*Reports on Geodesy and Geoinformatics*, 2025, Vol. 119, pp. 39–61

DOI: 10.2478/rgg-2025-0006
Received: 8 November 2024 / Accepted: 29 January 2025
Published online: 14 March 2025

ORIGINAL ARTICLE

# The comparison of distance metrics in descriptor matching methods utilised in TLS-SfM point cloud registration

Jakub Markiewicz [1*]

[1]Department of Photogrammetry, Remote Sensing and Spatial Information Systems, Faculty of Geodesy and Cartography, Warsaw University of Technology, 1 Politechniki Square, 00-661 Warsaw, Poland

*jakub.markiewicz@pw.edu.pl

## Abstract

Advanced measurement techniques, such as Terrestrial Laser Scanning (TLS), play a vital role in documenting cultural heritage and civil engineering structures. A key aspect of these applications is the accurate registration of point clouds. Conventional TLS methods often rely on manual or semi-automated correspondence detection, which can be inefficient for large or complex objects. Structure-from-Motion Terrestrial Laser Scanning (SfM-TLS) offers an alternative methodology, comprising two primary phases: correspondence search and incremental reconstruction. Descriptor matching in SfM-TLS typically employs the $L_2$ norm to measure Euclidean distances between features, valued for its simplicity and compatibility with algorithms like SIFT. This study investigates the influence of various distance metrics on descriptor matching during the correspondence search stage of SfM-TLS. Eight metrics were analysed: Bray-Curtis, Canberra, Correlation, Cosine, $L_1$, $L_2$, Squared Euclidean, and Standardised Euclidean. Synthetic data experiments highlighted challenges in keypoint detection and matching due to measurement angles, material characteristics, and 3D-to-2D transformations. Simulations incorporating Gaussian noise demonstrated that image rotation and skew significantly affected tie-point accuracy, more so than variations in intensity. In field applications involving cultural heritage sites and building interiors, the $L_1$ and Squared Euclidean metrics yielded higher accuracy, while the Canberra metric underperformed. Metric selection was found to have a greater impact on complex geometries, such as historical structures, compared to simpler forms. Consequently, this study recommends the $L_1$ and Squared Euclidean metrics for pairwise SfM-TLS registration, as they exhibit robustness in maintaining high accuracy and completeness across a variety of architectural scenarios.

**Key words**: distance metrics, descriptor matching, pairwise TLS registration, cultural heritage, public utilities

## 1 Introduction

Nowadays, one of the most important measurement techniques used in the inventory and measurement of architecture (Abbate et al., 2019; Arif and Essa, 2017; Giżyńska et al., 2022; Kuzyk, 2023), civil engineering, or industrial objects and sites (Kowalska and Kowalczyk, 2024; Mukupa et al., 2016; Rashidi et al., 2020) is Terrestrial Laser Scanning (TLS). It is widely applied in generating measurement documentation, primarily due to its main advantages, which include, among others, accuracy in data acquisition, data density, automation of point measurement, and non-destructive way of data acquisition (Grussenmeyer et al., 2012; Remondino and Stylianidis, 2016; Tobiasz et al., 2019). The process of generating measurement documentation based on TLS data is a multi-stage process, which includes: (1) planning scanner positions concerning the measured object, (2) data acquisition, (3) point cloud registration, and (4) generating the final documentation in the form of 3D models, orthoimages, or vector drawings (Berenyi et al., 2010; Cipriani et al., 2019; Markiewicz et al., 2015; Mukupa et al., 2016; Piermattei et al., 2019; Tobiasz et al., 2019). However, one of the critical stages determining the accuracy of the final measurement documentation is the point cloud registration process (Cheng et al., 2018).


Publisher: De Gruyter

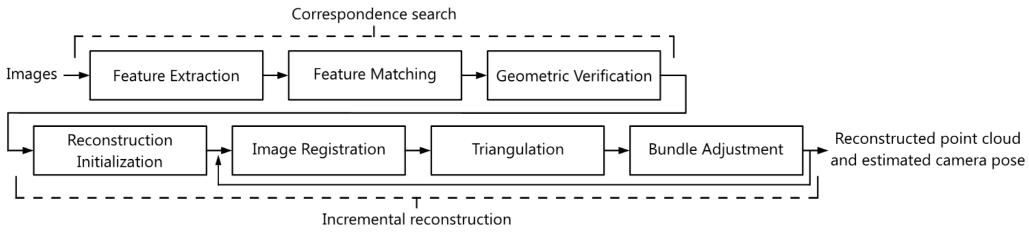



**Figure 1.** The Incremental SfM procedure (Bianco et al., 2018)

In general, the registration method is based on determining the mathematical relationship between the registered point cloud and the reference system corresponding to it. The number of different TLS point cloud registration methods can be found in the literature (Dong et al., 2018; Kowalska and Kowalczyk, 2024; Rashidi et al., 2020) but is generally divided into two groups, i.e. pairwise and multi-view registration. Recent research focuses on methods for automatic detection and matching feature points detected on point clouds, mainly based on the modified Structure-from-Motion (SfM) approaches (Alba et al., 2012; Han et al., 2018; Janßen et al., 2023; Kang et al., 2009; Markiewicz, 2024; Markiewicz et al., 2023; Moussa et al., 2012; Urban and Weinmann, 2015). The SfM method allows the determination of the relative orientation parameters between an unordered group of images and three-dimensional shape reconstruction based on these images. The SfM approach is a complex computational process consisting of multiple stages, which can be generally divided into two main components: (1) correspondence search and (2) incremental reconstruction (Figure 1).

The main idea of the *Corresponding search* part is to find robust and stable tie points according to the Tuytelaars and Mikolajczyk theory (Tuytelaars and Mikolajczyk, 2007). In the *Feature Extraction* phase, characteristic features, also named keypoints, are detected separately for each processing image, and the local characteristic of the image intensity gradients around each keypoint is described. *Feature matching* refers to finding corresponding features from two similar images based on similarity between descriptors. It should be emphasised that the feature matching accuracy depends on image similarity, complexity, and quality. The outcome of this step is a collection of images with a minimum of pairwise overlap, along with a corresponding set of feature matches. At this stage, points are identified and subjected to *Geometric Verification* and further processing in the SfM method. *Geometric Verification* is needed to eliminate outliers and improve the quality of tie points and, consequently, the quality of the final image orientation. Additionally, the relative image orientation parameters (using the homography method) and the 3D coordinates of tie points are determined and utilised in the final step – the incremental reconstruction. *Incremental reconstruction* allows the determination of image orientation parameters with camera calibration parameters (Karwel and Markiewicz, 2022). The *Reconstruction Initialization* is the crucial part because it leads to 3D model quality and final accuracy of data orientation. This process begins by selecting a pair of geometrically verified images with the densest matches (the highest number of tie points). These images provide the initial camera poses and common points, which serve as the foundation for the reconstruction. *Image registration* is used to calculate the pose (position and orientation) of newly added images using 2D-3D correspondences and solves the Perspective-n-Point (PnP) problem, often robustly optimized with RANSAC or its variants. *Triangulation* determines the 3D coordinates of additional points by leveraging epipolar constraints and solving reprojection errors, adding density to the point cloud. Finally, *Bundle Adjustment (BA)* refines both camera parameters and 3D points using the Levenberg-Marquardt algorithm, reducing accumulated errors.

The feature-based methods used for point cloud alignment from terrestrial laser scanning are based on a modified SfM method known in the literature as TLS-SfM. The main differences in the data processing workflow are related to the type of input data and the Geometric Verification stage. Since 2D detectors search for points on images, it is necessary to convert point clouds into the raster form. This is typically done by applying cartographic projections, for example, to create spherical images, where grayscale values of pixels are interpolated based on the intensity of laser beam reflection or the colour assigned to the point cloud. A depth map or X, Y, and Z coordinate maps are assigned for these spherical images, enabling the calculation of corresponding 3D coordinates in the point cloud based on detected 2D coordinates in the image. Another difference compared to the classical SfM method is the choice of the relative orientation model used during the geometrical verification stage. This method uses the 2D coordinates of point pairs detected during the descriptor-matching stage and the homography model. For the TLS-SfM method, a 6-parameter 3D transformation and the 3D coordinates of point pairs detected during the descriptor matching stage are used. Articles Markiewicz (2024); Markiewicz et al. (2023) detailed the TLS-SfM method and the individual data processing stages.

This article is a continuation of previous work (Markiewicz, 2024; Markiewicz et al., 2023), in which the TLS-SfM method was presented, and studies were conducted on the impact of detector selection on the accuracy and completeness of TLS point cloud registration. Most studies related to TLS point cloud registration based on a modified SfM method typically utilised the $L_2$ norm (Alba et al., 2012; Markiewicz, 2024; Moussa et al., 2012; Urban and Weinmann, 2015) or metrics similar to the $L_2$ norm during the descriptor-matching stage (Janßen et al., 2023). The choice of this norm is associated with the fundamental algorithm for keypoint detection and matching, which is based on the SIFT algorithm (Lowe, 2004). In this article, different strategies for similarity computation (Barycurtis, Canberra, Correlation, Coine, $L_1$, $L_2$, Seuclidean, Sqeuclidean) in the descriptor matching were presented, as well as assets' influence on the final registration accuracy.

As Test Sites, point clouds of interiors of buildings with historical surfaces of a decorative structure and interiors of public utilities (an office and an empty shop in a shopping mall) were chosen. The commonly used target-based registration method was compared with the proposed method.

This paper is divided into five main sections. Section 2 describes the descriptor matching methods. Section 3 describes the test sites, approach, and data analysis method. In Section 4, the results of the descriptor-matching assessments are summarised. In the conclusion (Section 5), future work is proposed, and the possibilities and limitations of matching approaches are summarised.

## 2 Feature matching

Feature matching, or generally image matching, is fundamental in many computer vision applications, namely image orientation, camera calibration, object recognition or tracking. The main idea of this approach is to establish correspondence between two images or features of the same scene. A common approach to feature matching (a part of the SfM's corresponding search phase) consists of a set of features, also named keypoints and the assignment of the local characteristic of the image intensity (description part). The feature-extraction part is performed on each image separately and based on the algorithms and methods which detect features invariant to image translation, scaling, and rotation, partially invariant to illumination changes, and robust to local geometric distortion such as SIFT (Bay et al., 2006; Harris and Stephens, 1988; Moussa et al., 2012; Tuytelaars and Mikolajczyk, 2007). Each detected feature is analysed for gradient change based on its nearest neighbour to assign unique features. In literature, many descriptors exist, such as SIFT, SURF or DAISY (Bay et al., 2006; Lowe, 2004; Tola et al., 2010). The SIFT (used in this investigation) descriptor's main idea is to calculate local image gradients at a selected scale around the region's key point under study. The descriptor's work is based on analysing histograms of $4\times4$ pixel neighbourhood orientations with 8 bins each. The histograms are derived from magnitudes and orientations sampled in a $16\times16$ region around the keypoint so that for each histogram, a $4\times4$ subregion of the original neighbourhood region is sampled. The magnitude and orientations of the image gradient are probed around the location of the keypoint, using the scale of the keypoint to select the image. To obtain orientation invariance, the descriptor coordinates and gradient orientations are rotated relative to the keypoint orientation (Karwel and Markiewicz, 2022).

Once the feature vector is obtained, the next stage of determining tie points on image pairs is the relative matching of keypoints. This is typically performed using a similarity measure for feature vectors. One of the simplest methods is based on the Brute-Force Matching algorithm. In this approach, a descriptor from one feature in the first set is compared to all features in the second set using distance calculations. The closest feature is then returned. This expansive solution guarantees getting the solution, but it does not guarantee that it will be optimal. Another more sophisticated approach is a FLAN-based matcher (Fast Library for Approximate Nearest Neighbours) that utilises a k-dimensional tree, a space-partitioning data structure used in computer science to organise points in a k-dimensional space. This method is based on the t-nearest neighbour search in large datasets and for high-dimensional features (OpenCV, 2018).

In data science, especially in Machine Learning, the similarity measure determines how data samples are related or close to each other. The similarity measure is usually expressed as a numerical value that allows it to assess whether it is correlated. Generally, for similarity function value analysis, it can be assumed that larger values indicate more significant similarity, while in distance functions, smaller values indicate more significant similarity (Aggarwal et al., 2015). Choosing a distance metric significantly influences the quality and correctness of description matching and tie point quality. Therefore, selecting the correct metric distance affects the final quality of image matching and TLS point cloud registration.

### 2.1 $L_n$ norm – Minkowski distance

One of the most commonly used distance methods for quantitative data matching is the $L_n$ norm between two vectors of data, respectively. $\overline{X} = (x_1, x_2, \ldots, x_s)$ and $\overline{Y} = (y_1, y_2, \ldots, y_s)$, which is determined by the following equation:

$$\text{Distance}\left(\overline{X},\overline{Y}\right) = \left(\sum_{i=1}^{s} |x_i - y_i|^p\right)^{\frac{1}{p}}. \tag{1}$$

One commonly used $L_n$ norm is $L_2$ also called the Euclidean distance method. This particular case derives its intuition from spatial application, where it has a physical interpretation. The main property of the Euclidean distance is the invitation of rotation, which is crucial, especially in the description matching case. The Manhattan distance, also known as Taxicab, Block Distance and $L_1$ norm, calculates the distance between two real-valued vectors and the sum of the absolute differences between two vectors.

The $L_1$ norm distance offers advantages such as robustness against outliers, encouragement of sparse solutions, geometric interpretability, and support for feature selection. However, it also comes with drawbacks, including a lack of smoothness, potential multiple solutions in the presence of correlated data, sensitivity to scaling, inefficiency for non-sparse data, and limited insight into relationships among non-zero coefficients. The $L_2$ norm distance has several advantages, including being differentiable at all points, making optimisation smoother; showing less sensitivity to outliers compared to the $L_1$ norm; yielding a unique solution in most cases, even with correlated data; being more efficient for non-sparse data; and capturing relationships between non-zero coefficients. However, it also has disadvantages, such as potentially weaker promotion of sparsity compared to the $L_1$ norm, sensitivity to feature scaling, a less intuitive geometric interpretation, and potential performance issues with datasets containing outliers.

In some cases, when the weight of some features is more important than others, it is possible to apply the weight of the features differently if domain-specific knowledge about the relative importance of different features is available. The generalised Minkowski distance (Eq. 1) is extended with the weight:

$$\text{Distance}\left(\overline{X},\overline{Y}\right) = \left(\sum_{i=1}^{s} a_i \cdot |x_i - y_i|^p\right)^{\frac{1}{p}}. \tag{2}$$

### 2.2 Normalized $L_1$ and $L_2$ norms

The normalised versions of the $L_1$ and $L_2$ norms are the fundamental metrics in the least square problems, linear algebra and Machine Learning applications. Applying normalised $L_1$ norm is also called Mean-Squared Error (MSE) (Eq. 3). The MSE is sensitive to the large outliers and allows for assets the quality of matching.

$$\text{Distance}\left(\overline{X},\overline{Y}\right) = \frac{1}{s}\sum_{i=1}^{s} |x_i - y_i|^2 \tag{3}$$

Another weighted method applied for $L_1$ is Bray-Cutris (also known as Braycurtis) distance (Eq. 4). It is often used for data scattered around an origin, as it is biased for measures around the origin and very sensitive for values close to zero (math.net, 2024). Compared to the $L_1$ norm, it is more robust regarding outlier influence.

$$\text{Distance}\left(\overline{X},\overline{Y}\right) = \sum_{i=1}^{s} \frac{|x_i - y_i|}{|x_i + y_i|} \tag{4}$$

The Braycurtis distance measure is advantageous for capturing relative abundances and handling sparse data effectively, but it can be sensitive to dominant species' influence and scaling differences.

The standardised Euclidean distance, also known as the "Seuclidean" distance, can measure the dissimilarity between data points while accounting for feature scaling:

$$\text{Distance}\left(\overline{X},\overline{Y}\right) = \sqrt{\frac{1}{s}\sum_{i=1}^{s} |x_i - y_i|^2}, \tag{5}$$

where $s$ is a 1-D array of component variances, it is usually computed among a larger collection of vectors.

This distance metric is advantageous as it normalises the data by dividing the squared differences between coordinates by the variances of each dimension. This normalisation process allows fairer comparisons among features with different scales, preventing features with larger ranges from dominating the distance calculation.

However, while addressing scaling issues, the seuclidean distance may still be sensitive to outliers or skewed distributions. It is advantageous when working with data where feature scales vary widely, helping to provide a more accurate representation of dissimilarity while considering the characteristics of individual features.

The squared Euclidean distance, often abbreviated as "sqeuclidean" distance, is a distance metric used to quantify the dissimilarity between two points in a multi-dimensional space. It is calculated by taking the sum of the squared differences between the corresponding coordinates of the two points:

$$\text{Distance}\left(\overline{X},\overline{Y}\right) = s\sum_{i=1}^{s}|x_i - y_i|^2, \tag{6}$$

where $s$ is the weight for each value in $u$ and $v$. Default is None, which gives each value a weight of 1.0.

This distance metric is advantageous for various applications due to its simplicity and computational efficiency. However, it can be sensitive to differences in the magnitudes of features, potentially leading to biased results when dealing with data with varying scales. Despite this limitation, the squared Euclidean distance remains a popular choice in various fields for its ease of calculation and ability to capture differences between points based on their coordinates.

### 2.3 Canberra distance

The Canberra distance is a quantitative measure to gauge dissimilarity between two sets of numerical attributes. It calculates divergence by summing the absolute differences between corresponding attributes in both sets and then normalising by summating their absolute magnitudes:

$$\text{Distance}\left(\overline{X},\overline{Y}\right) = \sum_{i=1}^{s}\frac{|x_i - y_i|}{|x_i| + |y_i|}. \tag{7}$$

This property makes it particularly suitable for data with varying scales. An advantage lies in its ability to capture both attribute magnitude and direction. Yet, this sensitivity can lead to undue influence from attributes with higher magnitudes, potentially distorting outcomes.

### 2.4 Cosine distance

The cosine similarity is defined as a cosine of angles between vectors of data, respectively $\overline{X} = (x_1, x_2, \ldots, x_s)$ and $\overline{Y} = (y_1, y_2, \ldots, y_s)$, which is the dot product of the vector divided by its length:

$$\begin{aligned}\text{Cosine similarity}\left(\overline{X},\overline{Y}\right) &= \cos(\theta) = \frac{dot(\overline{X},\overline{Y})}{\left\|\overline{X}\right\|\left\|\overline{Y}\right\|}\\ &= \frac{\sum_{i=1}^{s} x_i * y_i}{\sqrt{\sum_{i=1}^{s} x_i^2}\sqrt{\sum_{i=1}^{s} yx_i^2}}\\ \text{Distance}\left(\overline{X},\overline{Y}\right) &= 1 - \text{Cosine similarity}\left(\overline{X},\overline{Y}\right)\end{aligned} \tag{8}$$

The cosine similarity value is always ranges between [-1, 1]. If the value is equal to 1, two analysed vectors are similar; if the value is equal to 0, those vectors are orthogonal, and if it is -1, the vectors are negative. However, cosine similarity is mainly used in positive spaces, where outcomes are between [0,1]. Cosine distance (and similarity) is generally used as a metric for measuring distance when the magnitude of the vector does not play a key role. The main advantage of the cosine similarity is the low complexity, especially for sparse vectors.

### 2.5 Pearson Correlation distance

Another commonly used distance metric is based on the correlation coefficient. In principle, it allows determination of the strength of the relationship between two sets of numerical attributes. The covariance value is used to calculate the value of this distance metric:

$$\begin{aligned}\text{Correlation similarity}\left(\overline{X},\overline{Y}\right) &= \frac{\text{Covariance}(\overline{X},\overline{Y})}{\sqrt{\text{Variance}(\overline{X})}\sqrt{\text{Variance}(\overline{Y})}}\\ &= \frac{\sum_{i=1}^{s}(x_i - \frac{1}{s}\sum_{i=1}^{s}x_i)(y_i - \frac{1}{s}\sum_{i=1}^{s}y_i)}{\sqrt{\sum_{i=1}^{s}(x_i^2 - \frac{1}{s}\sum_{i=1}^{s}x_i)}\sqrt{\sum_{i=1}^{s}(y_i^2 - \frac{1}{s}\sum_{i=1}^{s}y_i)}}\\ \text{Distance}\left(\overline{X},\overline{Y}\right) &= 1 - \text{Cosine similarity}\left(\overline{X},\overline{Y}\right)\end{aligned} \tag{9}$$

The correlation similarity value always is between [-1, 1]. Similarly to the cosine similarity, it is assumed that values close to 1 indicate that the vectors are similar, values of 0 indicate no similarity, and values of -1 represent negative correlation. For this reason (as in the case of cosine similarity), the similarity values from the range [0,1] are used. The main advantages of cosine similarity are the straightforward interpretation of results, ease of calculation, and the ability to indicate whether there is a relationship between vectors and assess the quality of that relationship. Despite these advantages, the primary limitation of this similarity measure is its sensitivity to outliers.

## 3 Materials and methods

### 3.1 The overview of the approach

This research examined the impact of selecting distance metrics during the feature-matching stage for automatic TLS point cloud registration. For this purpose, the TLS-SfM method (Markiewicz et al., 2023), a multi-stage solution applied for point cloud registration, was used. It is based on the original software that utilised the OpenCV (OpenCV, 2018), NumPy (NumPy, 2024), SciPy (SciPy, 2024) libraries, and the Author's algorithms and methods. It consists of (1) data conversion from RAW point clouds into the raster form in the spherical projection with depth map (Markiewicz et al., 2023), (2) detection features by SIFT algorithm (Lowe, 2004), (3) descript detected keypoints by SIFT descriptor (Lowe, 2004), descriptor matching with Brute-Force Matching with the following distance measures: Barycurtis, Canberra, Correlation, Coine, $L_1$, $L_2$, Seuclidean, Sqeuclidean, (4) TLS pair co-registration with geometrical verification, (5) multiple-pair matching and (6) bundle adjustment (Figure 2).

The simulated data (Section 4.1) and real-world data (Section 4.2) were used to evaluate the selection of distance metrics during the feature-matching stage. To perform a complete analysis of the applying the strategy for feature matching, the following parameters were assets:

1. Data registration's completeness determines the distance metrics' robustness in the feature matching step. It is understood as the ability to register all pair scans with each other and determine the robustness and effectiveness of using specific metric distance in feature matching.
2. The number of correctly matched keypoints for a pair of TLS point clouds directly impacts the robustness, accuracy, and completeness of the TLS point cloud registration process (TP). At the same time, it also defines the effectiveness of distance metrics during the feature registration stage.
3. The number of incorrectly matched keypoints for a pair of point clouds that have correspondence to the second dataset (FP).
4. The number of correctly defined keypoints for which no matching points were found in the second dataset (TN).
5. The number of keypoints for which the correspondence on

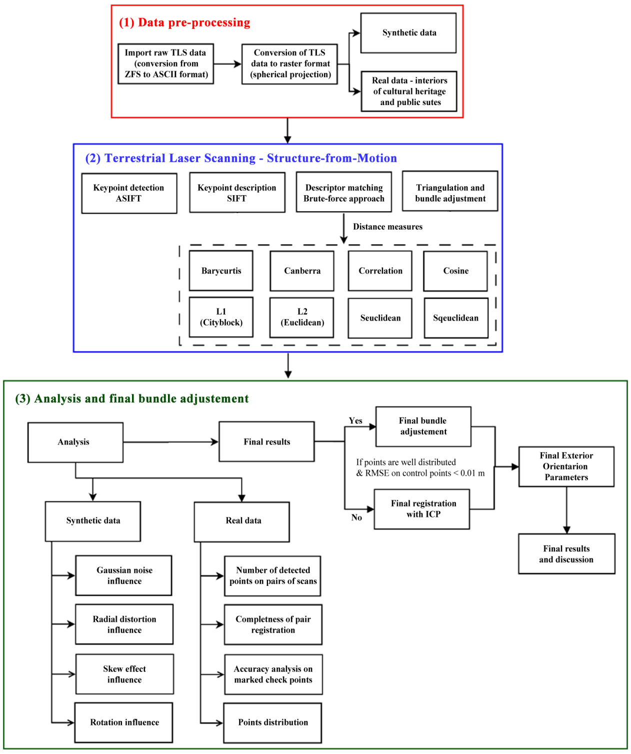



**Figure 2.** The diagram of the performed research

the second point cloud was found, but this prediction is incorrect (FN).

6. The accuracy parameter is used to measure how well data are matched. The accuracy is the proportion of correct predictions (both true positives and true negatives) among the total number of cases examined:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{10}$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

7. The precision determines how close the measurements are to each other. It is also defined as the proportion of true positives to all positive predictions, including false positives and true positives (also known as positive predictive value):

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{11}$$

8. The sensitivity (also known as Recall) determines the ability to identify points correctly. A high sensitivity means that more points are correctly detected, while a low sensitivity means that a lot of possible pairs of points are missing:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{12}$$

9. The F1 score is an overall measure of accuracy that combines precision and recall. If the F1 score is good (the high values), it seems that after the matching descriptor step, a low false positive and a low false negative exist:

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{13}$$

10. The Jaccard score (also known as the Intersection and Union) is used for evaluating the similarity and diversity of correct detected pairs of keypoints:

$$\text{Jaccard score} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} \tag{14}$$

11. The Root Mean Square Error (RMSE) and Sigma Median Absolute Deviation (SMAD) on signalised check points were used to assess the quality of point cloud registration and comparison of the results obtained from different distance measurements:

$$\text{RMSE} = \sqrt{\frac{\sum ((x_i - \mu)^2)}{N - 1}} \tag{15}$$

$$\text{SMAD} = 1.4826 \cdot \text{median}(|x_i - \delta|) \tag{16}$$

where $x_i$ is a value in the data set; $\mu$ is the mean, $N$ is the number of data points, and $\delta$ is the median.

## 3.2 Test Sites Description

In order to verify the impact of the selection of distance measurements, two types of test fields were used: (1) simulated data, for which:

- Gaussian noise and radial distortion,
- rotations and tilt were introduced,

and (2) real data (Figure 3):

- two decorated historical chambers at the Museum of King Jan III's Palace at Wilanów (Test Site I and II),
- a narrow office (Test Site III) located in the main hall of Warsaw University of Technology and
- a shopping mall, "Serenada", located in Krakow, Poland (Test Site IV), were selected.

The terrestrial laser scanning (TLS) data utilised in this investigation were acquired with use of the two phase-shift scanning instruments: the Z+F 5003 scanner with an angular scanning resolution 3.2 mm/10 m (deployed at Test Site I) and the Z+F 5006h scanner with resolution respectively 3.2 mm/10 m for Test Site II, 6.2 mm/10 m for Test Site III and 12 mm/10 m for Test Site IV. Those were acquired from different positions and heights (Figure 3). For the independent quality assessment, marked check points (not used for orientation parameters determination) were utilised. Table 1 provides summary information on the point clouds used in this investigation.

The selection of cultural heritage and public utility objects was caused by the characteristics of the acquired data, which exhibit heterogeneous structural attributes and surface geometries, allowing for the assessment of the efficacy and quality of determining tie points in the feature-matching steps using various distance metrics.

Test Site I is a complex geometric room with numerous ornaments, bas-reliefs, and facets. Alongside these, lavish gold-framed mirrors, an ornamental fireplace, and suspended fabrics grace the walls (refer to Figure 3a). In contrast to Test Site I, Test Site II does not contain ornaments and bas-reliefs, facets, or wall fabrics. Nevertheless, the walls still imitate the spatial effect due to the presence of wall paintings, as shown in Figure 3b. Test Site III comprises an office space featuring a slender lobby, sleek texture-free walls, and suspended lamps and power wires on the ceiling. Moreover, the floor is adorned with a dark carpet (Figure 3c). Test Site IV represents a standard empty retail space with untextured smooth walls. The concrete ceiling is decorated with lamps, electrical wires and a suspended air conditioner (Figure 3d).

**Table 1.** The list of the point clouds with parameters (Markiewicz, 2024)

| Test Site name | Scanner type | Angular resolution Horizontal | Angular resolution Vertical | Point scan resolution | No. scans | Avg. number of points per scan |
|---|---|---|---|---|---|---|
| I - "The Queen's Bedroom," in the Museum of King Jan III's Palace at Wilanów | Z+F 5003 | 1 scan -360°<br>5 scan -90° | 1 scan -320°<br>1 scan -180° | 3.2 mm/10 m | 6 | 1 scan – 42,308,262<br>5 scan – 126,913,021 |
| II - "The Chamber with a Parrot," in the Museum of King Jan III's Palace at Wilanów | Z+F 5006 | 360° | 320° | 6.3 mm/10 m | 4 | 40,320,455 |
| III – "The office room" in the main hall of Warsaw University of Technology | Z+F 5006 | 360° | 320° | 6.3 mm/10 m | 8 | 28,722,210 |
| IV - "Empty Shopping Mall" | Z+F 5006 | 360° | 320° | 12.1 mm/10 m | 7 | 13,677,292 |

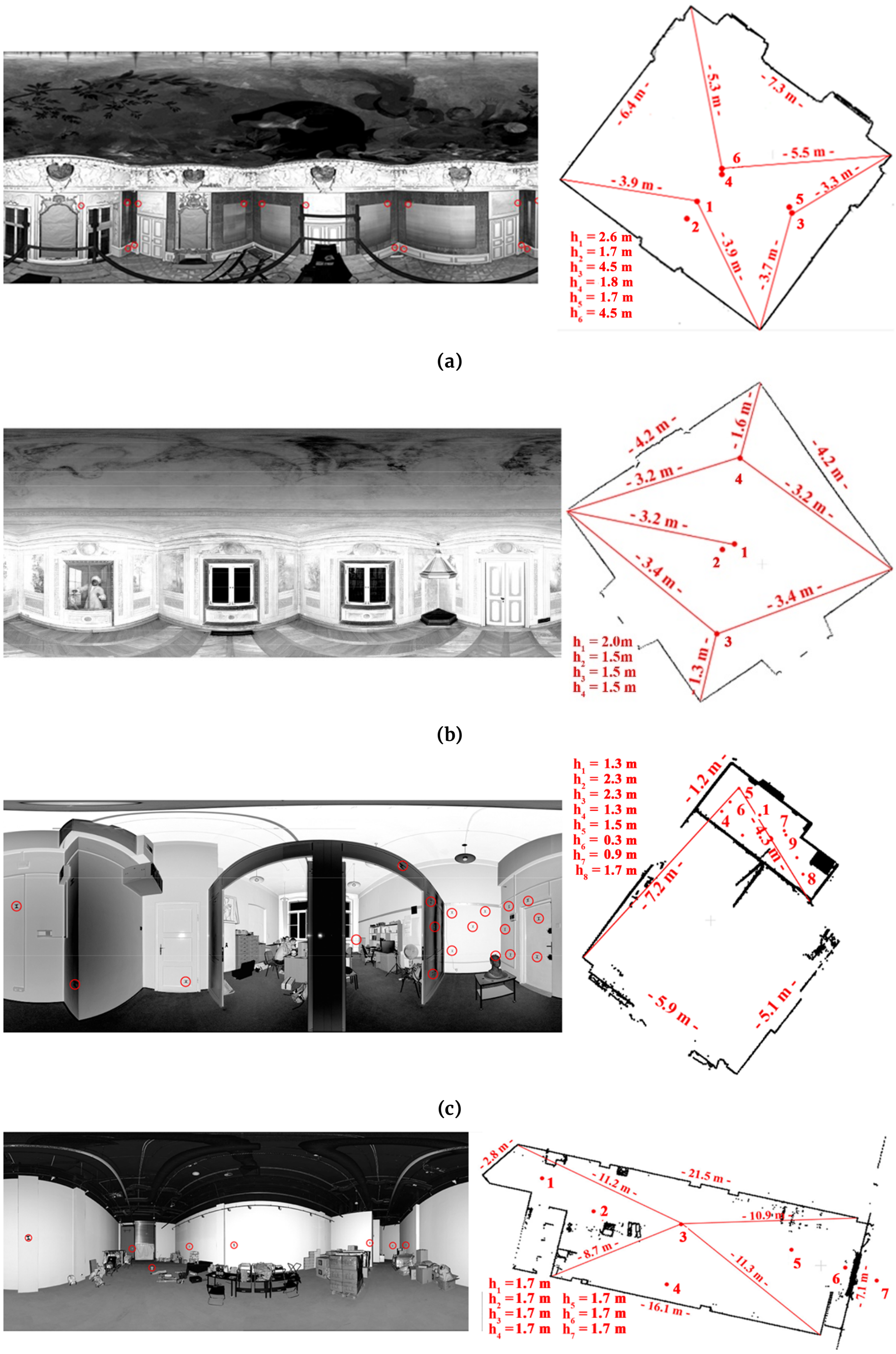



**Figure 3.** The point cloud in the spherical projection of: (a) Test Site I, (b) Test Site II, (c) Test Site III and (d) Test Site IV with marked points (red circles) (Markiewicz, 2024)

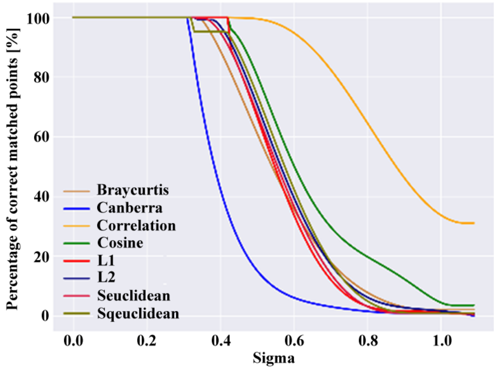



**Figure 4.** The Plot of the relationship between percentages correctly matched points and sigma value (Gaussian noise distortion) for all distance metrics

# 4 Results and Discussion

## 4.1 Synthetic data analysis

To assess the impact of the (1) variance of greyscale, (2) rotations and tilt, and (3) changes in image distortion on the quality of descriptor matching, synthetic data (a virtual test field) was prepared. The analysis of this abovementioned factor is a crucial step in keypoint matching because changes in the intensity of the laser beam reflection depend on scanning distances and beam incidence angles, which is unusual in images.

***Gaussian noise influence***

The initial investigations focused on assessing the effect of grey variance on descriptor-matching accuracy. The 128-parameter descriptor (equivalent to the size of the SIFT descriptor) was prepared for 2550 points consisting of random numbers between 0 and 1. To simulate the effect of grey variance, adding a Gaussian noise with a sigma value between 0 and 1.09 with a step of 0.01 was decided. This made it possible to asset descriptor matching accuracy between the output descriptor and the descriptor affected by Gaussian noise. Figure 4 shows the relationship between the percentage of correctly detected points and the sigma value for all the methods.

The results presented in Figure 4 for all distance measures indicate that the graphs will take on an inverted distribution function due to the representation of the value of correctly matched points on the graph rather than the result of incorrectly matched descriptors. This shows that the distance measures used for descriptors laden only with the variance of grey degree changes in the keypoint environment are robust to the occurrence of the image mentioned above's distortions. In contrast, differences are noticeable for sigma values, for which the fitted accuracy decreases (Figure 4).

Figure 4 shows that the measure least susceptible to mismatching descriptors is the correlation distance and the largest Canberra. Another distance measure to determine the tie points correctly is the Cosine measure. For the remaining distances, i.e. $L_1$, $L_2$, Seuclidean and Sqeuclidean, similar confounding effects of the sigma value on the correctness of the descriptor matching can be considered, as this is related to the specificity of the calculation of these measures.

To perform statistical evaluation (based on the measures described in Section 3.1), the F1-score, Jaccard index, and accuracy values were assessed (Figure 5).

Analysing the F1-score (Figure 5a) and Jaccard (Figure 5b) values, it can be seen that the shape of the approximation curves is similar to the curves shown in Figure 4. The only difference occurs for sigma values (the value on the x-axis shifted to the left relative to the original data), for which there is a significant decrease in the percentage correctness of the descriptor matching. When analysing the F1-score values for Canberra, it can be observed that a significant impact of variance intensity (Gaussian noise distortion) leads to a rapid decline in value. This indicates that using this matching metric results in low sensitivity (a high number of false negatives) and low precision (false positives). The highest independence of variance intensity (for the highest value of the sigma parameter) was obtained for correlation. The other values have similar magnitudes and comparable distributions. When evaluating Jaccard values, which do not account for True Negative values (unlike the F1-score), a similar trend can be observed as with the F1-score. This confirms that the number of false positives and false negatives is relatively low. The accuracy values (Figure 5c) approximately follow a Gaussian distribution, but differences can be seen in the peak height of the graph (y-axis) and the sigma values (x-axis). For the Barycurtis and $L_2$ measures, the lowest accuracy values were less than 10 per cent, and for sigma values, they were around 0.4. For the second group of distances, i.e. Cosine, $L_1$, Seuclidean and Sqeculidean, the accuracy values were around 15 per cent for sigma values between 0.45 and 0.5. For the Canberra measure, the lowest accuracy of 20% was the highest of the lowest accuracies obtained from all the measures. Still, it should be noted that descriptor matching using this measure is the least robust to the occurrence of Gaussian noise in the descriptors. The correlation measure had the second-highest accuracy, with the highest degree of noise robustness.

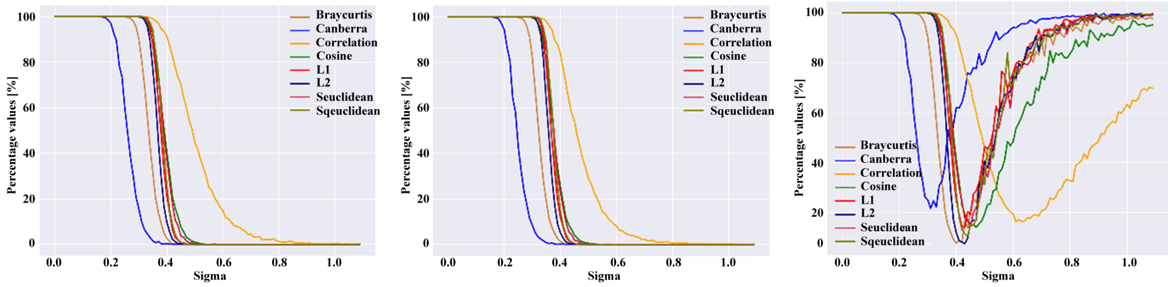
***Distortion effect influence***

One of the most common distortions of spherical images results from the conversion of point clouds to the raster form, based on the cartographic transformation that converts data from 3D to 2D. For this reason, it is essential to know how this distortion affects the choice of descriptor distance measurement. This translates into the correctness of detection and matching and the number and distribution of tie points. For this purpose, a simulation of radial distortion for both the "barrel" and the "pincushion" distortion cases was performed. Figures 6a and 6c show examples of the distorted images used in the analyses.

The distribution of the percentage of correctly matched points (Figure 6d) is symmetrical to 0 (the original image), which indicates that the effect of the "barrel" and the "pincushion" distortion on the accuracy of descriptor matching is the same, and the distribution resembles the graph of a homographic function. These distributions show that the highest percentage of correctly matched points was achieved with the Canberra distance measure. The values of the other measures are similar but, on average, about 10% lower than those of the Canberra distance measure.

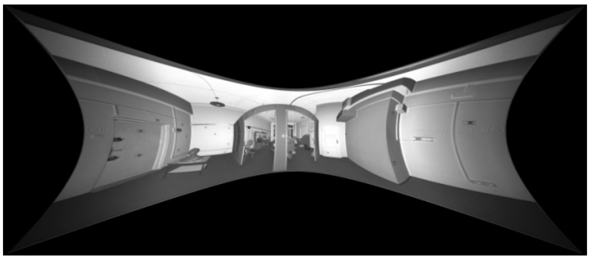
Analysing the effect of distortion on descriptor-matching accuracy reveals that all the methods are susceptible to its impact. This is shown in the low F1-score and Jaccard values, which for Cosine distance (the best-performing method) are 65% and 50%, respectively. Analysing the individual values for distance metrics reveals two distinct groups: those associated with the $L_1$, $L_2$ norms, Sececlidean, and Sqeuclidean, and those associated with Braycurtis, Canberra, Correlation, and Cosine (Figure 7a and 7b). Evaluating the results obtained for the first group of distance metrics indicates that they have similar values, differing by an average of about 1% and from Canberra by around 5%. Assessing the second group of distances reveals that they are higher, averaging around 5%. The best results were obtained with the Cosine method (directly related to the calculation method of this metric), while the worst results were observed for Canberra.

It should be noted, however, that all the metrics show low resilience to the significant impact of distortion (Figure 7c). At low distortion levels, they match tie points accurately. As image distortions increase, this accuracy rises due to the correct prediction and not due to determining of tie points. Evaluating the values obtained for each distance metric separately, it can be concluded that the most minor differences (and thus the best accuracy) were achieved




**(a)** F1-score **(b)** Jaccard **(c)** Accuracy

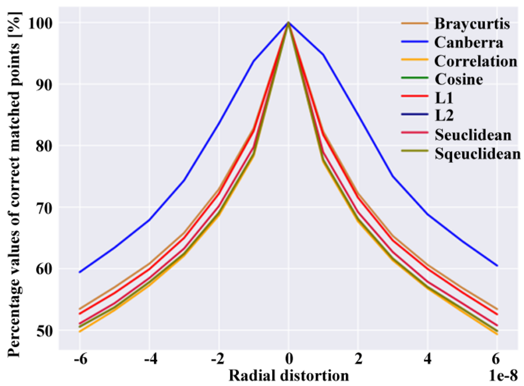
**Figure 5.** The plot of the relationship between percentage values of: (a) F1-score, (b) Jaccard and (c) Accuracy values and sigma value (Gaussian noise distortion) for all distance metrics



**(a)**

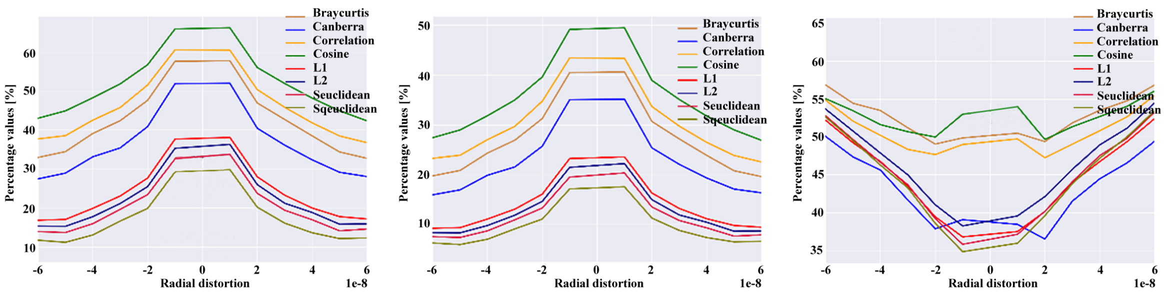
**(b)**

**(c)**




**(d)**

**Figure 6.** The example of: (a) a maximum "pincushion" distortion image, (b) an original image, (c) a maximum "barrel" distortion image, (d) a plot of the relationship between percentage values of corrected matched points and distortion values




**(a)** F1-score **(b)** Jaccard **(c)** Accuracy

**Figure 7.** The plot of the relationship between percentage values of: (a) F1-score, (b) Jaccard and (c) Accuracy values and distortion for all distance metricises

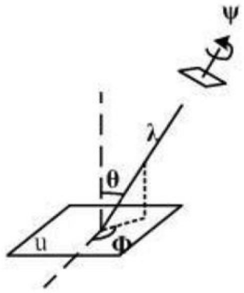



**Figure 8.** Geometric interpretation of affine decomposition (Yu and Morel, 2011)

with Cosine, Braycurtis, and Correlation. For the other measures, these values depend more on the impact of distortion, with the poorest results obtained for Sqeuclidean.

***Skew (tilt) and rotations effect influence***

The influence of the image's tilt and rotations on the correctness and accuracy was also checked using different distance metrics. For this purpose, we used a decomposition of the parameters mentioned above based on the method presented in the article *ASIFT: An Algorithm for Fully Affine Invariant Comparison* (Yu and Morel, 2011) and used in Affine-detectors for point detection and matching. In the ASIFT algorithm, each image is transformed by simulating all possible affine distortions caused by the change of the initial camera positions. For this purpose, affine decomposition is utilised to describe the transformation using the angle of rotation around the optical axis (spin, angle $\psi$), the skew angle $\theta$ (understood as the camera tilt angle), and $\phi$, which defines the rotation around the Z-axis (Figure 8). To perform this, Equations 17 and 18 are utilised (Yu and Morel, 2011):

$$u(x,y) \rightarrow u(ax+by+e, cx+dy+f) \tag{17}$$

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = H_\lambda R_1(\psi) T_t R_2(\phi) = \lambda \begin{bmatrix} \cos\psi & -\sin(\psi) \\ \sin(\psi) & \cos\psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin(\phi) \\ \sin(\phi) & \cos\phi \end{bmatrix} \tag{18}$$

where $\lambda > 0$ is the determinant of $A$ (affine transformation matrix), $R_i$ are the rotations, $\phi \in [0, \pi)$ and $T_t$ is the tilt, namely a diagonal matrix with first eigenvalue $t > 1$ and the second one equal to 1. It is possible to prepare the decomposition of the camera motion parameters into the viewing point angles (longitude ($\phi$) and latitude ($\theta = \arccos(\frac{1}{t})$)), spin of the camera ($\psi$) and zoom factor ($\lambda$). In the ASIFT algorithm, images undergo rotation with angle $\theta$, which is represented by tilt parameter $t = \frac{1}{\cos\theta}$.

From the results shown in Figure 9a, it can be observed that the image skew significantly decreases the accuracy of descriptor matching for distance measures. For a skew of 50 degrees, there is a noticeable trend of a rapid decrease in the accuracy of descriptor matching. Despite this, it should be noted that the least susceptible measure to incorrect descriptor matching for skew images is the Canberra distance, and the results for the other distance measures are similar.

Assessing the effect of the rotation of the image around the Z-axis on the correctness of descriptor matching (Figure 9b), it might be stated (similar to the skew effect) that the highest percentage of correctly matched descriptors was obtained for the Canberra measure. For other distance metrics, the matching accuracies are very close. Figure 9b shows that between 0 and $\pm 22.5°$, the decrease in accuracy of descriptor matching using all accuracy measures is marginal and does not exceed 5%. Above this value, a significant decline in accuracy is noticeable from a value of $\pm 80°$.

Rotation of the image around the optical axis (depth direction) has no negative impact on the accuracy of the descriptors as much as the influence of tilt and rotation around the Z-axis. The highest percentage of correctly matched tie points were obtained using the Canberra measure and successively for $L_1$ norm, Seuclidean, $L_2$ norm, Cosine, Sqeuclidean and Correlation. As the rotation of the image concerning the optical axis increases, a successive decrease in the percentage of correctly matched tie points can be observed. In the range from 0 to 45 degrees, these values do not exceed 90%; in the range from 45 to 90 degrees, 80%; in the range from 90 to 180 degrees, 60%; and in the range from 180 to 359 degrees, 40%.

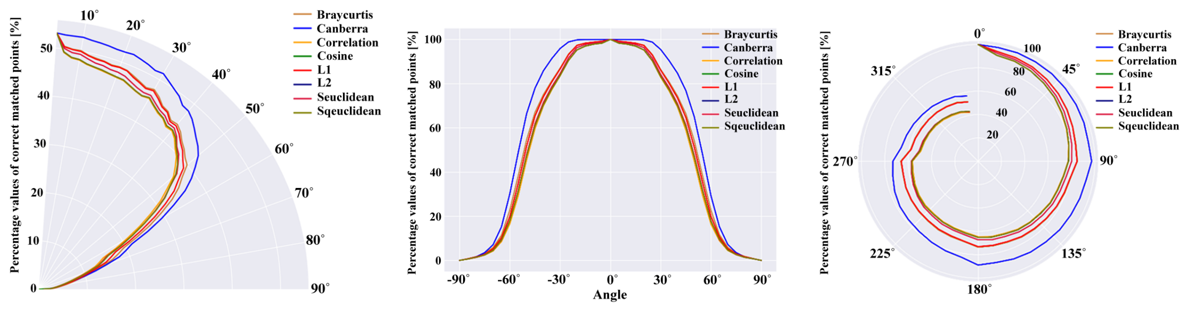
A classical approach using SIFT was used in this investigation to detect tie points. As shown in Figures 10a–10c, the most significant decrease in F1-score and Jaccard values is observed when the image skew exceeds 5 degrees. As previously described, this reduces the number of tie points and results in the missed detection of points with counterparts on the matched image. Assessing the Accuracy value (Figure 10c) reveals that, between 5 and 60 degrees, the Accuracy values remain stable at an average of 50% across all distance metrics, indicating considerable difficulty in accurate descriptor matching. From around 40 degrees onward, an increase in Accuracy is noticeable – tie points are correctly not matched.

A similar relationship was observed for the values obtained for rotation about the Z axis (Figure 10d and 10e). Evaluating the distribution of the F1-score and Jaccard values, it can be seen that there is a significant decrease in these values for about 15 degrees of rotation to the left and right. From about 30 degrees onwards, there is a value of approximately 0%, which indicates, at the very least, a lack of correct detection of the vantage points. Assessing the Accuracy values (Figure 10f), it can be seen that from around 5 degrees, there is a significant decrease in accuracy (no points detected) and from around 15 degrees, accuracy increases correct "mismatching" of tie points and a decline in the number of tie points detected. It should be noted that these values are similar for all distance metrics. It is impossible to divide them into groups, as was possible when analysing the impact of distortion.

The final assessment focused on the F1-score and Jaccard values obtained for rotation around the optical axis. The results indicate that this rotation has the most significant impact on descriptor matching accuracy. The data in Figure 10g show that all distance metrics, except for Canberra, exhibit low coefficients within the range of 0 to 12% for rotation angles between 10 and 360 degrees. Only for Canberra do these values remain stable until the rotation angle exceeds 80 degrees. Evaluating the Accuracy values (Figure 10i) reveals a linear increase in errors as the rotation angle grows, which, as with previous cases, is due to the correct "non-detection" of tie points.

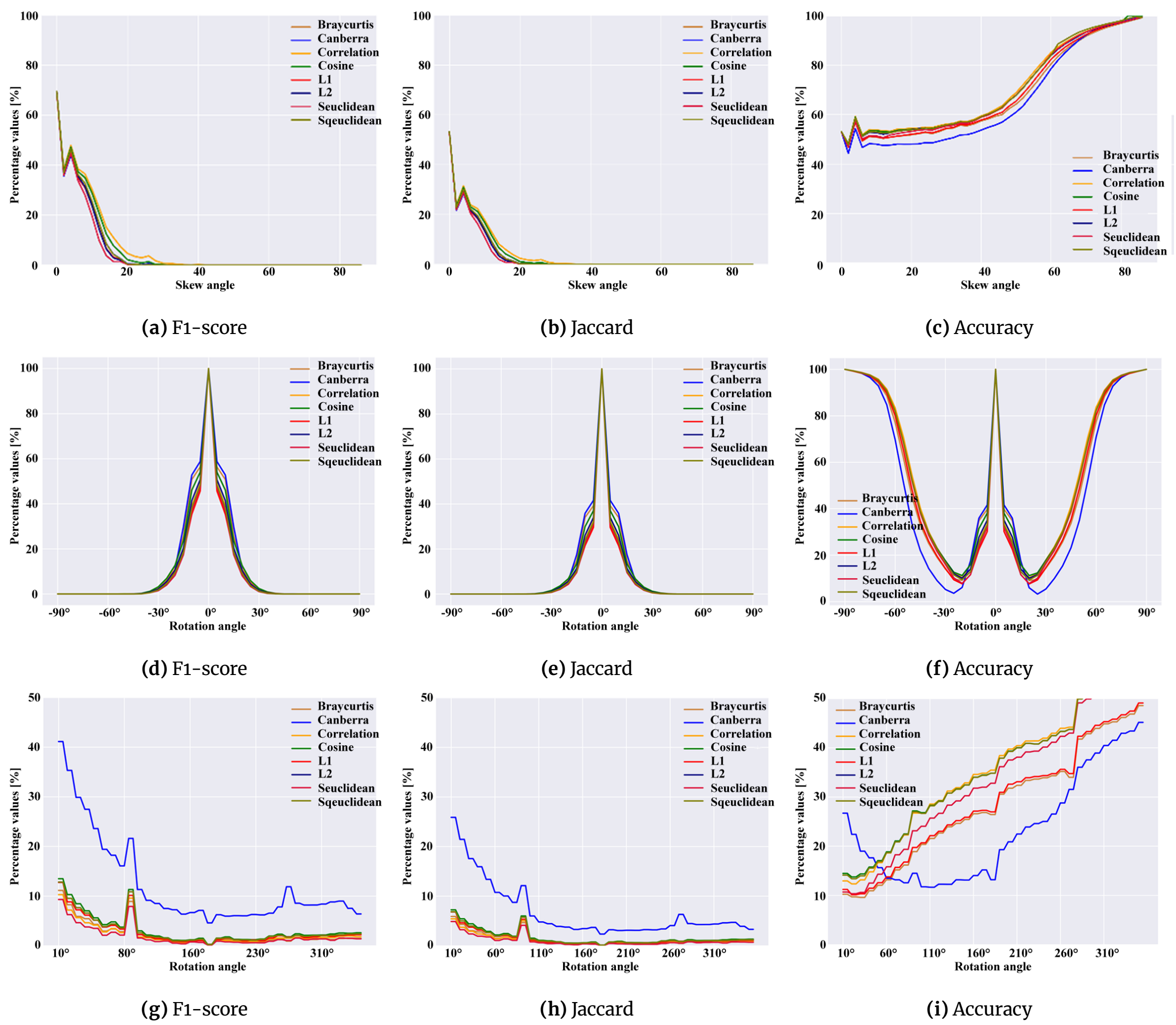
***The summary of analyses on synthetic data***

In detecting and matching keypoints on point clouds converted to the raster form, several issues may arise related to input data quality and cartographic transformations' effects on data conversion from the 3D to the 2D form – particularly noticeable in the TLS-SfM process. Problems associated with converting point clouds to the raster form have been detailed in Markiewicz and Zawieska (2019). These issues are related to (1) raw intensity deviations caused by the angle relative to the normal of the measured surface and the scanning distance and properties of the measured material; (2) deformations of a generated image, such as the effect of "distortion" and large deformations, which occur for large values of angles, i.e., in the upper and lower parts of the raster. These geometric failures considerably influence the number and the distribution of tie points detected by algorithms applied in image processing (Markiewicz and Zawieska, 2019). For this reason, it was decided to simulate the impact of intensity changes by introducing Gaussian noise into the reference descriptor and geometric distortion changes by generat-




(a) (b) (c)

**Figure 9.** The plot of the relationship between percentage values of corrected matched points and (a) skew angle, (b) rotation around the Z-axis and (c) rotation around the depth direction (optical) axis

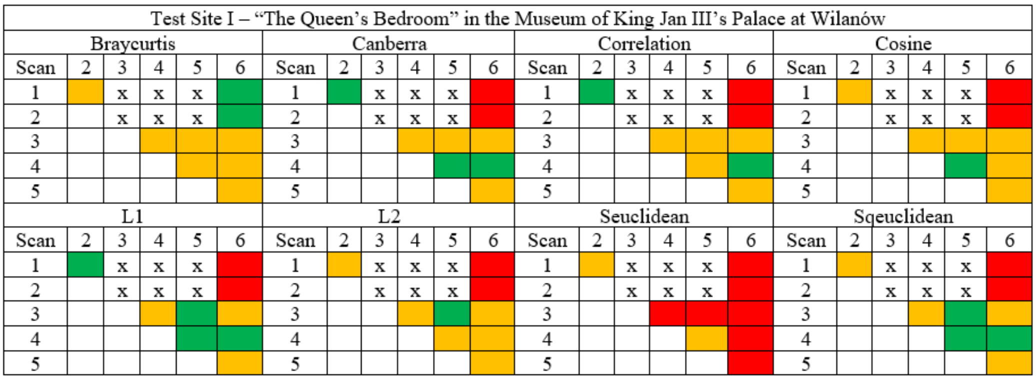



(a) F1-score (b) Jaccard (c) Accuracy

(d) F1-score (e) Jaccard (f) Accuracy

(g) F1-score (h) Jaccard (i) Accuracy

**Figure 10.** The plot of the relationship between percentage values of the F1-score, Jaccard and Accuracy values and (a)–(c) skew angle, (d)–(f) rotation around the Z-axis and (g)–(i) rotation around the depth direction (optical) axis



| Test Site I – "The Queen's Bedroom" in the Museum of King Jan III's Palace at Wilanów | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Braycurtis | | | | | | Canberra | | | | | | Correlation | | | | | | Cosine | | | | | |
| Scan | 2 | 3 | 4 | 5 | 6 | Scan | 2 | 3 | 4 | 5 | 6 | Scan | 2 | 3 | 4 | 5 | 6 | Scan | 2 | 3 | 4 | 5 | 6 |
| 1 | | x | x | x | | 1 | | x | x | x | | 1 | | x | x | x | | 1 | | x | x | x | |
| 2 | | x | x | x | | 2 | | x | x | x | | 2 | | x | x | x | | 2 | | x | x | x | |
| 3 | | | | | | 3 | | | | | | 3 | | | | | | 3 | | | | | |
| 4 | | | | | | 4 | | | | | | 4 | | | | | | 4 | | | | | |
| 5 | | | | | | 5 | | | | | | 5 | | | | | | 5 | | | | | |
| L1 | | | | | | L2 | | | | | | Seuclidean | | | | | | Sqeuclidean | | | | | |
| Scan | 2 | 3 | 4 | 5 | 6 | Scan | 2 | 3 | 4 | 5 | 6 | Scan | 2 | 3 | 4 | 5 | 6 | Scan | 2 | 3 | 4 | 5 | 6 |
| 1 | | x | x | x | | 1 | | x | x | x | | 1 | | x | x | x | | 1 | | x | x | x | |
| 2 | | x | x | x | | 2 | | x | x | x | | 2 | | x | x | x | | 2 | | x | x | x | |
| 3 | | | | | | 3 | | | | | | 3 | | | | | | 3 | | | | | |
| 4 | | | | | | 4 | | | | | | 4 | | | | | | 4 | | | | | |
| 5 | | | | | | 5 | | | | | | 5 | | | | | | 5 | | | | | |

**Figure 11.** The accuracy of the TLS registration for Test Site I

ing virtual images with the influence of radial distortion and image rotations and tilt.

When evaluating the impact of individual factors on the accuracy and reliability of tie-point matching, it can be observed that image rotation angles and skew have the most significant influence. In contrast, deviations in intensity values have the most negligible impact. Considering the effect of scanning surfaces at large angles and/or close distances, it must be concluded that distortion can occur on raster fragments (Figure 9a), and fragments can be rotated relative to each other in the Z-axis (Figure 9b). For this reason, the effect of distortion and rotation should be analysed together. When performing TLS measurements, it is assumed that the unit should be level and that the use of compensators will allow this condition to be met. An analysis of the rotation of the raster around the "optical axis" showed that the accuracy of descriptor matching when using the Canberra measure was close to 100%, and for the other measures was above 95%. For this reason, it can be assumed that the influence of this rotation can be considered negligible when selecting a descriptor-matching measure.

The effect of distortion is noticeable when processing point clouds are obtained from scanner positions close to walls, where the part of the objects are scanned with acute scan angles to the surface normal vector. With this in mind, it can be expected that fewer feature tie points will be detected and matched in such sections of point clouds. Therefore, it is advisable to consider the placement of TLS stations in a way that ensures not only accurate shape representation but also effective tie-point detection in the TLS-SfM process.

## 4.2 Real data analysis

The next investigation involved analysing the distance metrics selection on real data characterised by different geometric complexity and texture. For this purpose, the following parameters were checked: (1) the correctness of the pairwise point cloud registration (with the analysis used in Machine Learning), (2) the accuracy of the pairwise point cloud registration, and (3) the number and distribution of points used in the combined bundle adjustment.

***Evaluation of the Accuracy of Automatic Matching of Pairs of Scans***

Figures 11–14 present the results obtained for all distance metrics (Test Sites I–IV). To categorise the obtained results, the following colours were used: (1) full registration (green), where the RMSE for X, Y, and Z coordinates $\leq$ 0.005 m (for Test Sites I–III)/$\leq$ 0.01 m (for Test Site IV), and points are evenly distributed within the analysed area, (2) preliminary orientation parameters that should be used in the ICP (orange), and (3) no registration (red). The symbol "x" indicates that pairs of scans could not be registered due to the insufficient overlap.

From the results presented in Figure 11 for Test Site I, it appears that it is only possible to register all scans using the Barycurtis measure for all pairs of scans acquired from different distances, heights and angles. The worst results were obtained for the Seuclidean method, for which only 2 pairs out of 9 could be registered. The other methods allowed only 7 out of 9 pairs of scans to be registered. It was problematic to correctly register the point clouds for which the spherical images had significantly different "distortion" (scans 1, 3 and 19) for the corresponding fragments due to the impact of point cloud conversion to the spherical image form. It should be noted that for Test Site I, only scan 19 depicted the entire site, with the others only depicting individual walls.

The results obtained for Test Site II (Figure 12) show that it is possible to register all point clouds regardless of metricise distances. Compared to Test Site I, this is because all point clouds were acquired over the full angular range – all room walls, ceiling and floor were mapped. In addition, it should be mentioned that it is possible to use all measures for objects characterised by complex good and unambiguous textures and spherical images characterised by similar "distortions". A decisive aspect for selecting a specific solution is the number of points, their distribution and the achievable accuracies described in the following subsection.

Based on the results presented for Test Site III (Figure 13), the worst outcomes were observed for the Barycurtis, Correlation and Seuclidean distance metrics. In contrast, the best results were achieved using the L1 norm. It should be noted that in cases where the point clouds were acquired over the full angular range, the baseline between the point cloud pairs was small, and the distribution of the scanner position affected the significant effects of "distortion" on the spherical images, not all the methods allowed pair scans to be registered and only the Canberra method allowed the correct orientation of pairs 1–7. Despite this, all the methods allowed full final registration of all point clouds, as indicated by the number of pairs with pre-orientation (orange) and those without orientation (red) in all distance metrics. Similar to Test Site II, the choice of method depends on the number, distribution, and expected registration accuracy.

Significant differences in the selection of distance metrics when matching keypoints used to register point clouds can be seen for Test Site IV (Figure 14), for which point cloud pairs are acquired from a wide-range base, and the object is characterised by homogeneous texture or textureless areas. In this case, the best results (guaranteeing the best determination of tie points on point clouds) were obtained for the $L_1$ norm, Sqeuclidean, $L_2$ norm and Cosine method, respectively. It should be noted, however, that a more significant number of pre-preliminary registered pairs (orange colour) may contribute to fewer tie points for the final bundle adjustment of all TLS pairs and it may also affect the final accuracy of registration. These aspects are evaluated and discussed in the following subsections.

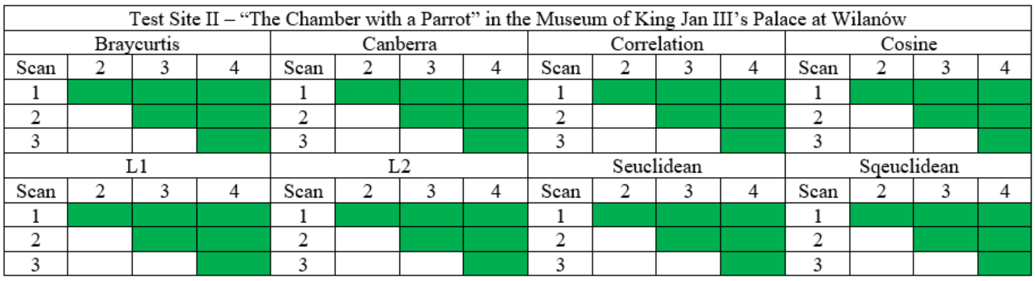



**Figure 12.** The accuracy of the TLS registration for Test Site II

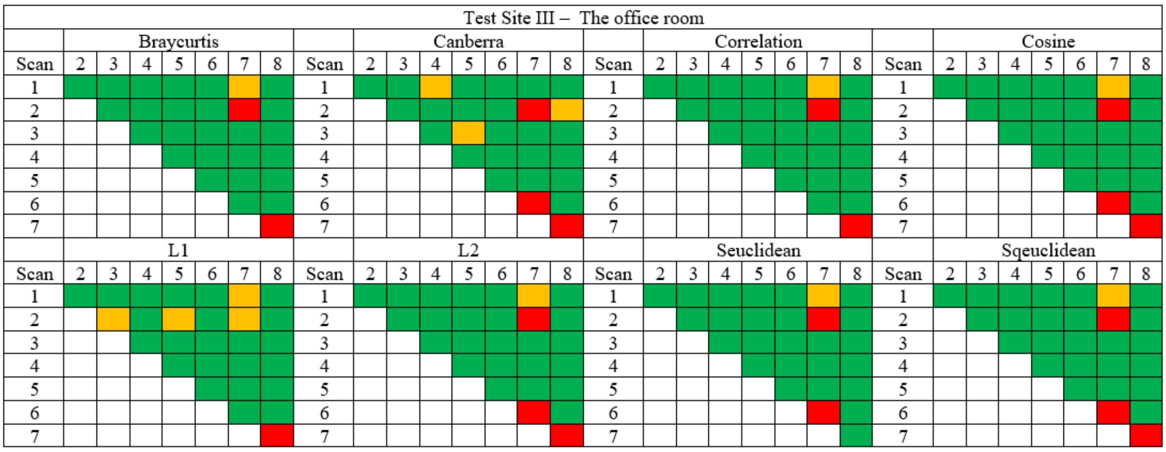



**Figure 13.** The accuracy of the TLS registration for Test Site III

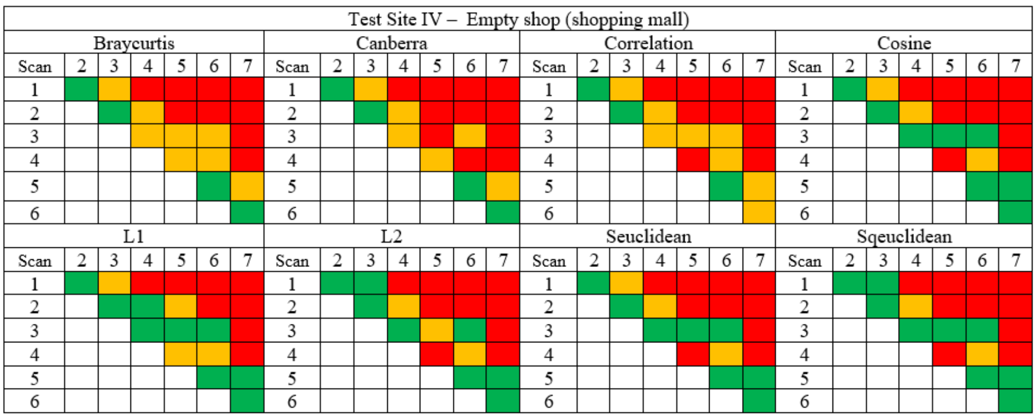



**Figure 14.** The accuracy of the TLS registration for Test Site IV

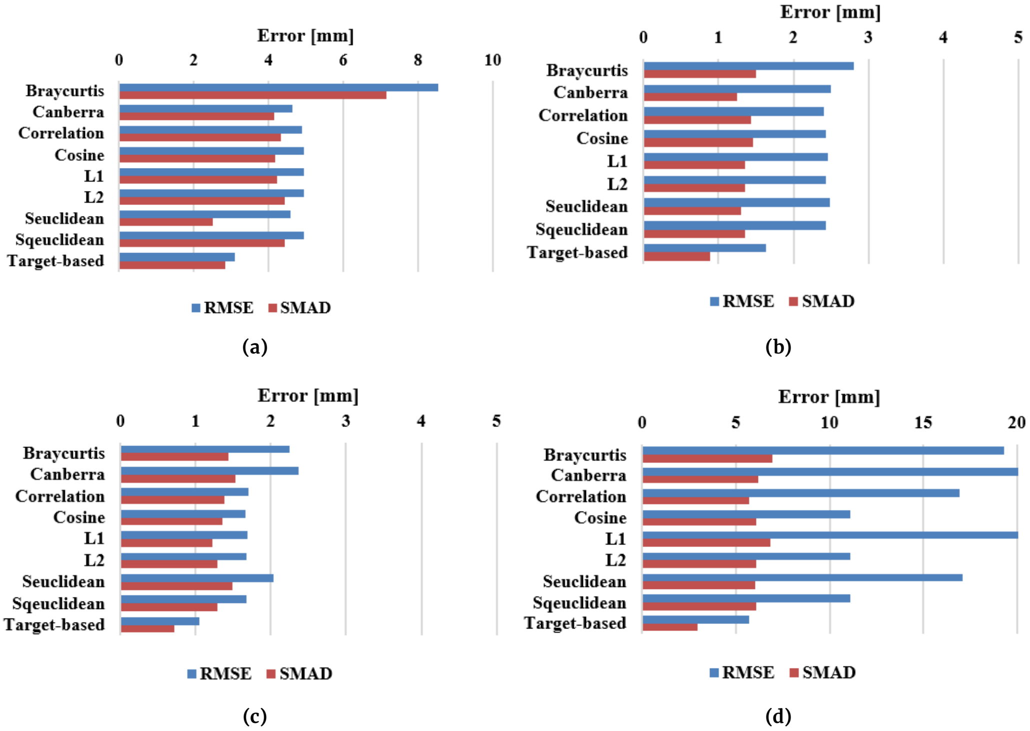



**Figure 15.** The plots of linear error values for RMSE and SMAD for: (a) Test Site I, (b) Test Site II, (c) Test Site III and (d) Test Site IV

*Accuracy Analysis of Signalised Check Points*

To assess the accuracy of the point cloud orientation process, values of deviations (from the full registration) of signalised (Test Site I, III and IV) and natural (Test Site II) check points were used; those points were used for independent quality assessment. To assess the accuracy of TLS point cloud registration, linear values of RMSE and SMAD deviations were used (Figure 15), along with the distribution of error values presented as boxplots (Figure 16).

Comparing the RMSE and SMAD values for Test Site I (Figure 15a), it can be concluded that the differences between them do not exceed (a) 0.5 mm for Canberra, Sqeuclidean, and Target-based; (b) 0.8 mm for Correlation, Cosine, $L_1$, and $L_2$; (c) 1.5 mm for Braycurtis; and (d) 2 mm for Seuclidean, indicating the absence of outliers in the tie points. The RMSE values for all the methods, except Braycurtis, fall within the range of 4.5 mm to 5 mm, with the RMSE for Braycurtis reaching 8.5 mm. Comparing these results with the RMSE values obtained for the Target-based method, the most negligible difference was found for Seuclidean (1.5 mm), while the largest was for Braycurtis (5.4 mm). It can be concluded that, for all the methods except Braycurtis, the results are similar to those of the Target-based method.

The RMSE values obtained for Test Site II (Figure 15b) show an improvement in point cloud registration accuracy compared to Test Site I. Although both sites are characterised by good texture (significant changes in grayscale gradients), two key factors contributed to the improved registration accuracy: the acquisition of point clouds at full angular resolution and smaller differences in the height of the registered scanner positions. However, like Test Site I, the worst results were obtained with Braycurtis. The differences between RMSE and SMAD do not exceed (a) 1 mm for Correlation, Cosine, and Target-based; (b) 1.2 mm for Canberra, L1, L2, Seculidean, and Sqeuclidean; and (c) 1.3 mm for Target-based method (with differences not exceeding 1.2 mm). Therefore, it can be concluded that they enable TLS point cloud registration that is comparable to the commonly used state-of-the-art methods.

For Test Site III (Figure 15c), similar trends can be observed as for Test Site II, except for the method with the highest RMSE value – Barycurtis for Test Site II and Canberra for Test Site III. Additionally, there are no significant differences between RMSE and SMAD; they do not exceed 0.8 mm for Barycurtis and Canberra, while for the other distance metrics, they remain below 0.5 mm. Comparing the obtained RMSE values with those from the Target-based method, it can also be concluded (as with Test Site II) that the TLS registration results are comparable to those of state-of-the-art approaches, with 1.5.mm differences being negligible.

The worst results were obtained for Test Site IV (Figure 15d). Comparing the RMSE and SMAD results, it is noticeable that (a) the RMSE error for Cosine, $L_2$, and Sqeuclidean is approximately twice as high, (b) about three times as high for Braycurtis, $L_1$, and Seuclidean, and (c) seven times fas high or Canberra. The highest RMSE values were achieved with Braycurtis, Canberra, and $L_2$, while the best results were obtained for $L_2$, Cosine, and Sqeuclidean. Although the errors for these distance metrics differ by approximately 5 mm from the Target-based method, these results should be considered acceptable, as they are lower than the scanning resolution of 12.1 mm at 10 m.

Due to the minor differences between the RMSE values for the various distance metrics, it is also essential to analyse the distribution of deviations presented in the boxplots (Figure 16).

When analysing the values presented in Figure 16a, variability in maximum and minimum values, as well as different sizes of IQR for the various distance metrics, were observed. The most significant spread of values was noted for Braycurtis, while the smallest was for Seuclidean (X, Y and Z-coordinates). However, considering that the Seuclidean method has the least number of registered scan pairs, this value should not be considered further, and the values for Correlation should be regarded as the smallest. For the Braycurtis and Correlation methods (across all axes), Canberra (for the Y and Z axes), and Cosine (for the Z axis), the median value exceeds 1 mm, indicating the presence of systematic errors. For the other methods, these values also do not equal 0, but the deviations do not exceed 1 mm. Evaluating the first quartile (Q1), the third quartile (Q3), and the interquartile range (IQR), it can be concluded that for the X axis, there are disparities in the distribution (uneven relative

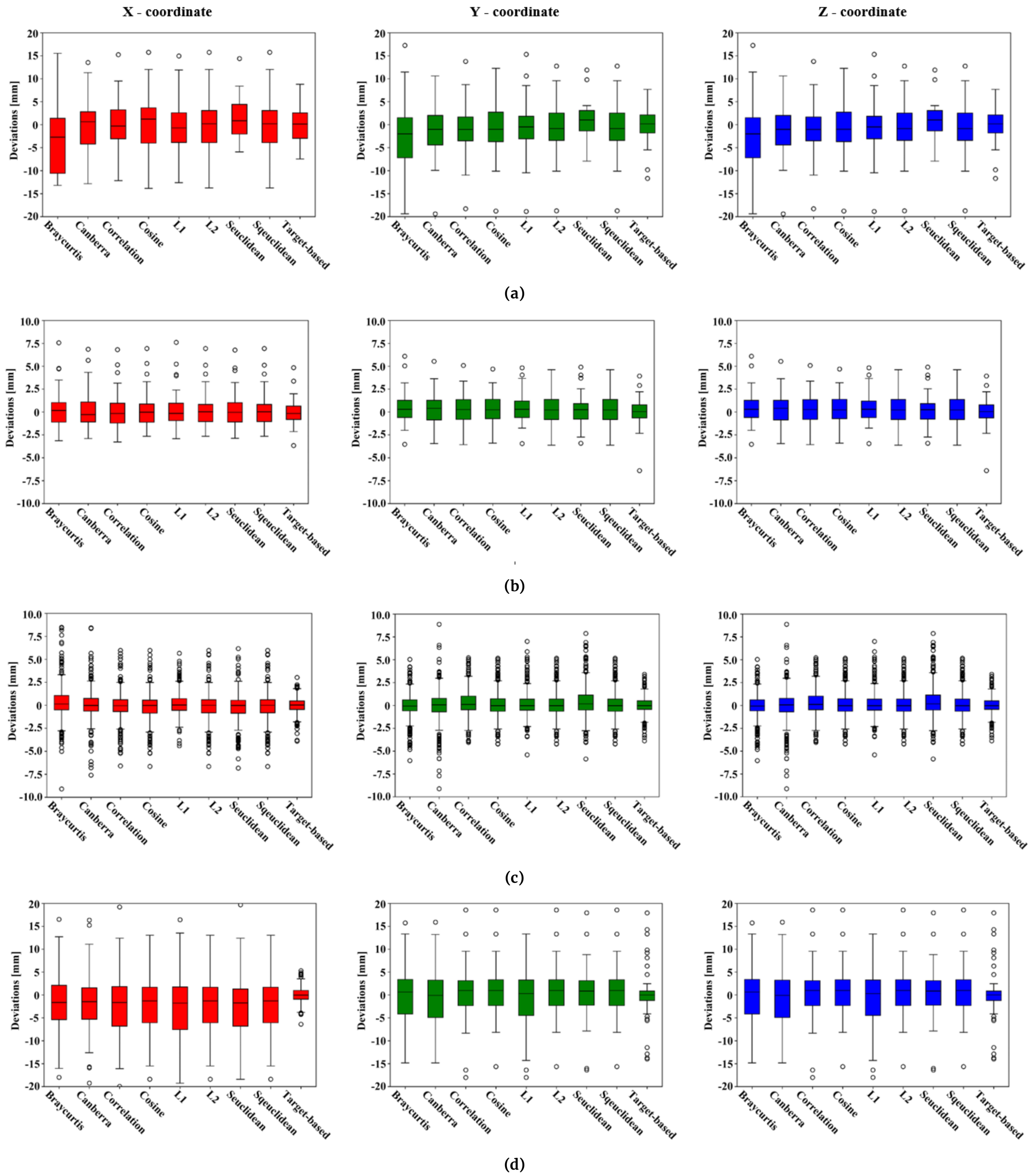



**Figure 16.** Box plots for the distribution of the deviations on signalised check points for all pairs of point clouds fully registered scans for: (a) Test Site I, (b) Test Site II, (c) Test Site III and (d) Test Site IV

to the median). Smaller disparities are noticeable for the Y and Z coordinates. Uniform distributions of values were obtained for the $L_1$, $L_2$, Sqeuclidean, and Target-based methods, with the most similar distribution achieved for the L1 method. On average (across all axes), the highest number of outliers (marked as circles) was obtained for Braycurtis (20 outliers), while the fewest were obtained for Canberra and Cosine, each with one outlier. For the remaining methods, there were two outliers each.

In contrast to Test Site I, for Test Site II (Figure 16b), the distribution of deviation values is uniform relative to the median for all the methods. For all distance metrics for the X, Y, and Z components, the median value does not exceed 0.3 mm, which allows us to conclude that there are no systematic errors in the data set. The smallest Q1 and Q3 values were obtained for Seuclidean, while the largest were for L2. The boxplot analysis shows that when registering point clouds over a full scanning range (360°) and processing objects with good texture, the choice of distance metric is less significant, as all the methods enable registration with high accuracy.

A similar uniform distribution of deviations as observed for Test Site II was achieved for Test Site III (Figure 16c). Although the scanned room is not characterised by the good texture and had flat white walls, and the scans were taken at close range to the walls, high registration accuracy comparable to the Target-based method was achieved. However, the near distance to the walls, while advantageous for achieving significant point cloud density, resulted in substantial distortions in the spherical images. This is evident from the large number of outliers marked as circles. Analysing all components (X, Y, Z), it can be seen that Canberra recorded the highest number of outliers. Assessing the remaining distance metrics, it can be clearly stated that the best results were achieved for $L_1$ and $L_2$.

The results presented for Test Site IV (an empty shop in a shopping mall) indicate the presence of systematic errors in the orientation of the data for the X coordinate, with all average values being approximately -1.5 mm. The boxplots show that the best results were obtained for $L_2$, Sqeuclidean, and Correlation. The distributions are uniform relative to the median, and the IQR values are similar. Similarly to Test Site III, there are many outliers in the data sets. The obtained maximum and minimum values range from ±15 mm, while 50% of the error values for the best distance metrics fall within the ±5 mm range. This, concerning the scanning resolution, indicates the correctness of the data registration process.

***Point distributions***

To assess the density and distribution of the identified tie points, the datasets were divided into an Octree with dimensions of $2 \times 2 \times 2$ meters (Figure 17) and the number of points in individual cubes was presented using bar charts (Figs. 18–21).

When assessing the number and distribution of points obtained for Test Site I (Figure 18), which features numerous architectural details, bas-reliefs, and facets, as well as the point clouds acquired, an uneven distribution of points within specific cubes can be observed, along with significant variation in the results obtained for different distance metrics,. This relationship might be caused by the utilisation of point clouds acquired from significantly different heights and capturing the same wall sections at substantially varying angles – an effect similar to the impact of distortion and rotation around the Z-axis, as discussed in section 4.1.

Evaluating the distribution (the filling of individual cubes with points; Figure 18), it can be observed that the best results were obtained for the Braycurtis metric. At the same time, the worst ones were obtained for the Squared Euclidean (Seuclidean) metric. Comparing the results for the remaining methods, it is noticeable that they are similar for all the methods except Canberra, for which no points were detected in cubes where tie points should have been located.

In the assessment of the results for Test Site II (Figure 19), which is characterised by simple geometry but featuring wall paintings that imitate a 3D effect, only the points obtained from point clouds acquired at different angles and distances relative to the measured walls were evaluated (without the first pair of scans). However, it should be emphasised that these differences in the scanning angle relative to the normal of the wall surfaces are not as significant as in the case of the point clouds acquired for Test Site I. Evaluating the distributions and the number of points in individual cubes, it can be observed that the results for all the methods are similar. Only for Canberra, in octree cube 85, a slightly higher number of points was observed.

The results presented in Figure 20 indicate (similarly to the previous Test Sites) that, for all the methods, the distribution of points within individual cubes is similar. Significant differences are noticeable for cubes from 35 to 40 and cube 55. It should be noted that all the methods enabled the detection of multiple points due to the small distance between the scanner position and the measured wall. The best results were obtained with Barycurtis and Canberra, respectively. However, when comparing these results to those for Test Site II, the number of points is significantly lower. This is because, unlike Test Site III, Test Site II features better texture, resulting in significant changes in grayscale gradients. These changes impact both the number of detected keypoints and the number of correctly matched descriptors.

Evaluating the distribution and number of matching points obtained for Test Site IV (Figure 21), it can be observed that the worst results (fewest points) were achieved with Barycurtis. In contrast, the best results were obtained with Cosine. It is also noticeable that all modifications of the Euclidean measure, i.e., $L_1$, $L_2$, Seculidean, and Sqeuclidean, exhibit similar distributions and point counts. Similar to Test Site III, there is a lower number of matching points for textureless objects compared to Test Sites I and II. This trend is associated with changes in grayscale gradients and the significant distance between the scanner positions and the measured wall segments, which results in a lower point cloud density and quality of the utilised spherical images.

***The statistical analysis of the performance of utilised metricises***

To evaluate the accuracy of pairwise point cloud registration, commonly used metrics in machine learning literature, such as accuracy, precision, recall, and F1-score, were selected for analysis. These metrics were applied, allowing a comprehensive assessment of the matched key points and the effectiveness of eliminating the impact of incorrect descriptor matching. Given that multiple pairs of point clouds were analysed, the results include the median values of these metrics and their minimum and maximum values. The results are presented in Table 2.

Evaluating the obtained Accuracy, Precision, Recall, and F1-score values, it can be unequivocally stated that regardless of the distance metrics used, the model has difficulties in correctly matching all possible tie points. This is evident in the F1-score, recall, and precision values. Therefore, it is essential to analyse these values for individual test fields separately, characterised by varying geometrical complexity and TLS scanner positions.

Evaluating the values obtained for Test Site I, characterised by geometrical complexity and a large number of architectural decorations acquired from significantly different heights and capturing the same wall sections at substantially varying angles, it can be stated that:

- High Accuracy values indicate many correctly unmatched keypoints (True Negatives), which is the dominant class – averaging 98.8% across all the methods. The average spread (understood as the difference between the maximum and minimum value) was approximately 1.2%, with the most significant spread for Correlation (1.5%) and the smallest for Seuclidean (0.2%). It should be noted that for Seuclidean, only 2 out of 9 point clouds could be relatively registered (preliminary orientation; Figure 11).

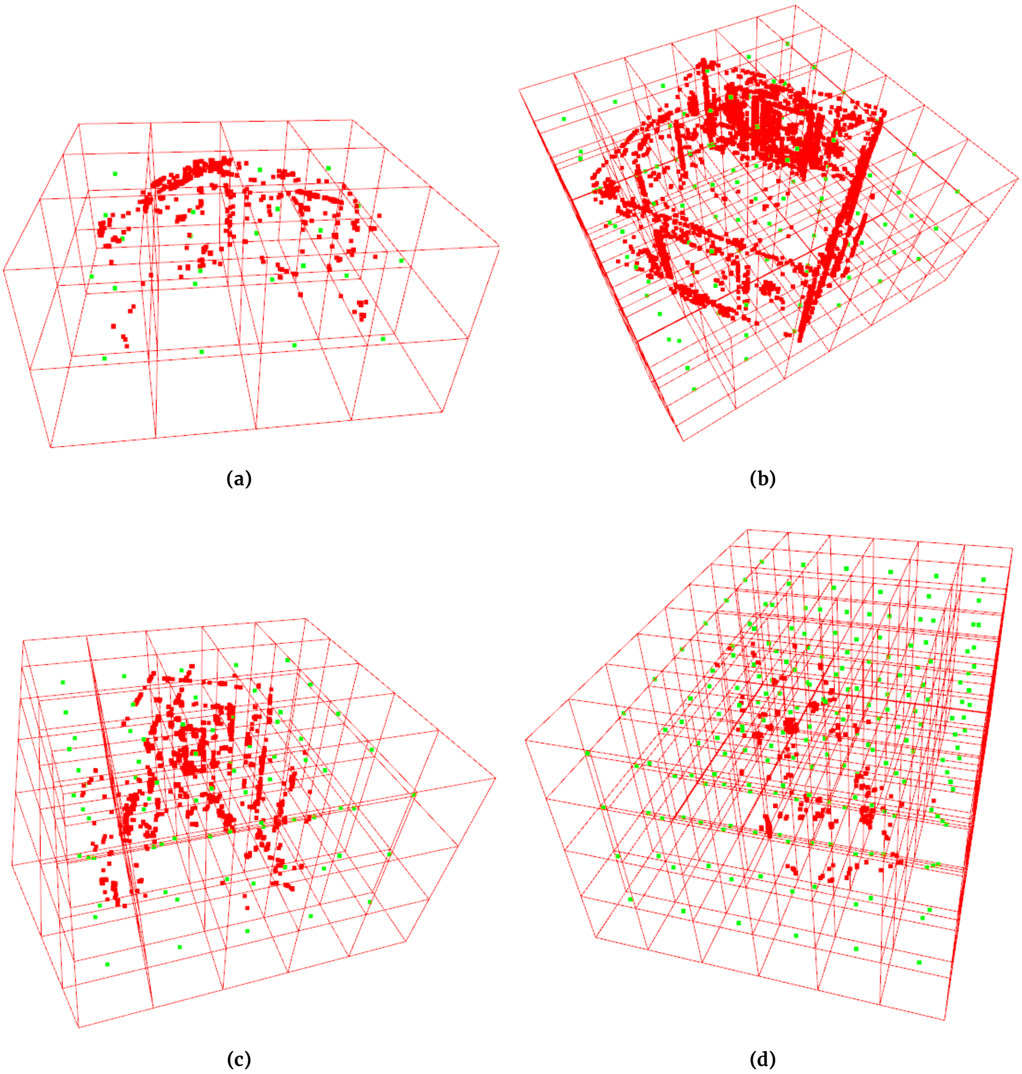



**Figure 17.** Point clouds with defined Octree for: (a) Test Site I, (b) Test Site II, (c) Test Sit III and (d) Test Site IV

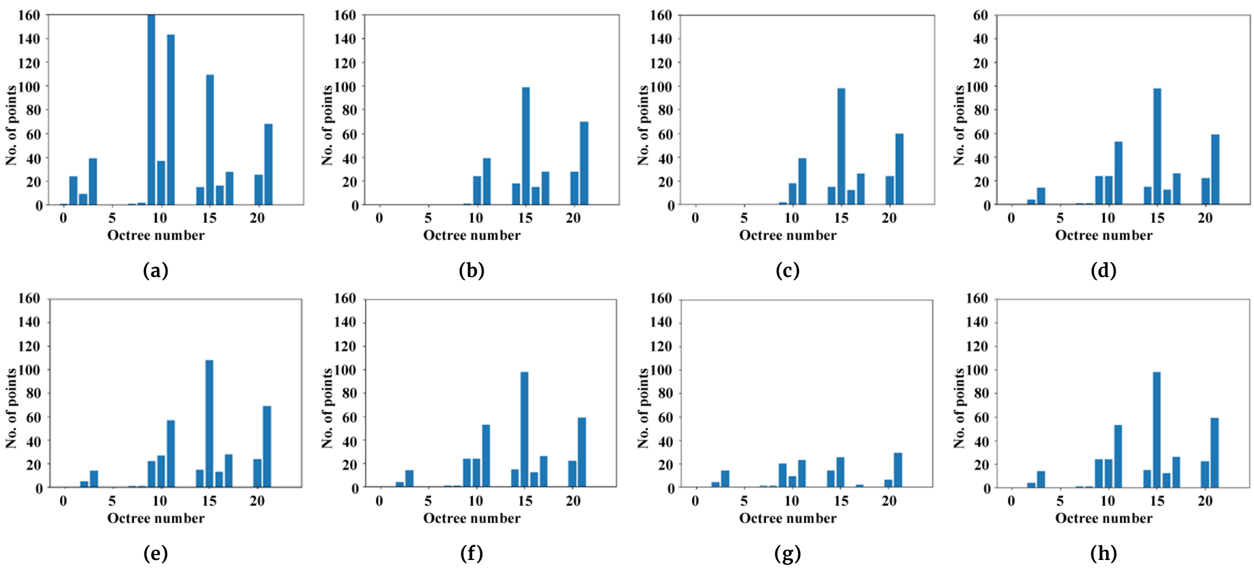



**Figure 18.** The bar chart of the points number in individual octree cubes for: (a) Barycurtis, (b) Canberra, (c) Correlation, (d) Cosine, (e) $L_1$, (f) $L_2$, (g) Seuclidean and (h) Sqeuclidean – Test Site I

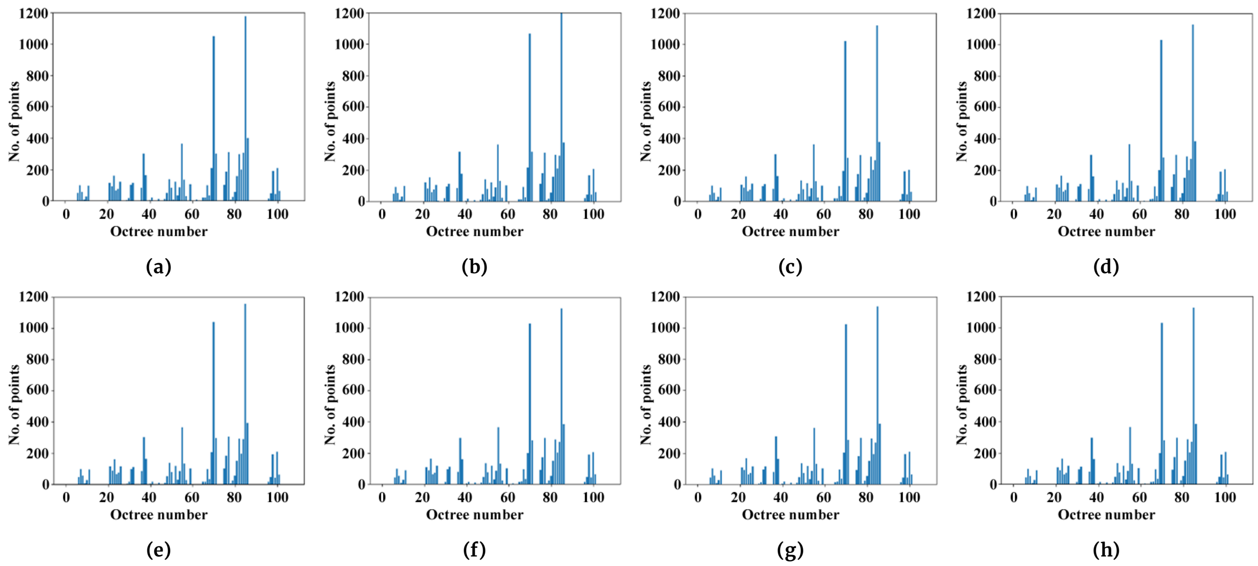



**Figure 19.** The bar chart of the points number in individual octree cubes for: (a) Barycurtis, (b) Canberra, (c) Correlation, (d) Cosine, (e) $L_1$, (f) $L_2$, (g) Seuclidean and (h) Sqeuclidean – Test Site II

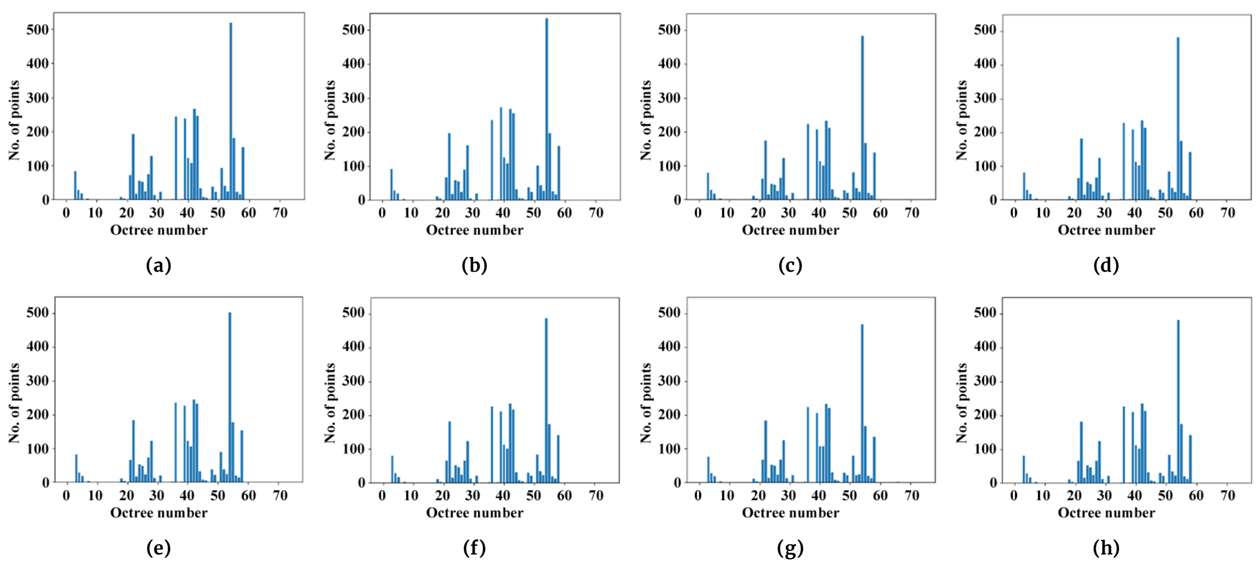



**Figure 20.** The bar chart of the points number in individual octree cubes for: (a) Barycurtis, (b) Canberra, (c) Correlation, (d) Cosine, (e) $L_1$, (f) $L_2$, (g) Seuclidean and (h) Sqeuclidean – Test Site III

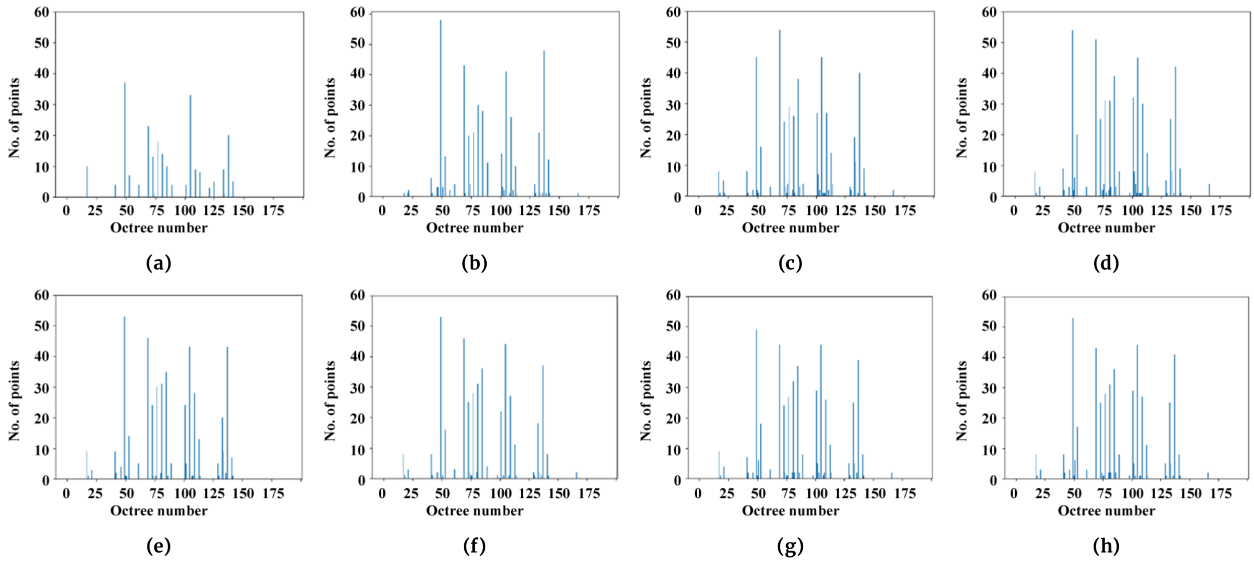



**Figure 21.** The bar chart of the points number in individual octree cubes for: (a) Barycurtis, (b) Canberra, (c) Correlation, (d) Cosine, (e) $L_1$, (f) $L_2$, (g) Seuclidean and (h) Sqeuclidean – Test Site IV

**Table 2.** The statistical analysis of pairwise TLS point cloud registration for all the Test Sites

| Test Site | Method | Accuracy [%] | | | Precision [%] | | | Recall [%] | | | F1 [%] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Min | Median | Max | Min | Median | Max | Min | Median | Max | Min | Median | Max |
| I | Braycurtis | 98.5 | 99.3 | 99.7 | 35.9 | 67.9 | 95.4 | 38.2 | 65.2 | 89.4 | 38.9 | 62.5 | 83.8 |
| | Canberra | 98.5 | 98.9 | 99.8 | 55.3 | 68.4 | 94.6 | 37.8 | 67.7 | 83.3 | 54.0 | 62.5 | 83.3 |
| | Correlation | 98.2 | 98.2 | 99.7 | 36.0 | 60.0 | 85.8 | 34.3 | 59.0 | 75.4 | 39.6 | 56.6 | 74.6 |
| | Cosine | 98.2 | 98.9 | 99.7 | 35.9 | 63.0 | 88.8 | 34.3 | 58.9 | 76.8 | 39.6 | 56.2 | 76.8 |
| | $L_1$ | 98.5 | 98.9 | 99.7 | 46.5 | 62.8 | 91.7 | 81.1 | 64.6 | 81.2 | 51.2 | 61.9 | 81.2 |
| | $L_2$ | 98.2 | 98.9 | 99.7 | 36.0 | 62.8 | 88.8 | 35.5 | 59.0 | 76.3 | 39.6 | 56.2 | 76.3 |
| | Seuclidean | 98.3 | 98.4 | 98.5 | 61.4 | 64.3 | 67.2 | 67.2 | 70.9 | 74.5 | 67.2 | 67.2 | 67.3 |
| | Sqeuclidean | 98.2 | 98.9 | 99.7 | 35.9 | 62.8 | 88.8 | 35.5 | 58.7 | 76.3 | 39.6 | 56.2 | 76.3 |
| II | Braycurtis | 84.7 | 89.9 | 98.8 | 56.2 | 61.7 | 71.6 | 49.8 | 74.7 | 82.7 | 52.8 | 67.6 | 74.5 |
| | Canberra | 85.2 | 89.5 | 94.1 | 56.6 | 60.2 | 72.6 | 52.1 | 72/9 | 82.0 | 55.2 | 65.9 | 75.4 |
| | Correlation | 82.5 | 88.8 | 93.4 | 52.8 | 57.9 | 67.3 | 46.8 | 70.0 | 79.6 | 49.6 | 63.5 | 69.9 |
| | Cosine | 82.9 | 89.1 | 93.4 | 52.9 | 58.9 | 68.1 | 46.9 | 71.0 | 80.9 | 49.7 | 64.6 | 70.8 |
| | $L_1$ | 83.9 | 89.6 | 93.8 | 55.6 | 60.7 | 70.1 | 49.9 | 73.2 | 81.4 | 52.9 | 66.5 | 72.9 |
| | $L_2$ | 82.9 | 89.1 | 93.4 | 52.9 | 59.0 | 68.1 | 46.9 | 71.0 | 80.9 | 49.7 | 64.6 | 70.9 |
| | Seuclidean | 83.6 | 89.2 | 93.5 | 53.4 | 59.2 | 69.5 | 47.4 | 71.8 | 80.4 | 50.2 | 64.9 | 72.3 |
| | Sqeuclidean | 82.9 | 89.1 | 93.4 | 52.9 | 59.0 | 68.1 | 46.9 | 71.0 | 80.9 | 49.7 | 64.6 | 70.8 |
| III | Braycurtis | 80.5 | 96.5 | 97.9 | 3.6 | 17.4 | 56.7 | 4.0 | 30.2 | 91.2 | 3.8 | 20.2 | 69.9 |
| | Canberra | 81.6 | 96.5 | 97.9 | 4.1 | 19.6 | 57.9 | 4.5 | 31.6 | 93.0 | 4.3 | 20.1 | 71.3 |
| | Correlation | 79.4 | 96.5 | 97.9 | 3.0 | 15.7 | 49.8 | 3.4 | 25.3 | 95.2 | 3.2 | 17.9 | 61.4 |
| | Cosine | 79.7 | 96.5 | 97.9 | 3.6 | 15.7 | 51.4 | 4.0 | 25.8 | 92.6 | 3.8 | 18.3 | 63.4 |
| | $L_1$ | 80.0 | 96.5 | 97.9 | 3.6 | 16.7 | 54.6 | 4.0 | 29.0 | 88.6 | 3.8 | 19.9 | 67.3 |
| | $L_2$ | 79.7 | 96.5 | 97.9 | 3.6 | 15.5 | 51.3 | 4.0 | 26.0 | 92.1 | 3.8 | 18.4 | 63.3 |
| | Seuclidean | 79.6 | 96.5 | 97.9 | 2.5 | 15.2 | 51.0 | 2.9 | 26.7 | 93.1 | 2.7 | 18.9 | 62.9 |
| | Sqeuclidean | 79.7 | 96.5 | 97.9 | 3.6 | 15.5 | 51.3 | 4.0 | 26.0 | 92.1 | 3.8 | 18.4 | 63.3 |
| IV | Braycurtis | 95.6 | 98.5 | 99.3 | 16.3 | 46.4 | 85.7 | 5.7 | 17.7 | 62.2 | 9.9 | 26.2 | 67.6 |
| | Canberra | 95.8 | 98.5 | 99.3 | 11.6 | 47.9 | 81.8 | 5.1 | 17.6 | 61.1 | 7.1 | 24.3 | 64.9 |
| | Correlation | 95.3 | 97.9 | 99.3 | 22.5 | 38.0 | 100.0 | 6.6 | 16.5 | 53.6 | 12.2 | 12.2 | 59.2 |
| | Cosine | 95.3 | 98.0 | 99.3 | 21.1 | 40.0 | 100.0 | 6.6 | 17.3 | 58.4 | 12.4 | 25.3 | 64.5 |
| | $L_1$ | 95.5 | 98.3 | 99.3 | 16.3 | 43.3 | 100.0 | 6.6 | 16.4 | 59.5 | 9.9 | 24.6 | 65.7 |
| | $L_2$ | 95.3 | 98.0 | 99.3 | 21.1 | 40.0 | 100.0 | 6.6 | 17.3 | 58.4 | 12.4 | 25.0 | 64.5 |
| | Seuclidean | 95.4 | 98.2 | 99.3 | 11.6 | 41.4 | 71.4 | 4.7 | 16.3 | 56.7 | 7.1 | 22.9 | 62.6 |
| | Sqeuclidean | 95.3 | 98.0 | 99.3 | 21.1 | 40.0 | 100.0 | 6.6 | 17.3 | 58.4 | 12.4 | 25.0 | 64.5 |

- Relatively low Precision values (averaging around 64%) suggest that using the Brute-Force Matching method and the tested similarity metrics contributes to the incorrect matching of tie points, classifying them as False Positives. The spread values differ significantly, with the worst case being 59.5% for Braycurtis and the best at 5.8% for Seuclidean. The second-best distance metric is Canberra, with a Precision value of 39.3%.
- The median Recall values, averaging around 64%, indicate that approximately 36% of all positive cases were not recognised, with about 36% omitted as False Negatives. The spread values are significantly lower than those for Precision, ranging from 40.8% (L2) to 49.5% (Braycurtis), not considering the results obtained for Seuclidean (7.4%).
- Evaluating the F1-score values (the harmonic mean of Precision and Recall), the spread of these values ranged from 29.3% for Canberra to 44.9% for Braycurtis. The highest median value was obtained for Seuclidean at 67.2%, while the lowest was for L2 at 56.2%.

Evaluating the values obtained for Test Site II (characterised by simple geometry but featuring wall paintings that create a three-dimensional effect), it can be observed that there is an average decrease of 8% in the median Accuracy values, similar median values for Precision, and an increase in Recall and F1-score compared to the results for Test Site I. Analysing the individual values, it can be stated that:

- The highest median Accuracy value was obtained for Braycurtis (89.9%), while the lowest was for Correlation (88.8%) – these are insignificant differences. The average spread values were 9.9%, with a minimum of 8.8% for Canberra and a maximum of 10.8% for Correlation.
- Similar to the Precision values obtained for Test Site I, the average median values were around 60% (approximately 64% for Test Site I). The spread between the maximum and minimum values was smaller than that for Test Site I, ranging from 14.4% to 16.0% for Correlation and Seuclidean, respectively.
- Similar to the precision values, the average median Recall value was also 8% higher than that of Test Site I, which stood at 72%. This indicates that for point clouds acquired from greater distances than those for Test Site I and with more minor changes in the heights of the scanner positions, it was possible to detect a more significant number of positive cases, including False Negatives (FN). Analysing the spread values, it can be observed that they range from 29.9% for Canberra to 34.0% for Cosine. The maximum value of 74.5% was obtained for Braycurtis, while the lowest was for Correlation at 70.0%. This is a 12% increase compared to the values obtained for Test Site I.
- For F1-score values, compared to Test Site I, the difference between the minimum (63.5% – Correlation) and maximum (65.3% – Bray-Curtis) values, as well as the spread values (19.9% – $L_1$ and 22.0% – Seuclidean), is relatively small but still observable. This indicates better effectiveness in determining tie points for point clouds acquired over a full angular range. It applies particularly to scanner positions that measure from greater distances and at smaller angles relative to the surface normal, reducing the impact of distortions in spherical images caused by converting point clouds into raster form.

In the case of Test Site III (office room), with scanner positions placed near walls, significantly lower Precision, Recall, and F1-score values were observed compared to Test Sites I and II:

- The average Accuracy values were 96.5%, with the spread ranging from 16.3% for Canberra to 18.5% for Correlation.
- In contrast to the previous Test Sites, the average median values are about four times lower, averaging only 16.4%. The differences between the maximum and minimum values are similar to those obtained for Test Site I, amounting to 46.8% for Correlation and 53.8% for Canberra, respectively.
- Evaluating the Recall results, it should be noted that for point clouds acquired close to walls and for which sections were scanned at acute angles, potential tie points were incorrectly detected during descriptor matching using various distance metrics. This resulted in instability, as seen in significant differences between the minimum and maximum values (84.6–91.7%) and the average median value (27.6%) across all the metrics. This instability may also have been influenced by the low texture quality of the object, which made the descriptors repetitive and thus impacted the accuracy of descriptor matching.
- Evaluating the F1-score values reveals an average threefold decrease compared to Test Sites I and II and reduced stability in point detection (values ranging from 58.2% to 67.1%). This is due to the placement of scanner positions relative to the walls and the characteristics of the Test Field itself, which contributes to the distortion effect on spherical images. This relationship was demonstrated in analyses of synthetic data. The best results (highest median F1-score) were obtained for Braycurtis at 20.2%, while the worst ones were for Correlation at 17.9%.

Test Site IV (an empty shop in a shopping mall) was characterised by the flat, textureless surfaces of the measured site and had characteristics similar to those of Test Site III. The difference lies in the scanner position and the chosen scanning resolution. In the case of Test Site III (as mentioned earlier), the scanner positions were placed close to the walls. Due to Test Site IV's dimensions, determining the positions farther from the walls was possible, resulting in less significant distortion in the spherical images compared to Test Site III. This is seen in the values presented in Table 2:

- The average median Accuracy values are 1.7% higher than for Test Site III, with 98.5% for Braycurtis and 97.9% for Correlation. The spread values are approximately 4.7 times as small, ranging from 3.5% (Canberra) to 4.0% (Cosine).
- For Precision values, the average median values are approximately 2.5 times as high as for Test Site III, with minimum and maximum values of 38.0% for Correlation and 47.9% for Canberra, respectively. The spread for Test Site IV is more significant than for Test Site III's, ranging from 59.8% for Seuclidean to 83.7% for $L_1$.
- Assessing the Recall values shows lower values than for Test Site III. This indicates that many points are incorrectly matched as False Negatives. The average median value is 17.1% across all the distance metrics, with individual methods ranging from 16.3% for Seuclidean to 17.5% for Braycurtis. Evaluating the spread from minimum to maximum values reveals that it is, on average, about 36% smaller, amounting to 47.0% and 56.5% for Correlation and Braycurtis, respectively.
- For F1-score values, a slight improvement of around 8% is observed compared to the results obtained for Test Site III. Analysing the individual values shows that the minimum and maximum median F1-score values were obtained for Seuclidean (22.8%) and Braycurtis (26.1%). However, it should be emphasised that these values are low and indicate ineffective detection and accurate matching of tie points. There are many points for which counterparts in the registered point cloud were not found, or these values were incorrectly matched. This is also evident in the spread values between the maximum and minimum values, which are 47.0% for Correlation and 57.8% for Canberra.

In summary, when selecting the best distance metric based on the values of Accuracy, Precision, Recall, and F1-score (without considering the number of points, their distribution, or the correctness of pairwise registration), two cases of point clouds should be considered separately: (1) point clouds of objects with good texture (significant differences in grayscale gradient values) and numerous architectural details (Test Sites I and II), and (2) point clouds of objects characterised by simple geometry and textureless surfaces (Test Sites III and IV).

- Comparing both cases, it can be observed that F1-score values are significantly higher for case 1 than case 2, although they still remain low. Assessing the Accuracy, Recall, and Precision values for case 1, it can be concluded that most points are classified as negative cases (a high percentage of True Negatives), which increases Accuracy but does not reflect the effectiveness in detecting all possible tie points. Precision and Recall indicators are not high due to False Positives and False Negatives, indicating that there may be difficulties in recognising the positive class during descriptor matching. Using spherical images at full scanning resolution (360° – Test Site II) increases the values of all the statistical metrics. After analysing the results presented for case 1, it is recommended that the $L_1$-norm be used.
- Evaluating the values for case 2, it should be emphasised that all indicator values are low. This is due to minimal changes in grayscale gradient values and incorrect descriptor matching when using various distance metrics. Another factor contributing to the lower Precision values is the distance between TLS positions and measured object sections and the acute angles relative to scanning to the normal vectors of measured surfaces, leading to significant spherical image distortion (Test Site III). For this reason, when using this method for tie point detection, one should bear in mind that many points may not be correctly matched. For the alignment of point clouds with poor texture and few architectural details, it is recommended to use the Braycurtis distance metric.

## 5 Discussion and summary

The research focused on examining the influence of selecting distance metrics during the descriptor matching stage in the TLS-SfM method on the accuracy of point cloud registration. For the analysis, commonly used distance metrics were selected, namely Barycurtis, Canberra, Correlation, Coine, $L_1$, $L_2$, Seuclidean and Sqeuclidean, which are utilised in machine learning.

Initial tests on synthetic data revealed challenges in detecting and matching keypoints on point clouds converted to the raster form, with major issues emerging from input data quality and effects of 3D-to-2D transformations, particularly in the TLS-SfM process. These issues include intensity deviations caused by measurement angles relative to the surface normal vector, scanning distance, material properties, and substantial geometric distortions at high angles, significantly affecting the upper and lower sections of rasters. These distortions impact the number and spatial distribution of detected tie points, necessitating simulations that apply Gaussian noise to represent intensity variation and virtual images to replicate radial distortion, rotation, and tilt. The analysis indicated that image rotation angles and skew greatly influenced tie-point accuracy, whereas intensity deviations had minimal effect. Further examination showed that raster fragments scanned at high angles or close distances could exhibit distortions and rotations, underscoring the need for joint consideration of these factors. Using the Canberra measure resulted in nearly perfect descriptor matching accuracy, suggesting that minor rotations are negligible when employing this measure, with other measures also achieving over 95% accuracy. These results highlight the importance of countering geometric distortion and rotation to ensure reliable keypoint detection, especially when scanning at close range or acute angles.

The experiments on real data were conducted at two types of test sites located in cultural heritage buildings and public utility facilities, allowing an independent analysis of data characterised by various architectural details, colours, and complex geometric features. Using these two types of test fields enabled an independent study of the impact of distance measurement selection on the accuracy of point cloud registration and the completeness of the entire process. For this reason, each type of test site was evaluated separately, and Table 3 presents a summary of the usefulness evaluation criteria, rated on a scale from 1 to 8. The individual observations were weighted to assess and differentiate the impact of individual components on the final evaluation and selection of the best distance metrics. The experiments resulted in the following conclusions:

- Research indicates that, despite the low values obtained for F1-score, Accuracy, Precision, and Recall, using the TLS-SfM method enables the correct orientation of point clouds. The achieved registration accuracy (comparable to the commonly used Target-based method), point distribution (lower impact), and pairwise registration allow complete TLS data registration.
- A two-stage approach could solve the issue of low F1-score, accuracy, precision, and recall. In the first stage, it is recommended to use the pairwise registration method to obtain the orientation parameters of the point clouds. In the next stage, using the k-NN method, the remaining tie points with correspondences in both datasets should be identified, and a final bundle adjustment should be performed based on them.
- A comprehensive analysis of the results (without considering the division into specific types of test fields characterised by varying geometric complexity, different textures, and scanner position distributions) indicates that the best results were obtained with the $L_1$ and Sqeuclidean methods. At the same time, the worst result was achieved with the Canberra method.
- Evaluating the differences in the total sum of points, it can be observed that more minor differences are noticeable for test fields with short baselines between TLS positions (Test Sites II and III) compared to those with larger baselines (Test Sites I and IV).
- Test Site I (the interior of a historical building) was characterised by complex geometry, many gilded elements, and architectural details, which contributed to a well-defined texture in the generated spherical images. The registered point clouds were captured at significantly different heights (with height differences between pairs of point clouds ranging from 0.1 m to 2.8 m), and the same sections were scanned at notably different, acute angles relative to the surface normal, resulting in various image distortions (not in full 360° resolution). For this scanner setup and test field type, the overall best results were obtained using the $L_1$-norm and Sqeuclidean methods, while the worst were achieved with Sececlidean. It should be noted, however, that only the Braycurtis method enabled full registration of all point clouds. However, with this method, the registration error was approximately three times as high as with the state-of-the-art Target-based approach using signalised check points. When evaluating the error values for the other approaches, it can be stated that the RMSE values were approximately twice as high as those for the target-based approach. However, they did not exceed 5 mm, which makes this solution acceptable.
- Test Site II, like Test Site I, is the interior of a cultural heritage object; however, unlike Test Site I, it features simple geometry. The walls were decorated with paintings that imitate a spatial illusion effect, and the point clouds were captured at similar heights. The point clouds were acquired in full angular resolution, resulting in only selected sections being scanned from close distances (with acute angles relative to the wall's normal plane vector). In contrast, the remaining areas were scanned from significantly longer distances. This setup led to notable distortions only in certain sections of the generated spherical images. This relationship results in only slight differences in the outcomes of point cloud alignment when using different distance metrics. All the methods allowed the process to be conducted accurately for pairwise and full registration. The differences are noticeable only in the RMSE values. For all the methods except Braycurtis, RMSE values did not exceed 2.5 mm, while for Braycurtis, RMSE was approximately 3.0 mm, compared to about 1.5 mm for the Target-based approach. When processing such interiors, one

**Table 3.** Summary of the evaluation criteria for point detection algorithms across all the Test Sites, rated on a scale from 1 to 8

| Test Site | Method | Evaluation Criteria: Completeness of pairwise registration | Completeness of full registration | RMSE on signalised check points | Deviations on marked check points – X-axis | Deviations on marked check points – Y-axis | Deviations on marked check points – Z-axis | Point distribution | ML statistical analysis | Total | Final Ranking |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Weight | 3 | 5 | 4 | 3 | 3 | 3 | 3 | 1 | 25 | |
| **Cultural Heritage Interiors** | | | | | | | | | | | |
| I | Braycurtis | 3 | 6 | 5 | 3 | 3 | 3 | 8 | 2 | 4.5 | VI |
| I | Canberra | 4 | 4 | 8 | 5 | 7 | 7 | 5 | 4 | 5.6 | III |
| I | Correlation | 3 | 3 | 7 | 7 | 7 | 7 | 5 | 2 | 5.3 | IV |
| I | Cosine | 2 | 3 | 7 | 6 | 5 | 5 | 5 | 2 | 4.6 | V |
| I | $L_1$ | 5 | 5 | 7 | 7 | 8 | 8 | 5 | 4 | 6.2 | I |
| I | $L_2$ | 2 | 3 | 7 | 6 | 7 | 7 | 5 | 2 | 5.0 | VII |
| I | Seuclidean | 1 | 1 | 8 | 8 | 6 | 6 | 1 | 2 | 4.2 | VIII |
| I | Sqeuclidean | 4 | 5 | 7 | 6 | 7 | 7 | 5 | 3 | 5.7 | II |
| II | Braycurtis | 8 | 8 | 6 | 5 | 8 | 8 | 6 | 4 | 6.9 | V |
| II | Canberra | 8 | 8 | 7 | 5 | 6 | 6 | 7 | 4 | 6.7 | VI |
| II | Correlation | 8 | 8 | 7 | 8 | 7 | 7 | 6 | 4 | 7.2 | II |
| II | Cosine | 8 | 8 | 7 | 6 | 7 | 7 | 6 | 4 | 7.0 | IV |
| II | $L_1$ | 8 | 8 | 7 | 8 | 7 | 7 | 6 | 4 | 7.2 | II |
| II | $L_2$ | 8 | 8 | 7 | 7 | 5 | 5 | 6 | 4 | 6.6 | VII |
| II | Seuclidean | 8 | 8 | 7 | 7 | 8 | 8 | 6 | 4 | 7.3 | I |
| II | Sqeuclidean | 8 | 8 | 7 | 7 | 5 | 5 | 6 | 4 | 6.6 | VIII |
| **Public Utilities Interiors** | | | | | | | | | | | |
| III | Braycurtis | 7 | 8 | 7 | 6 | 6 | 7 | 7 | 4 | 6.8 | V |
| III | Canberra | 5 | 8 | 7 | 6 | 6 | 6 | 6 | 4 | 6.4 | VIII |
| III | Correlation | 6 | 8 | 8 | 6 | 6 | 7 | 7 | 3 | 6.8 | V |
| III | Cosine | 6 | 8 | 8 | 6 | 7 | 7 | 7 | 3 | 7.0 | IV |
| III | $L_1$ | 8 | 8 | 8 | 8 | 8 | 7 | 7 | 3 | 7.6 | I |
| III | $L_2$ | 6 | 8 | 8 | 7 | 7 | 7 | 7 | 3 | 7.1 | II |
| III | Seuclidean | 7 | 8 | 7 | 7 | 6 | 6 | 6 | 3 | 6.7 | VII |
| III | Sqeuclidean | 6 | 8 | 8 | 6 | 8 | 7 | 7 | 3 | 7.1 | II |
| IV | Braycurtis | 5 | 6 | 4 | 3 | 6 | 5 | 5 | 4 | 4.9 | VII |
| IV | Canberra | 3 | 3 | 3 | 5 | 6 | 5 | 5 | 4 | 4.1 | VIII |
| IV | Correlation | 4 | 5 | 4 | 4 | 7 | 7 | 7 | 2 | 5.2 | VI |
| IV | Cosine | 7 | 6 | 5 | 4 | 7 | 7 | 7 | 4 | 6.0 | III |
| IV | $L_1$ | 8 | 8 | 4 | 3 | 6 | 5 | 5 | 4 | 5.6 | V |
| IV | $L_2$ | 6 | 6 | 6 | 4 | 7 | 7 | 7 | 4 | 6.0 | III |
| IV | Seuclidean | 7 | 7 | 3 | 3 | 8 | 8 | 8 | 4 | 6.1 | II |
| IV | Sqeuclidean | 7 | 7 | 6 | 4 | 7 | 7 | 7 | 4 | 6.4 | I |

could conclude that the choice of distance metrics does not significantly impact the accuracy and completeness of TLS point cloud registration.

- Test Site III is an office space with smooth-textured walls, ceiling-mounted lamps, electrical wiring, and a dark carpeted floor. Due to the office dimensions, the scanner positions were placed close to all walls. This resulted in acute scanning angles relative to the walls' normal planes and significant changes in intensity and distortions in the spherical images. However, the differences in the total scores for the individual distance metrics are negligible due to the short scanning distances and full angular resolution, similar to Test Site II. Analysing pairwise registration, it can be observed that the difference in the number of correctly matched values between the worst (Canberra) and the best (Correlation and Euclidean) metrics is only 3 pairs. This does not affect the completeness of the full registration. By evaluating the RMSE values linearly, it can be seen that for all approaches, except for Barycurtis and Canberra, the RMSE values do not exceed 2 mm. At the same time, for the two metrics mentioned above, it is 2.5 mm. Comparing this with the results from Target-based, which is 1 mm, it can be concluded that using all of these distance metrics, it is possible to perform fully automated TLS point cloud registration correctly.
- Test Site VI was the "Empty Shopping Mall" consisting of smooth walls, a concrete floor, overhead lighting, visible electric wires, and an air-conditioning system. Due to the dimensions and shape of the room, the TLS positions were placed further from the walls, and there were relatively long baselines between the individual positions. All point clouds were acquired at approximately the same height. It should be emphasised, however, that the scanning resolution was twice as low as that for the point clouds obtained from Test Sites II and III. Selecting the appropriate distance metric is crucial for this arrangement of TLS stations and the type of object being measured. It influences the number of correctly matched pairwise registrations. The best results were obtained with the $L_1$-norm, Sqeuclidean, $L_2$, Seuclidean, and Cosine metrics, which directly impact the quality and accuracy of the full registration. This is evident in the linear RMSE error values, which for $L_2$-norm, Cosine, and Sqeuclidean did not exceed 11 mm, being approximately twice as high as those for the Target-based method. However, considering the adopted scanning resolution (12.1 mm / 10 m), it can be concluded that the process was correctly done for the methods mentioned above.
- In summary, based on the values obtained for the two groups of point clouds – cultural heritage objects characterised by good texture (with significant grayscale gradient variations) and numerous architectural details and public utility objects with simple geometry and textureless surfaces – it is recommended to use $L_1$-norm and Sqeuclidean distance metrics, respectively, during pairwise registration.

Future research will focus on analysing the selection of distance metrics for point cloud alignment performed outdoors, as well as examining the impact of descriptor selection on the accuracy and completeness of TLS point cloud registration using the TLS-SfM method.

## Acknowledgements

This paper was co-financed under the research grant of Warsaw University of Technology supporting scientific activity in the discipline of Civil Engineering and Transport.

## References

Abbate, E., Sammartano, G., and Spanò, A. (2019). Prospective upon multi-source urban scale data for 3D documentation and monitoring of urban legacies. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W11:11–19, doi:10.5194/isprs-archives-xlii-2-w11-11-2019.

Aggarwal, C. C. et al. (2015). *Data mining: The textbook*, volume 1. Springer.

Alba, M., Barazzetti, L., Scaioni, M., and Remondino, F. (2012). Automatic registration of multiple laser scans using panoramic rgb and intensity images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII-5/W12:49–54, doi:10.5194/isprsarchives-xxxviii-5-w12-49-2011.

Arif, R. and Essa, K. (2017). Evolving techniques of documentation of a world heritage site in Lahore. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W5:33–40, doi:10.5194/isprs-archives-xlii-2-w5-33-2017.

Bay, H., Tuytelaars, T., and Van Gool, L. (2006). *SURF: Speeded Up Robust Features*, pages 404–417. Springer Berlin Heidelberg, doi:10.1007/11744023_32.

Berenyi, A., Lovas, T., and Barsi, A. (2010). Terrestrial Laser Scanning – Civil engineering applications. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38(Part 5):80–85.

Bianco, S., Ciocca, G., and Marelli, D. (2018). Evaluating the performance of Structure from Motion pipelines. *Journal of Imaging*, 4(8):98, doi:10.3390/jimaging4080098.

Cheng, L., Chen, S., Liu, X., Xu, H., Wu, Y., Li, M., and Chen, Y. (2018). Registration of laser scanning point clouds: A review. *Sensors*, 18(5):1641, doi:10.3390/s18051641.

Cipriani, L., Bertacchi, S., and Bertacchi, G. (2019). An optimised workflow for the interactive experience with cultural heritage through reality-based 3d models: Cases study in archaeological and urban complexes. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W11:427–434, doi:10.5194/isprs-archives-xlii-2-w11-427-2019.

Dong, Z., Yang, B., Liang, F., Huang, R., and Scherer, S. (2018). Hierarchical registration of unordered TLS point clouds based on binary shape context descriptor. *ISPRS Journal of Photogrammetry and Remote Sensing*, 144:61–79, doi:10.1016/j.isprsjprs.2018.06.018.

Giżyńska, J., Komorowska, E., and Kowalczyk, M. (2022). The comparison of photogrammetric and terrestrial laser scanning methods in the documentation of small cultural heritage object – case study. *Journal of Modern Technologies for Cultural Heritage Preservation*, 1(1), doi:10.33687/jmtchp.001.01.0013.

Grussenmeyer, P., Alby, E., Landes, T., Koehl, M., Guillemin, S., Hullo, J. F., Assali, P., and Smigiel, E. (2012). Recording approach of heritage sites based on merging point clouds from high resolution photogrammetry and terrestrial laser scanning. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B5:553–558, doi:10.5194/isprsarchives-xxxix-b5-553-2012.

Han, X.-F., Sun, S.-J., Song, X.-Y., and Xiao, G.-Q. (2018). 3D point cloud descriptors in hand-crafted and deep learning age: State-of-the-art. *arXiv preprint arXiv:1802.02297*.

Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *Procedings of the Alvey Vision Conference 1988*, AVC 1988. Alvey Vision Club, doi:10.5244/c.2.23.

Janßen, J., Kuhlmann, H., and Holst, C. (2023). Keypoint-based registration of tls point clouds using a statistical matching approach. *Journal of Applied Geodesy*, 18(2):267–284, doi:10.1515/jag-2022-0058.

Kang, Z., Li, J., Zhang, L., Zhao, Q., and Zlatanova, S. (2009). Automatic registration of terrestrial laser scanning point clouds using panoramic reflectance images. *Sensors*, 9(4):2621–2646, doi:10.3390/s90402621.

Karwel, A. K. and Markiewicz, J. (2022). The methodology of the archival aerial image orientation based on the SfM method. *Sensors and Machine Learning Applications*, 1(2), doi:10.55627/smla.001.02.0015.

Kowalska, M. and Kowalczyk, M. (2024). Modelling of the exterior surface of a historic and contemporary object by digital photos and laser scanning. *Journal of Modern Technologies for Cultural Heritage Preservation*, 3(3), doi:10.69534/jmtchp/193893.

Kuzyk, Z. (2023). The use of modern measurement methods in the inventory of endangered cultural heritage objects in lviv. *Journal of Modern Technologies for Cultural Heritage Preservation*, 2(2), doi:10.69534/jmtchp/191347.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, doi:10.1023/b:visi.0000029664.99615.94.

Markiewicz, J. (2024). Evaluation of 2d affine — hand-crafted detectors for feature-based tls point cloud registration. *Reports on Geodesy and Geoinformatics*, 117(1):69–88, doi:10.2478/rgg-2024-0008.

Markiewicz, J., Kot, P., Markiewicz, L., and Muradov, M. (2023). The evaluation of hand-crafted and learned-based features in Terrestrial Laser Scanning-Structure-from-Motion (TLS-SfM) indoor point cloud registration: the case study of cultural heritage objects and public interiors. *Heritage Science*, 11(1), doi:10.1186/s40494-023-01099-9.

Markiewicz, J., Podlasiak, P., and Zawieska, D. (2015). A new approach to the generation of orthoimages of cultural heritage objects—integrating tls and image data. *Remote Sensing*, 7(12):16963–16985, doi:10.3390/rs71215869.

Markiewicz, J. and Zawieska, D. (2019). The influence of the cartographic transformation of TLS data on the quality of the automatic registration. *Applied Sciences*, 9(3):509, doi:10.3390/app9030509.

math.net (2024). Distance metrics, https://numerics.mathdotnet.com/Distance.html.

Moussa, W., Abdel-Wahab, M., and Fritsch, D. (2012). An automatic procedure for combining digital images and laser scanner data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B5:229–234, doi:10.5194/isprsarchives-xxxix-b5-229-2012.

Mukupa, W., Roberts, G. W., Hancock, C. M., and Al-Manasir, K. (2016). A review of the use of terrestrial laser scanning application for change detection and deformation monitoring of structures. *Survey Review*, pages 1–18, doi:10.1080/00396265.2015.1133039.

NumPy (2024). Numpy – the fundamental package for scientific computing with python, https://numpy.org/.

OpenCV (2018). http://opencv.org/.

Piermattei, L., Karel, W., Wang, D., Wieser, M., Mokroš, M., Surový, P., Koreň, M., Tomaštík, J., Pfeifer, N., and Hollaus, M. (2019). Terrestrial structure from motion photogrammetry for deriving forest inventory data. *Remote Sensing*, 11(8):950, doi:10.3390/rs11080950.

Rashidi, M., Mohammadi, M., Sadeghlou Kivi, S., Abdolvand, M. M., Truong-Hong, L., and Samali, B. (2020). A decade of modern bridge monitoring using Terrestrial Laser Scanning: Review and future directions. *Remote Sensing*, 12(22):3796, doi:10.3390/rs12223796.

Remondino, F. and Stylianidis, E. (2016). *3D recording, documentation and management of cultural heritage*, volume 2. Whittles Publishing Dunbeath, UK.

SciPy (2024). Scipy – fundamental algorithms for scientific computing in python, https://scipy.org/.

Tobiasz, A., Markiewicz, J., Łapiński, S., Nikel, J., Kot, P., and Muradov, M. (2019). Review of methods for documentation, management, and sustainability of cultural heritage. case study: Museum of king jan iii's palace at wilanów. *Sustainability*, 11(24):7046, doi:10.3390/su11247046.

Tola, E., Lepetit, V., and Fua, P. (2010). DAISY: An efficient dense descriptor applied to wide-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5):815–830, doi:10.1109/tpami.2009.77.

Tuytelaars, T. and Mikolajczyk, K. (2007). Local invariant feature detectors: A survey. *Foundations and Trends® in Computer Graphics and Vision*, 3(3):177–280, doi:10.1561/0600000017.

Urban, S. and Weinmann, M. (2015). Finding a good feature detector-descriptor combination for the 2D keypoint-based registration of TLS point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-3/W5:121–128, doi:10.5194/isprsannals-ii-3-w5-121-2015.

Yu, G. and Morel, J.-M. (2011). Asift: An algorithm for fully affine invariant comparison. *Image Processing On Line*, 1:11–38, doi:10.5201/ipol.2011.my-asift.