\Delta H, \Delta V)$ denote the original horizontal and vertical angle deviation of the same target center, which can be computed using:

$$\begin{cases} \Delta H = H_{\mathrm{LiDAR}} - H_{\mathrm{RGB}} \\ \Delta V = V_{\mathrm{LiDAR}} - V_{\mathrm{RGB}} \end{cases}. \tag{2}$$

However, due to the spherical sampling pattern of TLS scanners, ΔH becomes exaggerated with higher elevation angles. Thus, we use the sine of the vertical angle acting as a scaling factor to correct such an exaggeration. Let $(\Delta\hat{H}, \Delta\hat{V})$ denote the homogenized horizontal and vertical angle deviation of the same target center, and we can compute them using:

$$\begin{cases} \Delta\hat{H} = \Delta H \sin V_{\mathrm{LiDAR}} \\ \Delta\hat{V} = \Delta V \end{cases}. \tag{3}$$

3.4.2 Angular deviation We compute joint angular deviation using:

$$\Delta\hat{A} = \sqrt{\Delta\hat{H}^2 + \Delta\hat{V}^2}. \tag{4}$$

The beam divergence, image resolution, and scan resolution can reflect the achievable level of detail (LoD) of each scan. Hence,

to assess the magnitude of these angular deviations relative to the achievable LoD, we introduce the maximum of the beam divergence, image and scan resolution as a threshold for classifying deviations as outlying. To represent all three factors in the same units, we neglect the beam waist and consider only the beam divergence instead of the physically more meaningful beam diameter. The beam divergence, image and scan resolution values for all investigated scanners are listed in Table 1.

4. EVALUATION OF THE TARGET DETECTION AND EXTRACTION

In this section, we describe the dataset we used for training target detectors, implementation details, and evaluate the performance of the target detection and extraction.

4.1 Dataset

As manual annotation of targets is labor-intensive, the training dataset comprises both manually annotated and synthetic targets. The dataset comprises 2D images with standardized dimensions of 800 pixels (vertically) x 1333 pixels (horizontally). The targets in these 2D images contain four categories: checkerboard, rotated checkerboard, round checkerboard, and BOTA8, as shown in Figure 4.

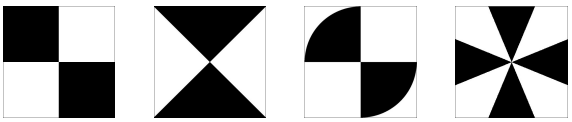


Figure 4. Types of targets used during the model training stage. From left to right: checkerboard, rotated checkerboard, round checkerboard, and BOTA8 (Janßen et al., 2019).

4.2 Implementation details

We conduct all experiments on a single RTX 3090 Ti GPU and implement all model training using PyTorch (Paszke et al., 2017). We train four different target detectors based on Faster R-CNN, Mask R-CNN, YOLOX (Ge et al., 2021) and YOLOv7 (Wang et al., 2022). Target center estimation is performed in MATLAB R2022a, primarily using the source code provided by Janßen et al. (2019), which are further integrated into the automatic procedure. Specifically, we implement Faster R-CNN and Mask R-CNN using MMDetection toolbox (Chen et al., 2019), while we implement the other two using the source code from official documentations. We train all these detectors using corresponding pre-trained models, which means we can easily achieve satisfactory results with much less training time compared to training from scratch. Specifically, we train Faster R-CNN with linear 500 warm-up iterations, with a batch size of 16 using a SGD optimizer with learning rate 0.02 and momentum 0.9, and weight decay 0.0001. We train the network in 24 epochs, and use a pre-trained model ResNet101. We train Mask R-CNN using the same hyperparameters as Faster R-CNN. Further, we train YOLOX with a batch size of 6 in 300 epochs, based on a pre-trained model YOLOX-x. We train YOLOv7 using a batch size 4, dataloader workers 4, and 300 epochs, based on a pre-trained model YOLOv7-E6E.

4.3 Performance

We train different target detectors on the training dataset. To compare the performance, we test all detectors on the same test set, i.e., a test set that contains only manually annotated targets. In addition, we use mean Average Precision (mAP) (Ren et al., 2015), a commonly used metric for object detection, to compare these detectors. Among these detectors, the detector based on YOLOv7 performs best, but the difference among them are very small. We further address how efficient our procedure for detecting and extracting targets compared to manual extraction. Through testing on a point cloud with 50 million points and roughly 250 targets, the manual method using CloudCompare needs approximately 50 minutes to extract approximately 90% targets, whereas our automatic procedure detects and extracts the same number of targets within 3 minutes. The improved efficiency of our proposed procedure enables us to conduct later experiments at scale sufficient for generalizable assessment of the alignment quality in commercial TLS scanners.

5. EXPERIMENTS

We utilize the proposed procedure to process data from two types of alignment experiments: (i) Experiment A, which assesses the alignment quality using eight TLS scanners, and (ii) Experiment B, which evaluates the alignment stability with changes of the ambient temperature using four TLS scanners.

5.1 Experiment A: alignment quality assessment

We set up this experiment in a room of approximately 10 x 10 x 3 m^3. Roughly 250 artificial targets (checkerboard, rotated checkerboard, and round checkerboard printed on A4 paper) are approximately uniformly distributed on the ceiling, walls, and floor. To obtain more generalizable results, we use eight TLS scanners, including two Faro Focus3D X330 (indicated as Faro Focus3D X330 (a) and (b) subsequently). We place these scanners on a tripod in the middle of the room, one at a time, and scan the entire FoV for each setup.

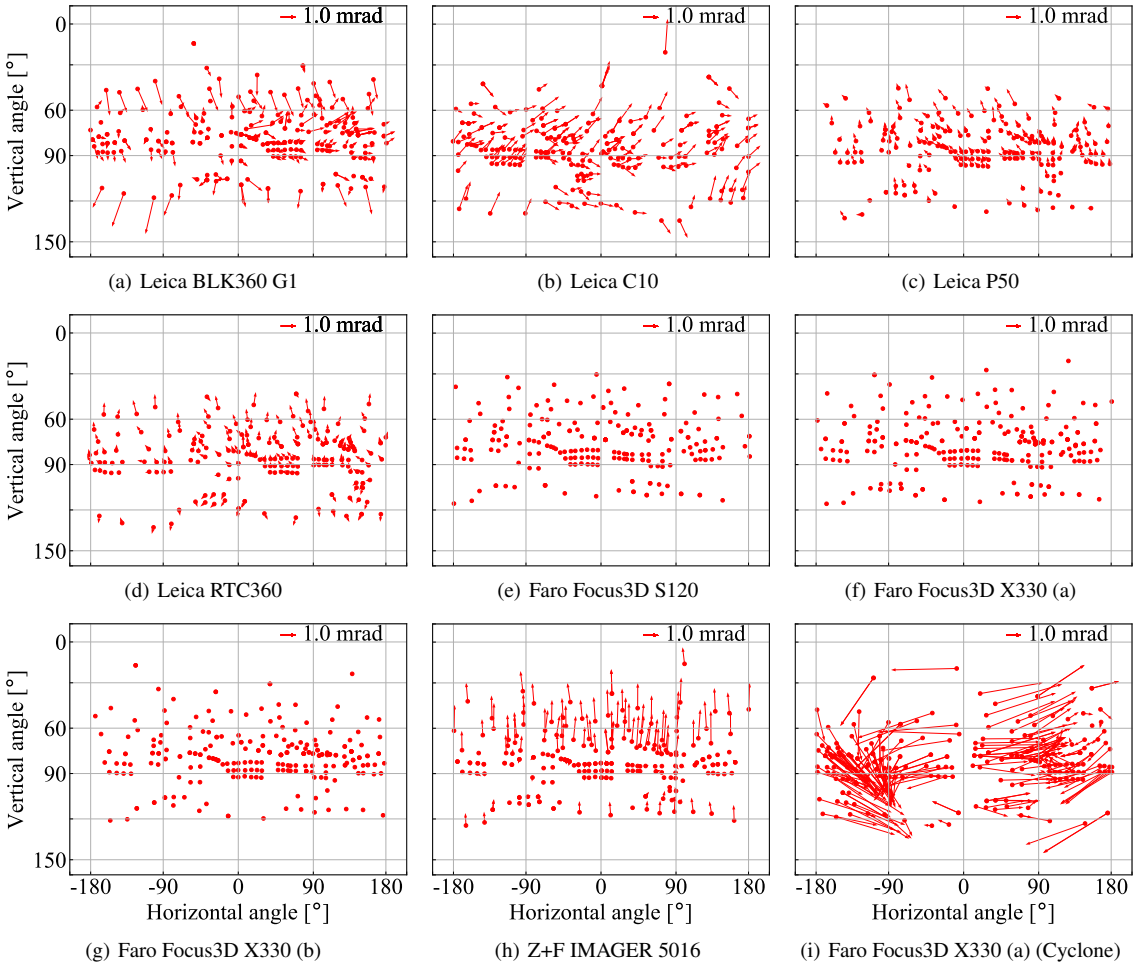
5.1.1 Scan settings The scan and image resolution are detailed in Table 1. The image resolutions are typically fixed, while the scan resolutions can be chosen by the user. For most scanners we choose scan resolutions that closely match the respective image resolution. For Leica C10, however, considering the time efficiency, we choose a scan resolution of 0.50 mrad, instead of 0.14 mrad.

Scanner	Beam div. (mrad)	Scan res. (mrad)	Image res. (mrad)
Leica BLK360 G1	0.68	0.50	0.40
Leica C10	0.14	0.50	0.15
Leica P50	0.39	0.16	0.15
Leica RTC360	0.50	0.30	0.28
Faro Focus3D S120	0.54	0.61	0.69
Faro Focus3D X330	0.54	0.61	0.69
Z+F IMAGER 5016	0.60	0.63	0.66

* According to scanner manuals, Leica BLK360 G1 and Leica RTC360 are equipped with three non-co-axial RGB cameras, while other scanners have a single co-axial RGB camera.

Table 1. Configurations of used scanners. div.: divergence, res.: resolution. Note that the beam divergence and image resolution values are based on data sheets, and the beam divergence values are converted to the full angle (Gaussian beam, $1/e^2$ points).

5.1.2 Result analysis Note that the scanners we used have not been calibrated and have not been in maintenance with the manufacturer before the experiments and thus the results do not



(a) Leica BLK360 G1 (b) Leica C10 (c) Leica P50

(d) Leica RTC360 (e) Faro Focus3D S120 (f) Faro Focus3D X330 (a)

(g) Faro Focus3D X330 (b) (h) Z+F IMAGER 5016 (i) Faro Focus3D X330 (a) (Cyclone)

Figure 5. Visualization of homogenized horizontal and vertical angle deviations on vertical-horizontal angle maps. The arrow direction represents the deviation direction. Note that the small arrow at the right-upper corner represents the scale of the arrows.

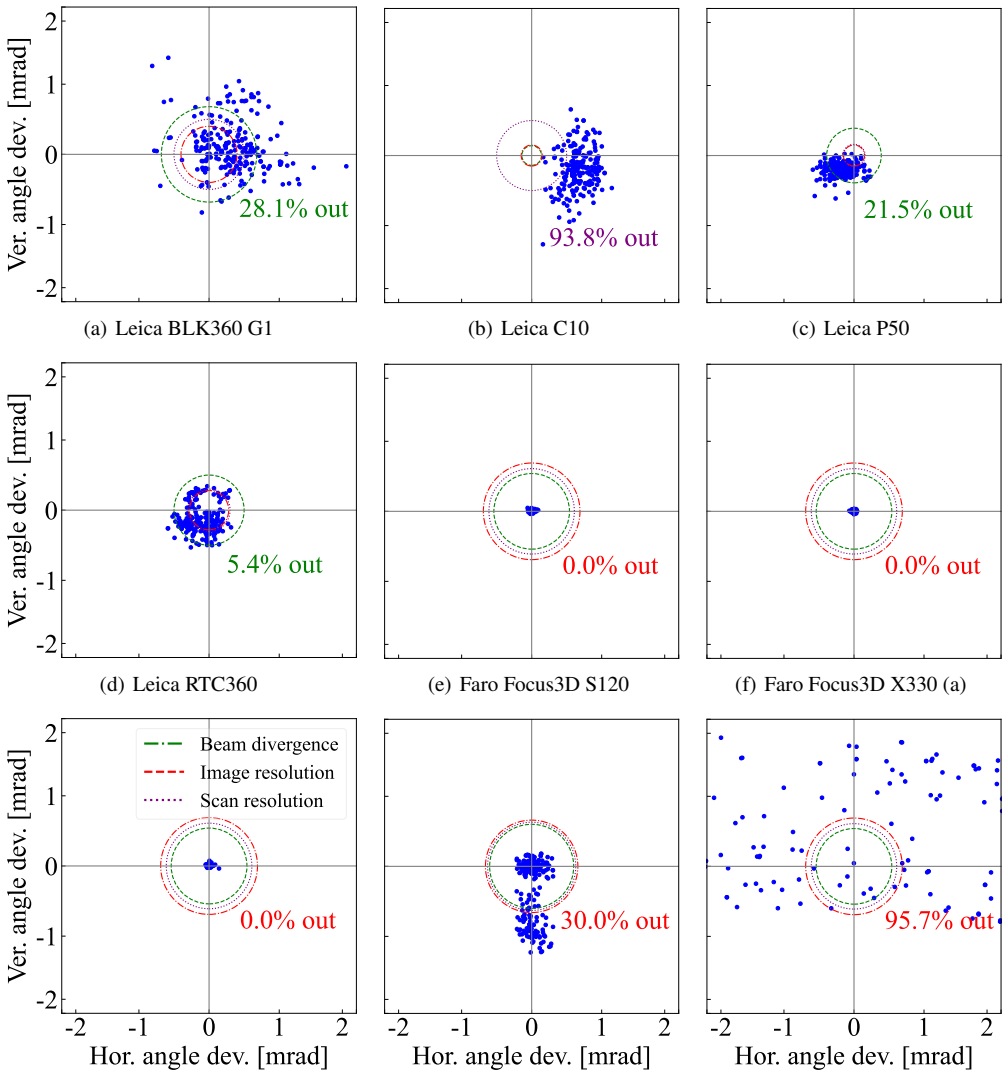
represent the quality of the scanner series in optimal working condition. Rather, our results present a snapshot of the alignment quality for the specific instruments used herein at the specific point in time.

We first visualize the homogenized horizontal and vertical angle deviations on vertical-horizontal angle maps, as described in Figure 5(a) - 5(h) (Note that Figure 5(i) and also Figure 6(i) will be explained in the last paragraph of this section). The results demonstrate that the Leica C10 exhibits deviations that primarily point towards the right direction in the horizontal angle. The Z+F IMAGER 5016 shows larger deviations predominantly oriented upwards in the vertical angle. For the Z+F IMAGER 5016, there are three distinct groups aligning with different elevation angle ranges. Specifically, almost all targets above 70° exhibit nearly constant deviation in vertical angle, almost all targets within approximately 15° from the horizon exhibit virtually no deviations, and the targets at very low elevations (around 120°) exhibit again nearly constant positive vertical deviations. The Leica BLK360 G1, Leica P50 and Leica RTC360 exhibit angle deviations of varying magnitudes, whereas the Faro Focus3D S120 and two Faro Faro Focus3D X330 indicate merely negligible angle deviations. It is worth noting that misalignment between cameras may occur in scanners with multiple RGB cameras, e.g., the Leica RTC360 and BLK360 G1. However, we leave a potential analysis of different deviations per camera for future works.

We further compare angular deviations with beam divergences, image resolutions, and scan resolutions, as illustrated in Figure 6(a) - 6(h). Notably, the Leica C10 exhibits angular deviations surpassing the maximum threshold of the beam divergence, image resolution and scan resolution in 93.8% of cases. This implies that the majority of the target centers exhibit a large misalignment between their geometry and colors, with deviations extending to at least one point/pixel spacing. In contrast, the Z+F IMAGER 5016 displays 30.0% of angular deviations exceeding the maximum threshold. All the horizontal angle deviations for the Leica C10 are positive. For the Z+F IMAGER 5016, all the larger deviations are negative vertical angle deviations. These patterns are also visible in Figure 5(b) and Figure 5(h), and suggest the presence of systematic errors of the camera to LiDAR unit alignment used for colorizing the point clouds of these scanners.

The Leica BLK360 G1 has 28.1% of angular deviations surpassing the maximum threshold. For the Leica P50, we observe that most deviations are in the negative vertical angle direction, with 21.5% of angular deviations surpassing the maximum threshold. The Leica RTC360 tested herein has only 5.4% of angular deviations surpassing our outlier threshold, and the distribution is overall spatially more random, although Figure 5(d) suggests that there may be some systematics related to the three different cameras. Conversely, the results of the Faro Focus3D S120 and two Faro Focus3D X330 exhibit only negligible angular deviations, all of which are by far smaller than the beam divergence, image and scan resolution. While this could

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-1-W1-2023-597-2023 | © Author(s) 2023. CC BY 4.0 License.



(a) Leica BLK360 G1 (b) Leica C10 (c) Leica P50

(d) Leica RTC360 (e) Faro Focus3D S120 (f) Faro Focus3D X330 (a)

(g) Faro Focus3D X330 (b) (h) Z+F IMAGER 5016 (i) Faro Focus3D X330 (a) (Cyclone)

Figure 6. Distribution of angular deviations. Angular deviations are computed according to Equation 4. The three different circles represent the value of the beam divergence, image resolution, and scan resolution. For each individual scanner, the percentage (e.g., 0.0% out in Figure 6(g)) indicates the percentage of angular deviations that exceed the maximum threshold of the beam divergence, image resolution, and scan resolution for each individual scanner.

be related to excellent RGB camera to LiDAR unit calibration and very stable relative orientation of the two, the results (also of the temperature tests, see Section 5.2) and a patent search as well as the analysis of Faro data exported using a different software (see next paragraph) strongly suggest that the images and point clouds are well aligned numerically in the point cloud colorization process within the Faro SCENE software.

In addition, we notice that Leica software (e.g., Cyclone REGISTER 360) allows users to import raw scan projects from scanners of other manufacturers. Thus, we utilize the Cyclone REGISTER 360 to colorize the point cloud obtained from the Faro Focus3D X330 (a), and present the results in Figure 5(i) and Figure 6(i). Notably, the colored point cloud colorized using the Cyclone REGISTER 360 exhibits significant deviations, particularly in the horizontal angle direction, when compared to the result colorized using the software (i.e., Faro SCENE) provided by its own manufacturer as shown in Figure 5(f). Furthermore, Figure 6(i) shows that 95.7% of angular deviations exceed the maximum threshold, whereas the result in Figure 6(f) indicates that no angular deviation exceeds the maximum threshold. This comparison indicates (1) that the assessment of the alignment quality refers to the system of scanner and all software used to output or produce the colorized point clouds, not just the scanner itself, and (2) that the Faro SCENE either uses numerical parameters for the alignment which are not used by or not accessible to the Cyclone REGISTER 360, or the Faro SCENE includes a data driven alignment of RGB images and point clouds.

5.2 Experiment B: alignment stability evaluation

We conduct an additional analysis to evaluate how stable TLS scanner alignment is under varying ambient temperatures. The experiment is carried out in a climate chamber with dimensions of 4.0 x 3.0 x 2.2 m^3. We raise the temperature incrementally from 5 °C to 40 °C, and do scanning at each 5 °C interval. In this experiment, we use four scanners, including Leica BLK360 G1, Leica RTC360, Faro Focus3D X330 (a), and Z+F IMAGER 5016, and ten artificial targets, including four rotated checkerboard targets on tripods and six checkerboard targets on the chamber surface.

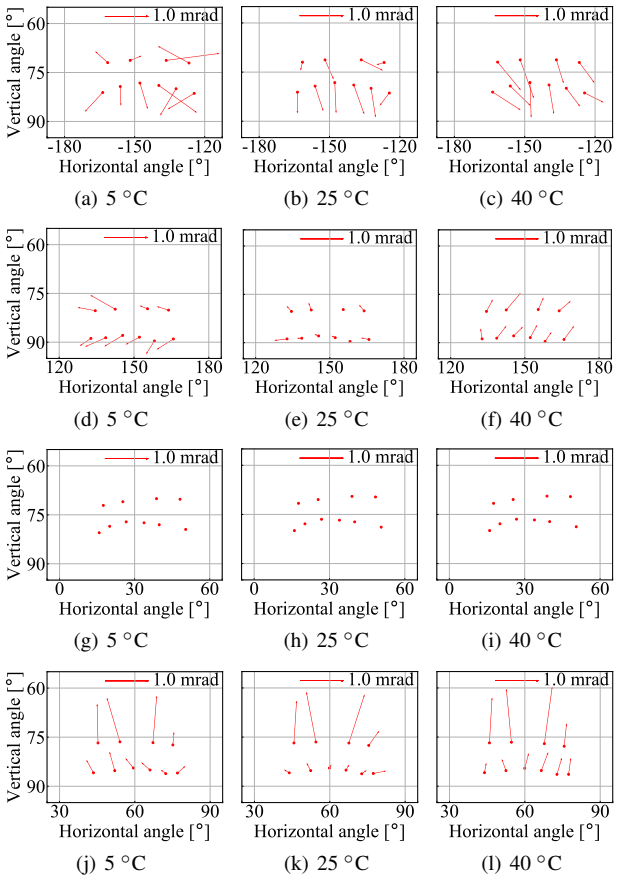


Figure 7. Visualization of homogenized horizontal and vertical angle deviations of four scanners on vertical-horizontal angle maps at three different ambient temperatures (i.e., 5, 25, and 40 °C). First row to last row: Leica BLK360 G1, Leica RTC360, Faro Focus3D X330 (a), and Z+F IMAGER 5016.

In Figure 7, we observe that the Leica BLK360 G1 exhibits quite similar magnitudes of deviations at all three temperatures, but the deviation directions change considerably between 5 and 40 °C. The Leica RTC360 shows relatively small deviations at 25 °C, while the deviations at 5 and 40 °C exhibit relatively large changes. In contrast, the Faro Focus3D X330 (a) remains stable with no distinct deviation detected at different temperatures. The Z+F IMAGER 5016 shows both larger and smaller deviations at 25 °C, with its larger deviations becoming steeper and smaller deviations becoming larger at both 5 and 40 °C.

For scanners that exhibit significant changes in either magnitudes or directions of deviations with varying temperatures, we infer that the variable temperatures affect the alignment stability between these scanners and their built-in RGB cameras. Conversely, scanners that demonstrate a stable alignment across varying temperatures may achieve this due to their stable design or due to effective data driven alignment during the colorization process within the scanner or in a separate software.

6. CONCLUSIONS

In this study, we present an artificial target detection and estimation procedure for quantifying the quality of the alignment between the TLS point cloud geometry and colors. The code used herein for the entire analysis from the colorized point cloud to the plots and numeric output of the angular deviations will be provided on GitHub. Our experiments verify the efficiency and effectiveness of the proposed procedure. More importantly, our experimental results provide valuable insights into both the quality and the stability of the alignment across multiple commercial TLS scanners. The primary observations from the experiments are as follows:

- The alignment of RGB colors and point cloud geometry depends on the scanner and on the software used for the colorization of the point cloud. Using different software for colorizing the point cloud can lead to vastly different alignment quality.
- While we found only negligibly small misalignments of RGB colors and point cloud geometry for all three Faro scanners used in combination with the Faro SCENE software in our experiments, some point clouds produced by other scanners and software exhibited misalignments exceeding the beam divergence, image and scan resolution.
- The maximum deviations were about 2 mrad perpendicular to the line-of-sight when colorizing the point clouds with the respective manufacturer's software or scanner inbuilt functions, while they were up to about 5 mrad when using a different software.
- For one scanner almost all deviations (approximately 94%) attributed to misalignment of the RGB colors exceeded the beam divergence, image and scan resolution. For several of the scanners used herein, changing ambient temperature (tested at 5, 25 and 40°C) resulted in significant changes of the misalignments.

Given the widespread use of point cloud geometry and colors from RGB-D cameras, we are convinced that exploiting color information in addition to geometry and LiDAR intensity can yield valuable benefits beyond the visualization purpose also for TLS point clouds. However, our findings suggest that there may be significant misalignments and these may have to be addressed when utilizing the colored point clouds for high-accuracy applications (e.g., millimeter-level deformation monitoring). To tackle these issues, one can either (i) use scanners and software assuring sufficiently accurate alignment—e.g., identified using the approach presented herein, (ii) adopt additional alignment methods, such as utilizing an iterative optimization method (Rotstein et al., 2022), or (iii) apply a calibration approach between TLS and an external RGB camera (Omidalizarandi et al., 2019a). For scanners that exhibit significant alignment instability due to ambient temperature changes, only data-driven optimization approaches may be applicable unless it is possible to ensure the temperature remains stable during scanning. Our proposed procedure can be employed to assess the current alignment status and even evaluate the effectiveness of adopted alignment methods.

Our future work will address the possibility of integrating RGB colors with point cloud geometry to establish more reliable correspondences. Potentially, we can utilize these correspondences for point cloud registration and deformation analysis in geodetic monitoring.

ACKNOWLEDGEMENTS

The first author was supported by the China Scholarship Council (CSC). The dataset used for training target detectors was annotated by Jon Allemand. The authors also thank Robert Presl and Thomas Posur for their assistance in organizing the experimental sites.

References

Aiello, E., Valsesia, D., Magli, E., 2022. Cross-modal learning for image-guided point cloud shape completion. *Advances in Neural Information Processing Systems*. https://openreview.net/forum?id=bi1BTcXa8Q.

Balado, J., González, E., Rodríguez-Somoza, J. L., Arias, P., 2023. Multi feature-rich synthetic colour to improve human visual perception of point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196, 514–527. doi.org/10.1016/j.isprsjprs.2023.01.019.

Basso, F., Menegatti, E., Pretto, A., 2018. Robust intrinsic and extrinsic calibration of RGB-D cameras. *IEEE Transactions on Robotics*, 34(5), 1315–1332. doi.org/10.1109/TRO.2018.2853742.

Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Xu, J., et al., 2019. MMDetection: Open MMLab Detection Toolbox and Benchmark. *arXiv preprint arXiv:1906.07155*. doi.org/10.48550/arXiv.1906.07155.

Feng, M., Hu, S., Ang, M. H., Lee, G. H., 2019. 2D3D-MatchNet: Learning to match keypoints across 2D image and 3D point cloud. *Proceedings of the IEEE International Conference on Robotics and Automation*, 4790–4796. doi.org/10.1109/ICRA.2019.8794415.

Fischler, M. A., Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381–395. doi.org/10.1145/358669.358692.

Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J., 2021. YOLOX: Exceeding YOLO series in 2021. *arXiv preprint arXiv:2107.08430*. doi.org/10.48550/arXiv.2107.08430.

Guizar-Sicairos, M., Thurman, S. T., Fienup, J. R., 2008. Efficient subpixel image registration algorithms. *Optics letters*, 33(2), 156–158. doi.org/10.1364/OL.33.000156.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN. *Proceedings of the IEEE international conference on computer vision*, 2980–2988. doi.org/10.1109/ICCV.2017.322.

Janßen, J., Medic, T., Kuhlmann, H., Holst, C., 2019. Decreasing the uncertainty of the target center estimation at terrestrial laser scanning by choosing the best algorithm and by improving the target design. *Remote Sensing*, 11(7), 845. doi.org/10.3390/rs11070845.

Julin, A., Kurkela, M., Rantanen, T., Virtanen, J.-P., Maksimainen, M., Kukko, A., Kaartinen, H., Vaaja, M. T., Hyyppä, J., Hyyppä, H., 2020. Evaluating the quality of TLS point cloud colorization. *Remote Sensing*, 12(17), 2748. doi.org/10.3390/rs12172748.

Kweon, H., Yoon, K.-J., 2022. Joint learning of 2D-3D weakly supervised semantic segmentation. *Advances in Neural Information Processing Systems*. https://openreview.net/forum?id=4Q9CmC3ypdE.

Liang, M., Yang, B., Wang, S., Urtasun, R., 2018. Deep continuous fusion for multi-sensor 3d object detection. *Proceedings of the European conference on computer vision*, 641–656. doi.org/10.1007/978-3-030-01270-$0_3$9.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 936–944. doi.org/10.1109/CVPR.2017.106.

Omidalizarandi, M., Kargoll, B., Paffenholz, J.-A., Neumann, I., 2019a. Robust external calibration of terrestrial laser scanner and digital camera for structural monitoring. *Journal of Applied Geodesy*, 13(2), 105–134. doi.org/10.1515/jag-2018-0038.

Omidalizarandi, M., Paffenholz, J.-A., Neumann, I., 2019b. Automatic and accurate passive target centroid detection for applications in engineering geodesy. *Survey Review*, 51(367), 318–333. doi.org/10.1080/00396265.2018.1456001.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A., 2017. Automatic differentiation in PyTorch. https://openreview.net/forum?id=BJJsrmfCZ.

Pham, Q.-H., Uy, M. A., Hua, B.-S., Nguyen, D. T., Roig, G., Yeung, S.-K., 2020. LCD: Learned cross-domain descriptors for 2D-3D matching. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34, 11856–11864. doi.org/10.1609/AAAI.V34I07.6859.

Pleskacz, M., Rzonca, A., 2016. Design of a testing method to assess the correctness of a point cloud colorization algorithm. *Archiwum Fotogrametrii, Kartografii i Teledetekcji*, 28. doi.org/10.14681/afkit.2016.007.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788. doi.org/10.1109/CVPR.2016.91.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 91-99. doi.org/10.5555/2969239.2969250.

Rotstein, N., Bracha, A., Kimmel, R., 2022. Multimodal colored point cloud to image alignment. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6656–6666. doi.org/10.1109/CVPR52688.2022.00654.

Steffey, K., Becker, R., Becker, B.-D., Gittinger, J., 2018. Laser scanner. U.S. Patent 10 132 611 B2.

Vollrath, J. E., Ossig, M., 2018. Balancing colors in a scanned three-dimensional image. U.S. Patent 10 116 920 B2.

Wang, C.-Y., Bochkovskiy, A., Liao, H.-Y. M., 2022. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*. doi.org/10.48550/arXiv.2207.02696.

Zaidi, S. S. A., Ansari, M. S., Aslam, A., Kanwal, N., Asghar, M., Lee, B., 2022. A survey of modern deep learning based object detection models. *Digital Signal Processing*, 126, 103514. doi.org/10.1016/j.dsp.2022.103514.

Zhang, Y., Yu, J., Huang, X., Zhou, W., Hou, J., 2022. PCR-CG: Point cloud registration via deep explicit color and geometry. *Proceedings of the European conference on computer vision*, 443–459. doi.org/10.1007/978-3-031-20080-$9_2$6.