Keypoint-based 4-Points Congruent Sets – automated marker-less registration of laser scans

Pascal Willy Theiler, Jan Dirk Wegner, and Konrad Schindler[a]

[a]*Photogrammetry and Remote Sensing group, Institute of Geodesy and Photogrammetry, ETH Zurich, 8093 Zurich, Switzerland, e-mail: {pascal.theiler, jan.wegner, schindler}@geod.baug.ethz.ch*

Abstract

We propose a method to automatically register two point clouds acquired with a terrestrial laser scanner *without* placing any *markers* in the scene. What makes this task challenging are the strongly varying point densities caused by the line-of-sight measurement principle, and the huge amount of data. The first property leads to low point densities in potential overlap areas with scans taken from different viewpoints while the latter calls for highly efficient methods in terms of runtime and memory requirements.

A crucial yet largely unsolved step is the initial coarse alignment of two scans without any simplifying assumptions, that is, point clouds are given in arbitrary local coordinates and no knowledge about their relative orientation is available. Once coarse alignment has been solved, scans can easily be fine-registered with standard methods like least-squares surface or *Iterative Closest Point* matching. In order to drastically thin out the original point clouds while retaining characteristic features, we resort to extracting 3D keypoints. Such clouds of keypoints, which can be viewed as a sparse but nevertheless discriminative representation of the original scans, are then used as input to a very efficient matching method originally developed in computer graphics, called *4-Points Congruent Sets* (4PCS) algorithm. We adapt the 4PCS matching approach to better suit the characteristics of laser scans.

The resulting *Keypoint-based 4-Points Congruent Sets* (*K-4PCS*) method is extensively evaluated on challenging indoor and outdoor scans. Beyond the evaluation on real terrestrial laser scans, we also perform experiments with simulated indoor scenes, paying particular attention to the sensitivity of the approach with respect to highly symmetric scenes.

Preprint submitted to ISPRS Journal of Photogrammetry and Remote Sensing October 27, 2015

Keywords:
point cloud registration, terrestrial laser scanning, 3D keypoint extraction, congruent point sets, geometric matching

1. Introduction

Airborne, mobile, and static terrestrial laser scanners are standard devices to acquire 3D data for a wide range of applications. In this work we deal with LiDAR point clouds of static terrestrial laser scanners (TLS) that are commonly used in surveying, archeology, manufacturing etc. Typically, multiple scans from different viewpoints are needed to fully cover large outdoor objects or complex indoor facilities in full detail. A prerequisite for any further processing of such data, like semantic classification or complete 3D reconstruction, is the *relative orientation* of all scans. The registration puts all scans into a common coordinate frame, thus assembling the scanned object of interest from the individual scans.

The industry standard at present is to place artificial markers in the scene during the measurement campaign. Corresponding markers (or targets) that are visible in at least two scans are then extracted either manually or automatically (e.g., Akca, 2003; Franaszek et al., 2009) to determine the relative orientation of the scans. This procedure is rather time-consuming, markers must remain stable during the measurement campaign and they should be distributed in such a way that no ill-defined geometric constellations arise. In addition, they inevitably occlude small parts of the scene and they usually have to be removed during post-processing because they are not acceptable in the final product. To circumvent artificial markers altogether, quite some effort has been spent on finding fully automated marker-less methods for LiDAR point cloud registration. In contrast to point clouds computed via dense matching of images, the scale is inherently known in LiDAR point clouds, since the sensor directly measures distances. Therefore, registration of LiDAR point clouds can be solved with a rigid-body transformation with six degrees of freedom. Typically, such transformation parameters are estimated in a two-step procedure: An initial coarse registration roughly aligns scans with a precision that avoids the following fine registration to get stuck in a local minimum. It turns out that the first step (coarse registration) is much harder than the second (fine registration).

Various solutions for *fine registration* of scans exist. The most common approach today certainly is the *Iterative Closest Point* (ICP) algorithm in-

troduced in the seminal works of Besl and McKay (1992), Chen and Medioni (1992) and Zhang (1994), and since then refined in different ways (e.g., Bergevin et al., 1996; Bae and Lichti, 2004; Minguez et al., 2006; Censi, 2008). The basic principle of ICP is to minimize the Euclidean distances between nearby points. ICP solves this problem iteratively, by first establishing point-to-point correspondences, and second estimating the transformation parameters. After applying the transformation to all points, point correspondences are sought again and a new, refined set of transformation parameters is estimated. This procedure is then repeated until convergence.

A general property of all such methods is that they involve a non-convex objective function which is optimized locally. Clearly, fine registration without a sufficiently precise coarse alignment is therefore prone to get stuck in local minima. More precisely, if raw scans are not well aligned initially, a sufficiently precise transformation into a common reference frame can hardly be found because the convergence basin of the non-convex objective function is too small. Please refer to Pottmann et al. (2006); Bae (2009) for comprehensive investigations on the convergence properties of such techniques. An idea that naturally comes into mind is thus to first coarsely align both raw scans so that this initial, rough solution is inside the ICP convergence basin. ICP then takes over and accomplishes fine-registration.

Here, we follow this line of thought and propose a fully automated, marker-less, *coarse registration* approach, where the coarsely aligned point clouds serve as input to standard ICP. What makes this task challenging are *(i)* the huge amount of data (millions of points), which calls for computationally efficient techniques, *(ii)* the typically large baselines between adjacent scanning viewpoints (to limit time and costs in the field) and *(iii)* the quadratic decrease of the point density with distance from the scanner (due to the angular sampling of scanning LiDAR devices). The latter causes different point densities on the same surfaces in different scans. As we will see later on in this paper, this property calls for custom-designed processing steps: standard approaches from computer vision or computer graphics cannot be applied directly because they usually assume approximately even point distributions within and between different scans.

In this paper we adapt the *4-Points Congruent Sets* (4PCS) approach for coarse registration, proposed by Aiger et al. (2008), to the aforementioned challenges. To keep registration computationally tractable we downsample raw scans and represent them with a sparse cloud of 3D keypoints. We test two different kinds of keypoints, the 3D *Difference-of-Gaussians* (DoG)

detector and the 3D *Harris* corner detector. The first relies on LiDAR intensities, whereas the second fires at distinct geometrical structures. In contrast to heavily downsampling point clouds at random, less aggressive downsampling followed by keypoint extraction better preserves salient features of the scene and thus offers better repeatability across scans. We call the combined method, which utilizes 3D keypoints as input to a (slightly modified) 4PCS algorithm *Keypoint based 4-Points Congruent Sets* (*K-4PCS*). We plan to release the test data and the source code of the proposed method – integrated in the open-source *Point Cloud Library* (PCL, Rusu and Cousins, 2011) – after publication.

2. Related work

Coarse, marker-free registration of point clouds typically follows a two-step strategy. First raw point clouds are reduced to sparse sets of features, and second corresponding features are sought in overlapping areas to estimate transformation parameters that align the point clouds sufficiently well. The result then serves as input to standard fine-registration like ICP (Besl and McKay, 1992). A common approach to feature extraction are 2D keypoints that are well known from image processing and can be applied to either intensity or range images of scans (Böhm and Becker, 2007; Kang et al., 2009). A drawback of pure 2D features is that they cannot cope well with strong viewpoint changes. 3D features are generally more robust (Allaire et al., 2008; Lo and Siebert, 2009; Flitton et al., 2010; Flint et al., 2007) and we will thus use them in our work. Another popular line of thought, particularly in man-made environments, is to derive features from planar surfaces (Brenner and Dold, 2007; Brenner et al., 2008; Theiler and Schindler, 2012). More general strategies that can handle arbitrarily shaped surfaces rely on salient directions (Novák and Schindler, 2013) or on a combination of salient directions and 2D features after an ortho-rectification process (Zeisl et al., 2013). It should be noted that approaches purely based on geometric object properties are limited to scenes of medium complexity due to self-occlusions, which occurr when very complex surfaces are acquired from different viewpoints.

Sampling and matching of extracted features often follows strategies related to RANSAC (Fischler and Bolles, 1981). A popular variant called *Sample Consensus Initial Alignment* was proposed by Rusu et al. (2009). They match corresponding point triplets to align scans, which however implies runtime complexity per sample of $\mathcal{O}(n^3)$.

An elegant method to do coarse and fine registration in one step was proposed recently by Yang et al. (2013). They develop a globally optimal version of ICP coined Go-ICP, which avoids mismatches due to local minima by global optimization with a clever branch-and-bound scheme. It runs efficiently on rather small point clouds of up to a few thousand points, but does not scale up to the typical size of laser scans.

We propose a completely automated method for marker-less coarse registration of TLS point clouds of arbitrary relative orientation based on (i) a sparse representation via discriminative 3D keypoints that are (ii) matched efficiently with a variant of the 4PCS method of Aiger et al. (2008). They showed that the computational burden of point cloud matching can be reduced to $\mathcal{O}(n^2)$, by adding a fourth point in such a way that the four points are roughly coplanar. They achieve high efficiency and success rates if matching uniformly distributed point clouds.

Here, we build upon our preliminary work (Theiler et al., 2013), but give a more in-depth description of the standard method and its extensions, add a second, purely geometric keypoint detector, conduct extensive experiments on synthetic data concerning robustness against repetitive structures, and present much extended experiments on real indoor and outdoor data. Experimental results show that *K-4PCS* reaches high success rates above 95% in moderately difficult scenarios and $\approx 65\%$ success even for extremely challenging scenes, with accuracies easily high enough to initialize standard ICP.

3. Conceptual overview

The proposed method combines ideas from image processing and computational geometry. On the one hand we make use of *Difference-of-Gaussian* (DoG) keypoints comparable to the ones introduced by Lowe (1999) – also known as SIFT keypoints – but adapted to 3D. Alternatively, we use keypoints found by a 3D *Harris* corner detector. The two methods are representative of 3D keypoint extraction: DoG takes into account point intensities, whereas Harris is purely geometry-based (see details in 4). On the other hand, the original 4PCS algorithm of Aiger et al. (2008) is adapted to handle the resulting sparse clouds of keypoints. This algorithm (see 5) is an efficient matching strategy based on corresponding keypoint quadruples and random sampling. Fig. 1 shows the overall workflow of the algorithm, which goes as follows:

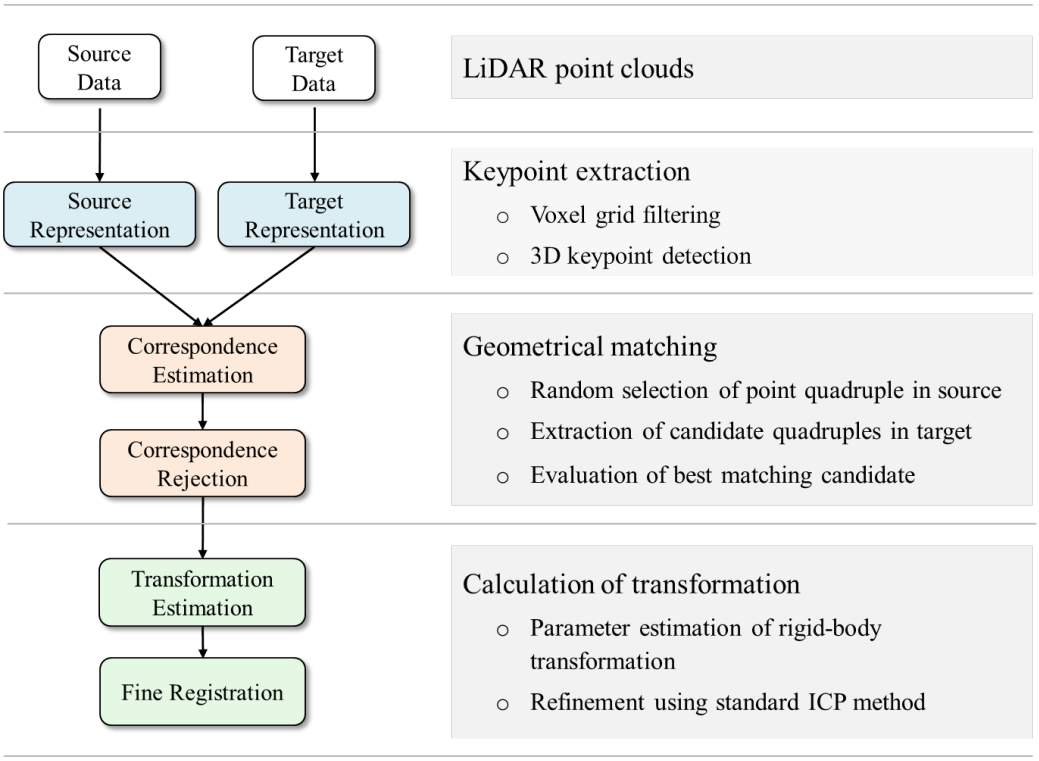

Figure 1: General workflow of the proposed *K-4PCS* method.

- Given two raw laser scans we first apply a voxel grid filter to roughly even out the strongly varying point distribution across point clouds that stems from the angular sampling measurement principle.
- After filtering we apply the 3D version of the DoG or Harris keypoint extractor to obtain a sparse, but nonetheless discriminative representation of the original point cloud.
- These *clouds of 3D keypoints* are matched using the 4PCS method, tuned to take into account the special characteristics of the keypoint sets.
- As a result we get quadruples of corresponding keypoint pairs, which are used to compute a rigid-body transformation.
- Finally, coarsely aligned point clouds are fine-registered with ICP.

Note that the described workflow does not depend on the type of extracted keypoints, because the matching is only based on geometrical information. In particular, we do not match descriptor vectors of keypoints, but solely rely on their relative positions in 3D space. The entire framework is thus generic and independent from a particular kind of 3D keypoint detector. In the following sections, we describe first the extraction of 3D DoG as well as 3D Harris keypoints and second the keypoint-based matching called *K-4PCS*.

4. Keypoint extraction

Standard TLS point clouds have tens of millions of unevenly distributed points, which makes coarse registration of the raw point clouds computationally very expensive. In order to reduce a point cloud to those points that are the most useful (and discriminative) for registration, we extract 3D keypoints. The *cloud of 3D keypoints* is a sparse representation with a high repeatability (i.e., stability). This property greatly increases the chance of finding a corresponding point in a second cloud of 3D keypoints. A visual example of the original point cloud and its sparse representation via a cloud of 3D keypoints is given in Fig 2.

In order to avoid the near-field bias inherent in regular angular sampling, and to achieve an initial reduction of the point count, a voxel grid filter is first applied. It divides the three-dimensional space into a regular grid of blocks (or voxels) of size τ. All scan points within each such block are determined, and their centroid is computed to henceforth serve as representative of the respective voxel. Note that working with the centroid instead of the more commonly used voxel center better preserves the original spatial layout of points. The filtered point cloud serves as input to the 3D keypoint detectors. We use 3D versions of *Difference-of-Gaussian* (DoG) and *Harris* keypoint detectors. Both are available in the open-source Point Cloud Library (PCL, Rusu and Cousins, 2011). While conceptually the generalization to 3D is straight-forward, the technical details are not widely known, so we will briefly introduce them in the following two sections.

4.1. 3D Difference-of-Gaussian keypoints

Nowadays, *Difference-of-Gaussian* (DoG) keypoint extraction in 2D is a standard tool for matching tasks in image processing. Major advantages of DoG keypoints are their invariance to scaling, rotation, and translation. In the original framework of Lowe (1999) keypoints are not only detected but

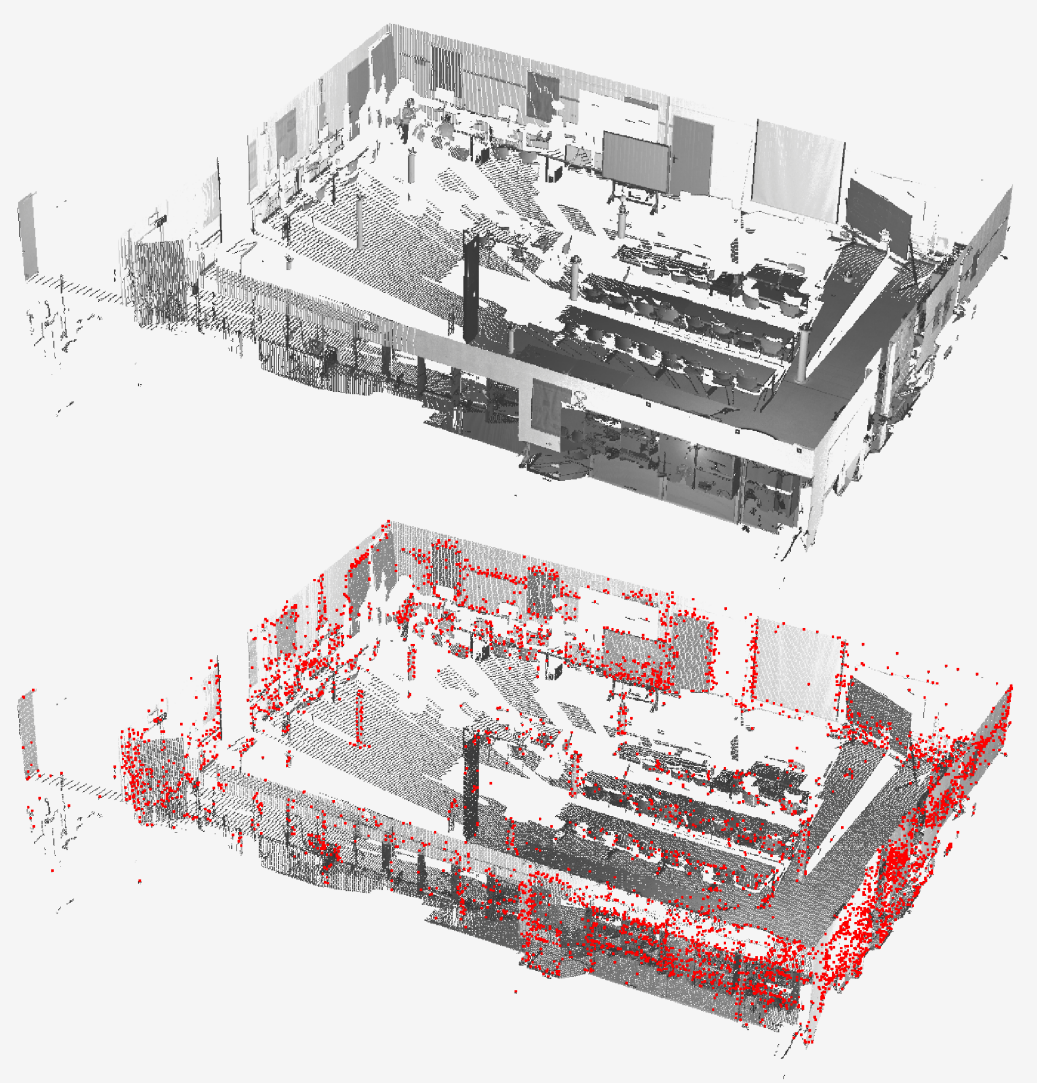

Figure 2: 3D keypoint extraction. The raw scan cloud (top) is sampled using a voxel grid based filter (bottom, gray), and the result serves as input for the detection of keypoints (here: DoG keypoints) which results in a sparse but highly discriminative keypoint cloud (bottom, red). The keypoints are visualized on top of the voxel gridded scan cloud. For better visibility, the room ceiling has been removed.

also encoded in a descriptor, which results in the well-known SIFT features. It should be noted that we do not make use of the descriptor and only employ DoG points for geometrical constraint matching. We thus prefer to use the

term DoG (instead of SIFT).

In our framework the extraction of DoG keypoints is carried out directly in 3D. The reason is that in 2D many high-contrast edges are on object silhouettes and depth discontinuities, leading to keypoints that are unstable across view points. A 3D detection scheme avoids such unstable points by selecting only points which have high contrast to their *neighbors in 3D space*. The DoG detector is an efficient approximation of the scale-normalized Laplacian. In 2D it is found by repeatedly blurring an image with Gaussian filters of increasing scale. The subtraction of images with adjacent blur scales results in a DoG response, in which local minima and maxima are then detected. For the 3D version, the same principle is based on the LiDAR return intensities. Although narrow-band LiDAR responses are quite different from image intensities (i.e., objects can have the same color but different reflectance properties and vice versa), the effectiveness of the detector does not seem to suffer significantly. Comparable to the 2D approach, the detection of 3D DoG keypoints is based on computing in each blur level τ_k (with $k = 1 \ldots m$) a Gaussian response G for each point (taking into account all neighbors in a given radius $r_k = 3 \cdot \tau_k$), subtracting the responses of adjacent scales at each point to obtain DoG responses R^G (Eq. (1)), and finally detecting keypoints as local minima and maxima in the DoG scale space.

$$R_i^G(x, y, z, \tau_k) = G_i(x, y, z, \tau_{k+1}) - G_i(x, y, z, \tau_k) \tag{1}$$

A valid keypoint is found, if the DoG response of the point is bigger (maxima) or smaller (minima) than the responses of each neighbor and additionally, if the absolute value of the points response exceeds a given threshold R_{min}. The calculation is repeated q times, doubling the base scale τ_1 of the voxel grid in each iteration (octave). The number of octaves as well as the number of scales per octave are user parameters and influence the number of extracted keypoints.

4.2. 3D Harris keypoints

As an alternative to the DoG detector we use the *Harris* corner detector introduced by Harris and Stephens (1988) and adapted to 3D space by Rusu and Cousins (2011). In comparison to the original idea of detecting corners and edges based on image gradients, the extraction in 3D is based on the local normals of the point cloud. It only uses geometrical properties and unlike DoG does not need intensities. First, a local normal is computed for

each point from all points within a neighborhood N. The search radius r to locate neighbors is chosen as $r = 3 \cdot \tau$, with τ as the voxel size of the basic grid. From the normals $\mathbf{n}_i = (n_{xi}, n_{yi}, n_{zi})^\top$ in the neighborhood a covariance matrix COV is calculated for each point in the input point cloud,

$$COV = \frac{1}{|N|} \cdot \sum_{i \in N} \mathbf{n}_i \cdot \mathbf{n}_i^\top \tag{2}$$

The covariance matrix delivers a point-wise response R^H based on determinant Det and trace Tr,

$$R^H = Det(COV) - \gamma \cdot Tr(COV)^2 \tag{3}$$

with γ an appropriately chosen constant. As for DoG, corners are extracted by searching for local maxima in the response space. The response must additionally exceed a given threshold R_{min} for a keypoint to be accepted. Note, the two thresholds are different since DoG and Harris responses do not have the same scale.

5. Keypoint-based 4-Points Congruent Sets matching

The matching of extracted keypoints is based on the *4-Points Congruent Sets* (4PCS) algorithm of Aiger et al. (2008). 4PCS was originally designed for aligning partially overlapping but rather evenly distributed, sparse point clouds with arbitrary orientation. It is an efficient variation of the obvious brute-force method to randomly sample congruent point triplets. Let us first examine the brute-force solution: a point triplet is extracted randomly from the *source* data set S, and a congruent triplet is found in the *target* data set T. From the congruent triplets, the alignment (i.e., rigid-body transformation) is determined and applied to the source point cloud to obtain S'. The alignment is verified by counting the number of point pairs from S' and T with small enough residuals. Random selection of base triplets in S is repeated until an alignment with large enough support is found (or a maximum number of iterations is reached). The problem of the brute-force approach is that it is hardly tractable, having computational complexity of at least $\mathcal{O}(n^3 \log n)$ with n as the number of points in T (Irani and Raghavan, 1996). The main insight of 4PCS is that sampling four approximately coplanar points instead of a minimal set of three points reduces computational

complexity to $\mathcal{O}(n^2)$, making it possible to align point clouds of reasonable size. Alignment proceeds in the same way (Fig. 3): given two input data sets S and T, randomly sample a four-point base set $B(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) \in S$ and search for a corresponding set $M(\mathbf{p_1}, \mathbf{p_2}, \mathbf{q_1}, \mathbf{q_2}) \in T$ with high support. However, as will be seen shortly, the corresponding sets can be found much more efficiently. To increase the geometrical stability and reliability of the transformation, the algorithm is biased towards selecting base sets with large point-to-point distances. In *K-4PCS* the maximum base length l_{max} is either defined by the user, or derived from the point cloud diameter and an estimate of the scan overlap.

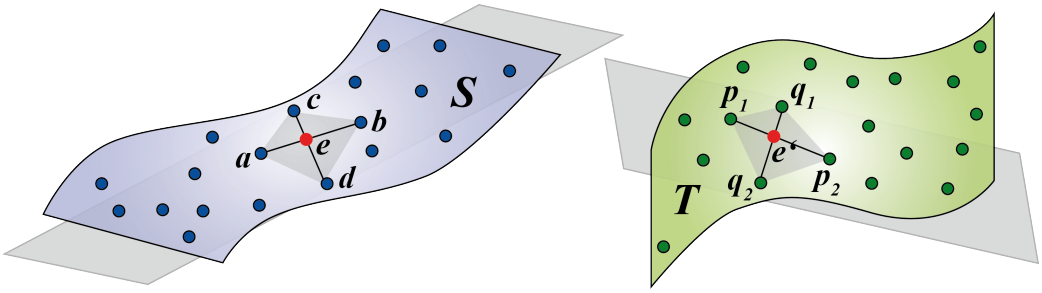

Figure 3: Basic principle of 4PCS with base set $B(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) \in S$ and a corresponding congruent point set $M(\mathbf{p_1}, \mathbf{p_2}, \mathbf{q_1}, \mathbf{q_2}) \in T$.

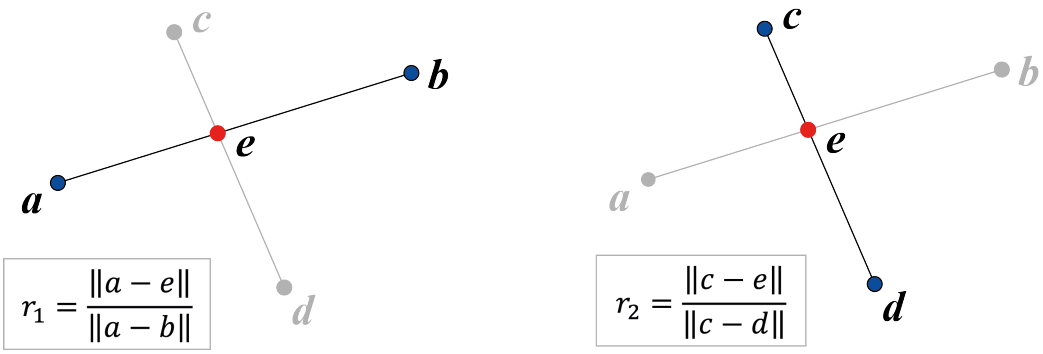

Figure 4: Illustration of the intersection point $\mathbf{e}$ using the base set $B(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) \in S$.

4PCS exploits the rule that intersection ratios of the diagonals in an arbitrary planar quadrangle are invariant under affine transformation (Huttenlocher, 1991). Having picked a base set of four approximately co-planar

points, the intersection point $\mathbf{e}$ of the diagonals as well as the corresponding intersection ratios r_1 and r_2 can be computed (Fig. 4). Finding a congruent set M then amounts to calculating two intersection points $\mathbf{e_1}, \mathbf{e_2}$ per point pair $\mathbf{p_1}, \mathbf{p_2} \in T$, using r_1 and r_2 of the current base $B \in S$ (Eq. (4)). A valid match M consists of two point pairs $P\{\mathbf{p_1}, \mathbf{p_2}\}$ and $Q\{\mathbf{q_1}, \mathbf{q_2}\}$ whose respective intersection points coincide. In practice we need to verify $\|\mathbf{e_1}(P) - \mathbf{e_2}(Q)\| < \delta_1$ or $\|\mathbf{e_2}(P) - \mathbf{e_1}(Q)\| < \delta_1$. This check is done with an efficient approximated nearest neighbor search with tolerance level δ_1.

$$\begin{aligned} \mathbf{e_1} &= \mathbf{p_1} + r_1 \cdot (\mathbf{p_2} - \mathbf{p_1}) \\ \mathbf{e_2} &= \mathbf{p_1} + r_2 \cdot (\mathbf{p_2} - \mathbf{p_1}) \end{aligned} \tag{4}$$

Up to this point, the method is able to cope with full affine transformation. However, in case of laser scan registration the task is to find a rigid-body transformation. In this case one can add additional constraints, as suggested by Aiger et al. (2008). First, instead of calculating intersection points for each point pair of T, possible pairs are filtered, based on the length of the diagonals in the base set B (Eq. (5)), and intersection points are computed only for valid point pairs.

$$\begin{aligned} \|\mathbf{p_1} - \mathbf{p_2}\| - \|\mathbf{b} - \mathbf{a}\| < \delta_2 \\ \|\mathbf{q_1} - \mathbf{q_2}\| - \|\mathbf{d} - \mathbf{c}\| < \delta_2 \end{aligned} \tag{5}$$

Second, given a matched set $M(p_1, p_2, q_1, q_2)$, the side length of the base set $B(a, b, c, d)$ can be used to verify the congruency of the irregular rectangles, that is we check

$$\begin{aligned} \|\mathbf{a} - \mathbf{c}\| - \|\mathbf{p_1} - \mathbf{q_1}\| < \delta_3 \quad , \quad \|\mathbf{a} - \mathbf{d}\| - \|\mathbf{p_1} - \mathbf{q_2}\| < \delta_3, \\ \|\mathbf{b} - \mathbf{c}\| - \|\mathbf{p_2} - \mathbf{q_1}\| < \delta_3 \quad , \quad \|\mathbf{b} - \mathbf{d}\| - \|\mathbf{p_2} - \mathbf{q_2}\| < \delta_3. \end{aligned} \tag{6}$$

The best-matching M and B are defined as those which result in the largest number of matching points (the largest overlap) between the transformed source cloud S' and the target cloud T. In practice the overlap is evaluated only for a fixed random subsample $S'_{sub} \in S'$ of 1000 points. Transformation parameters for each valid candidate match are computed using singular value decomposition and applied to S'_{sub}. Then, a nearest neighbor in T is searched for each point. The overlap is found as the fraction of nearest neighbors that

lie within a threshold δ_4. Like in standard RANSAC, base set sampling, matching and evaluation is repeated L times. The details about the calculation of L can be found in (Aiger et al., 2008, Eq. (1)). The method returns the first solution which has a support larger than a given threshold t, or in case this threshold is not reached the solution with the highest support after all L repetitions. t can be directly set by the user, or it is automatically set equal to the estimated scan overlap.

As described above, 4PCS is based on a number of tolerances δ_i. However they are all strongly correlated and can be related to the mean density of the input point clouds.[1] But these considerations do not directly apply to clouds of keypoints, which are *not* distributed uniformly. Recall that keypoints occur at characteristic and discriminative object parts but not on homogeneous surfaces. Consequently, we can hardly derive tolerances from the mean point density. Moreover, the probability of finding corresponding keypoints is much higher than that of finding correspondences in a random set of the same size, since keypoints are chosen to be repeatable across scans.

In our framework the stability, repeatability and number of extracted keypoints depend mainly on the voxel size τ of the initial voxel grid filter, which at the same time serves as base scale for DoG keypoint extraction, respectively defines the neighborhood radius of the Harris keypoint detector. Therefore, we propose the following scheme to set the parameters in a data-driven manner: we define $\delta_1, \delta_2, \delta_3$ w.r.t. the voxel grid spacing and set

$$\begin{aligned} \delta_1 &= \delta_3 = 4 \cdot \tau \\ \delta_2 &= \tau \end{aligned} \tag{7}$$

The inlier threshold $\delta_4 = \rho^2$ is still estimated from the point cloud density ρ, because *K-4PCS* allows to verify a match either using the keypoint cloud itself, or using the (usually voxel grid filtered) input point cloud. However, in the experiments described in 6 and 7.3, the support of a match is solely based on the resulting overlap of the keypoint cloud. We found that with these modified parameters 4PCS works much better with keypoints than with randomly decimated point clouds of the same size.

[1] More details about the calculation of these tolerances of the 4PCS algorithm can be found in Aiger et al. (2008) and associated open-source code.

6. Experiments

We test efficiency, success rate, the sensitivity of parameters with respect to different scenes, and robustness of *K-4PCS* on four different TLS data sets. In the following two subsections we describe and analyze results of two indoor and two outdoor data sets. The data sets address different challenges; in general one can say that the indoor data sets have rather large overlaps but a high degree of symmetry, whereas the outdoor scans have lower overlap.

To generate reference registrations we manually aligned all scans in a data set, followed by a refinement with standard ICP. Residuals between corresponding point pairs of the transformed source point cloud S' and the target point cloud T are used to estimate the accuracy of the manual registration, further called σ_0. All point correspondences which have been used in the final ICP iteration form the overlap and are henceforth called *true correspondences* $C_0 = \{c_s, c_t\}$.

The goal of *K-4PCS* is to coarsely align source and target point clouds, in such a way that the solution falls into the convergence basin of a fine registration method. To define a successful trial, the processing pipeline for experiments has thus been extended to include a refinement of the initial solution based on the standard ICP algorithm. Therefore, the success Υ of a trial can directly be evaluated based on the accuracy of the converged ICP solution ($\mathrm{RMSE}_{\mathrm{ICP}}$) with respect to manual registration accuracy σ_0 (Eq. (8)). Note, that the RMSE is only calculated using the true correspondences C_0.

$$\Upsilon = \begin{cases} 1 & \text{if} \quad \mathrm{RMSE}_{\mathrm{ICP}} \leq 3 \cdot \sigma_0 \\ 0 & \text{if} \quad \mathrm{RMSE}_{\mathrm{ICP}} > 3 \cdot \sigma_0 \end{cases} \tag{8}$$

K-4PCS is a randomized matching strategy. To even out fluctuations due to the randomization, the success is evaluated on a series of $n = 50$ trials. Our evaluation criteria is the success rate, defined by

$$\dot{\Upsilon} = \frac{1}{n} \cdot \sum_i \Upsilon \ , \quad n = 1 \ldots 50. \tag{9}$$

In addition to the success rate, the metric accuracy of a coarse alignment is addressed. The more precise we estimate an initial alignment, the lower the chance that fine registration converges towards a wrong local optimum plus convergence requires fewer iterations and registration is faster. All statistical values are computed with respect to the manually aligned references

(with accuracy σ_0). On the one hand, the transformation parameters - estimated using the four correspondences received from *K-4PCS* without an ICP refinement step - are directly compared to the reference. Differences are split into *Mean Angular Error* (MAE) and *Mean Translation Error* (MTE). On the other hand, the metric accuracy is represented by the RMSE between target T and transformed source S' - applying the transformation parameters of the coarse alignment - while considering only the true correspondences C_0. Naturally, the metric accuracy is only calculated on successful trials, ignoring falsely matched pairs.

$$\mathrm{RMSE_M} = \sqrt{\frac{1}{|C_0|} \sum_{i \in c_s, j \in c_t} (p'_i - q_j)^2} \quad , \quad p' \in S', \;\; q \in T \tag{10}$$

Finally, we also record runtimes of the keypoint matching method, which is implemented as a multi-threaded algorithm using *OpenMP*. That is, the L repetitions (base selection, candidate matching and validation) are distributed over different threads/cores and run in parallel. All experiments are carried out on a standard 64 Bit Windows 7 desktop computer with 12 cores (3.2 GHz) and 16 GB RAM.

As described in 4 and 5, *K-4PCS* has a few parameters that have to be set by the user. To find good settings and test parameter sensitivity across different types of scenes we perform extensive parameter studies. Preliminary tests have shown that some parameters did not need to be changed, but can be set constant over all tests. We thus focus on those that have to be adapted to the task. Given the expected large influence of the basic voxel grid resolution τ, all tests were done for three different voxel sizes $\tau = \{200\,mm, 100\,mm, 50\,mm\}$. Limited spatial extent and large overlap of indoor scans result in lower runtimes, allowing us to additionally evaluate a voxel size of $25\,mm$. For DoG keypoint extraction the number of octaves as well as the number of scales per octave was set to five. This proved to be a good trade-off between computation time and keypoint quality. A more crucial parameter, which strongly influences the number of extracted DoG and Harris keypoints, is the minimum response threshold R_{min}. The impact of this threshold on the success rate was evaluated on one of the data sets (see 6.1). Regarding *K-4CPS* parameters, the minimum support threshold for a registration to be deemed successful is always set to $t = 1$. This ensures that the matching method always runs for the full L iterations

and runtime remains comparable across different setting, which otherwise would strongly depend on the estimated overlap. Note however, this is the most conservative setting possible, and runtimes could often be significantly reduced with a more realistic and informed threshold. We have conducted additional tests with one indoor and one outdoor data set to evaluate varying estimations of the overlap. The maximum distance l_{max} of the base samples during matching is in general automatically calculated using the estimated overlap.

6.1. Indoor applications

We have tested *K-4PCS* on two different indoor data sets with different scene structures posing various challenges. The first data set (I) consists of five scans acquired in a laboratory room with dimensions $\approx 15\,m \times 10\,m \times 3\,m$. Five scans have been acquired with a *Zoller+Fröhlich TLS Imager 5006i* that has a field of view of 360° horizontally and 150° vertically, and a maximum range of $\approx 79\,m$. This room mainly contains standard office furniture like tables, chairs, shelves and blackboards and additionally features several cylindrical pillars (Fig. 5).

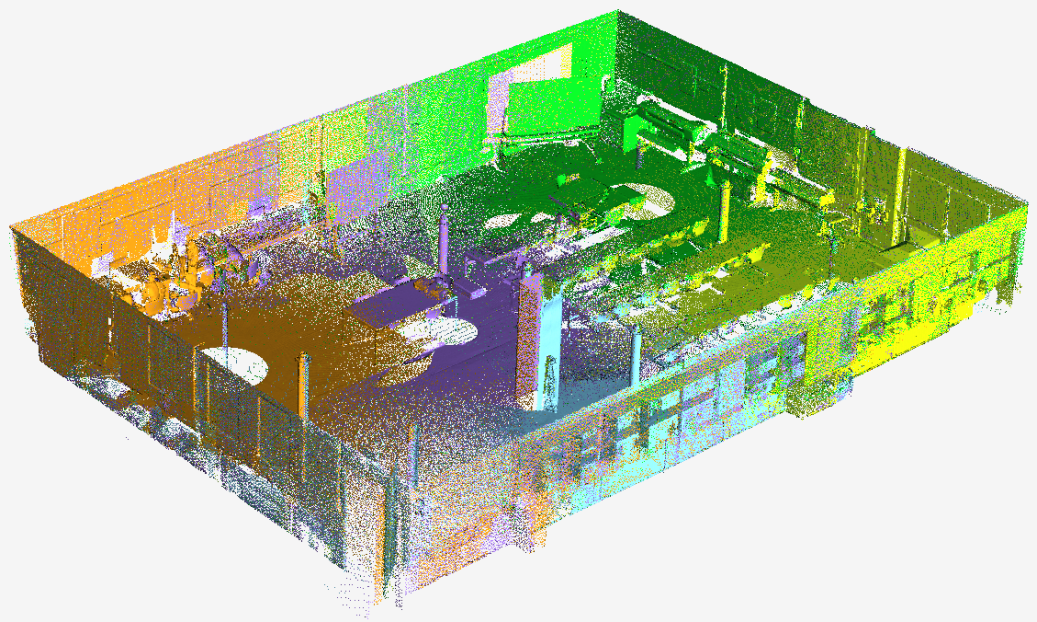

Figure 5: Data set I acquired in a laboratory room. The ceiling has been removed and only 33% of the data are displayed, for a better visual impression.

Although the room has a rather simple geometrical layout, automated alignment is demanding due to multiple rotational symmetries. On the other hand, its limited spatial extent yields large scan overlaps in spite of the limited scan range, which facilitates point cloud registration.

A first test on data set I has been carried out with an estimated overlap of 80%. Recall that the maximum base length is automatically calculated from this setting. Both keypoint detectors (DoG and Harris) were applied to point clouds of four different voxel sizes. Minimum response thresholds of valid keypoints are set to $R^G_{min} = 0.01$ for DoG, respectively $R^H_{min} = 0.0001$ for the Harris keypoint detector. Tab. 1 shows results regarding success rate $\dot{\Upsilon}$, as well as runtime T of keypoint matching. All values are averaged over all trials of all different scan pairs. Furthermore, the mean number of extracted keypoints k is given. Due to the large standard deviation of k of up to 180, the values are rounded to the nearest multiple of hundred.

	DoG				**Harris**			
$\boldsymbol{\tau}$ in mm	200	100	50	25	200	100	50	25
$\dot{\boldsymbol{\Upsilon}}$ in %	63.8	96.2	97.6	96.8	84.4	87.8	90.2	82.2
$\boldsymbol{T}$ in s	1	8	12	21	$\ll 1$	$\ll 1$	$\ll 1$	1
$\boldsymbol{k}$	400	1200	3200	8200	100	300	800	2300

Table 1: Results of data set I: mean success rate $\dot{\Upsilon}$, matching time T and number of extracted keypoints k over all scan pairs and trials. All tests were run with four different voxel sizes τ and are based on either DoG *or* Harris keypoints.

The registration success rate is generally high and reaches $\approx$ 98% with optimal parameter settings. A close inspection of failure cases revealed that these are predominantly caused by the large degree of symmetry, leading to false registrations that are 180° rotated. One property of matching that biases results towards rotated solutions (in case of high degrees of scene symmetry) is its tendency to favor solutions with minimal translations, because even after voxel grid filtering more keypoints are detected in the scanner's near field. In case of scan pairs in opposite room corners, this causes the matching to fail more often (e.g., s1-s2; see Tab. 2).

We can also observe the expected correlation between voxel size τ of the input point cloud and the number of extracted keypoints k in Tab. 1. Decreasing τ better preserves details and naturally results in more detected keypoints. This in turn leads to a higher repeatability and thus to higher success rates $\dot{\Upsilon}$. We always detect less Harris than DoG keypoints because

the geometrical structure of data set I is rather simple (while it does exhibit significant texture on the walls). As a consequence, the runtime with Harris keypoints is much lower compared to DoG. The slight drop of the success rate at the smallest τ-level ($25\,mm$) for both keypoint detectors is most probably caused by the following effect: if τ is chosen smaller than the mean point density of the input point cloud in the overlap area, keypoint detection is biased towards extracting more keypoints in dense, non-overlapping areas. This increases the chance of selecting bases that are located outside the scan overlap, and consequently reduces the probability of finding a correct match. In conclusion, highest success rates are achieved by setting τ in the range of the estimated point density in the overlapping area (i.e., here $\approx 50\,mm$). Runtime is again a function of the voxel size: smaller τ generally yields higher success rates, at the cost of longer computation time.

The metric accuracy was tested using a voxel size $\tau = 50\,mm$ and DoG keypoints. The results are shown in Tab. 2. Note, that the standard deviation of the mean value is referring to the average standard deviation over the single scan pairs. It can be seen, that the angular error is $\approx 0.6°$ and does not change significantly between scan pairs. The translation error as well as the $\mathrm{RMSE_M}$ are $< 15\,cm$, whereas slightly worse solutions are reached in case of less successful scan pairs. After ICP refinement, the RMSE considering the true correspondences is reduced to $\approx 1\,cm$, which is on the order of the scanner's measurement accuracy. We noted that all standard deviations (1σ) are in the same order as the mean values. In conclusion, the reachable accuracy of the coarse alignment is quite variable, but in most cases (here $\approx 98\%$) is still good enough to put the scans into the convergence basin of ICP.

In a further test we have assessed the influence of minimum response threshold R_{min} of both keypoint detectors. Like the voxel size τ, this threshold has a direct impact on the number of detected keypoints. Tab. 3 compares the number of detected DoG keypoints and the resulting success rates at varying $R^G_{min} \in \{0.005, 0.01, 0.02, 0.04\}$.

It comes as no surprise that the number of keypoints increases with a decreasing minimum response R_{min} and with a decreasing voxel size τ. Naturally, $\check{\Upsilon}$ highly depends on the number of detected keypoints. Recall that runtime quadratically increases with the number of keypoints and thus larger keypoint numbers than necessary are in general not desirable.

The matching part of *K-4PCS* depends on 3 parameters (cf. 5). While the minimum scale τ has been varied for each previous test, the maximum

	$\dot{\Upsilon}$	MAE	MTE	$\mathbf{RMSE_M}$	$\mathbf{RMSE_{ICP}}$
	in %	in °	in mm	in mm	in mm
s1-s2	86.0	0.8 ± 0.7	169 ± 94	168 ± 65	16 ± 0
s1-s3	100.0	0.5 ± 0.4	128 ± 58	125 ± 49	11 ± 0
s1-s4	98.0	0.7 ± 0.6	141 ± 99	148 ± 72	4 ± 0
s1-s5	98.0	0.6 ± 0.6	133 ± 99	138 ± 84	11 ± 0
s2-s3	100.0	0.3 ± 0.2	72 ± 32	73 ± 23	11 ± 0
s2-s4	100.0	0.6 ± 0.5	112 ± 88	114 ± 63	9 ± 0
s2-s5	100.0	0.5 ± 0.4	132 ± 71	124 ± 42	9 ± 0
s3-s4	98.0	0.6 ± 0.6	99 ± 39	132 ± 54	6 ± 0
s3-s5	100.0	0.5 ± 0.4	100 ± 78	122 ± 64	10 ± 0
s4-s5	96.0	0.8 ± 0.6	146 ± 64	173 ± 65	10 ± 0
Mean	**97.6**	$\mathbf{0.6 \pm 0.5}$	$\mathbf{122 \pm 72}$	$\mathbf{131 \pm 58}$	$\mathbf{10 \pm 0}$

Table 2: Metric accuracy represented by the *Mean Angular Error* (MAE), *Mean Translation Error* (MTE) and *Root Mean Square Error* between true correspondences before ($\mathrm{RMSE_M}$) and after ICP refinement ($\mathrm{RMSE_{ICP}}$). Results are shown for data set I, using $\tau = 50\,mm$ and DoG keypoints.

	$R_{min} = 0.04$		$R_{min} = 0.02$		$R_{min} = 0.01$		$R_{min} = 0.005$	
$\boldsymbol{\tau}$	**k**	$\dot{\Upsilon}$	**k**	$\dot{\Upsilon}$	**k**	$\dot{\Upsilon}$	**k**	$\dot{\Upsilon}$
in mm		in %		in %		in %	in %	
200	100	26.6	300	58.4	400	63.8	500	62.4
100	200	20.2	700	75.6	1200	96.2	1500	98.0
50	300	15.0	1700	82.6	3200	97.6	4200	98.4
25	800	18.0	4300	81.2	8200	96.8	11000	98.6

Table 3: Success rate $\dot{\Upsilon}$ and number of detected DoG keypoints k with different minimum response thresholds R_{min} applied on data set I.

sampling distance for the base selection l_{max} as well as the estimated overlap were so far set to constant values derived from preliminary tests. Although l_{max} is usually automatically derived from the estimated overlap, additional tests were performed to assess success rates while varying both matching parameters. Note, to be independent of the absolute point cloud size, the maximum baseline is again given in percent of the point cloud diameter. It turns out that the success rate is not significantly altered if choosing different parameters in a range of $0.6 \ldots 1.0$, which is probably caused by the large overlap of scans in data set I. Note, the number of sampling iterations L

is derived from the estimated overlap (higher overlap → fewer iterations). Hence, tests with high overlap have shorter runtime, but can result in lower success rates.

The second indoor data set (II) has been acquired in a large workshop that houses construction experiments of the civil engineering department at ETH. It consists of large, regular structures and contains some big machinery. To cover the large scene with dimensions of $\approx 100\,m \times 30\,m \times 12\,m$ eight scans were required, in spite of rather small overlaps ($\approx 40\% - 60\%$). The scans have been captured with a *Faro Focus3D* TLS scanner with a field of view of $360° \times 152.5°$ and a maximum measurement distance of $150\,m$. Constrained by the tube-like geometry of the room, only immediately adjacent scans have a reasonable overlap, leading to seven scan pairs in total (Fig. 6).

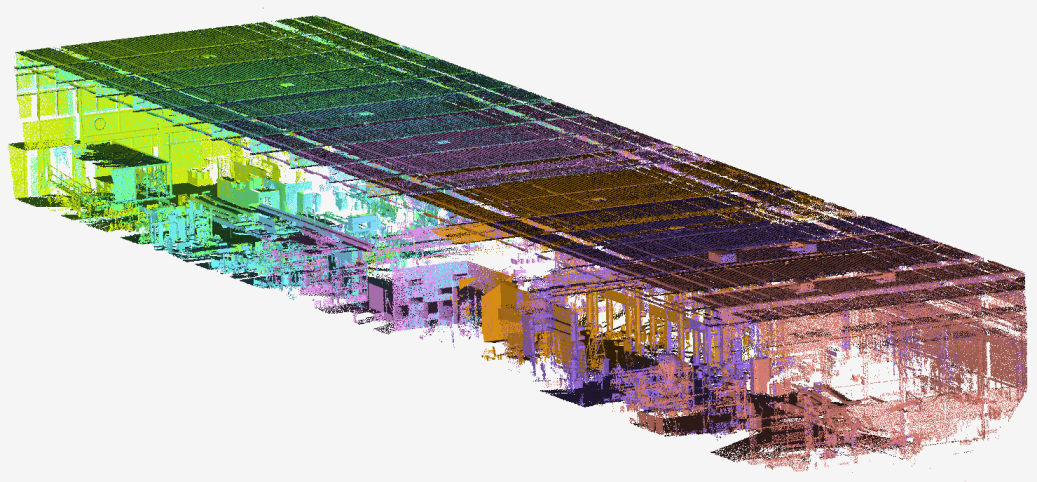

Figure 6: Data set II of a big workshop: all eight scans have been acquired in a consecutive way and are displayed with a point count of 25%. The wall facing the viewer has been removed for better visualization.

Again, all tests were conducted with four different voxel sizes τ. Due to the very large extent of the room and to keep keypoint numbers at a manageable level, the minimum response threshold was increase to $R^G_{min} = 0.02$ for Dog keypoints and $R^H_{min} = 0.002$ for Harris keypoints, respectively. Estimated overlaps are set to 0.6 for the first 3 scan pairs, which were acquired with higher overlaps to increase the level of detail in the first part of the hall. The 4 other scan pairs have been tested with a value of 0.5.

Compared to data set I the success rate $\dot{\Upsilon}$ is much lower, reaching a max-

	DoG				Harris			
$\boldsymbol{\tau}$ in mm	200	100	50	25	200	100	50	25
$\dot{\boldsymbol{\Upsilon}}$ in %	60.6	60.9	53.1	42.9	45.4	47.4	47.1	15.1
$\boldsymbol{T}$ in s	8	17	38	124	2	6	11	80
$\boldsymbol{k}$	1100	2500	5600	13200	600	1500	2800	10200

Table 4: Results of data set II: mean success rate $\dot{\Upsilon}$, matching time T and number of extracted keypoints k over all scan pairs and trials. All tests were run with four different voxel sizes τ and are based on either DoG *or* Harris keypoints.

imum of $\approx 61\%$ based on DoG keypoints and a voxel size of $\tau = 100\,mm$. The explanation of this lower $\dot{\Upsilon}$ lies in the challenging tube-like geometry. As already mentioned previously, the verification by the number of matched points tends to return solutions where the translation between scan pairs is small. In case of scenes with repetitive structures like those in the elongated workshop, it tends to place scans at the same location, with no displacement. A close inspection of Fig. 7 supports this finding: the vast majority of false registrations occurs in the middle of the hall (s5, s6). Here, scans cover only two walls on the long sides of the workshop, where the structure is repetitive (Fig. 6). The low success rate $\dot{\Upsilon}$ of the 3 scan pairs containing s5 or s6 ($< 50\%$) lowers the overall success rate as shown in Tab. 4, although the remaining pairs are mostly successfully matched (on average in 96% with $\tau = 100\,mm$ and based on DoG keypoints). The diagnosis is reinforced by comparing results of the DoG and Harris detectors. Overall, matching based on DoG keypoints is more successful than with Harris keypoints, although total keypoint numbers are comparable: the purely geometric Harris keypoints are ambiguous, whereas contrast due to wall texture, floor markings etc. generates some useful DoG keypoints.

Considering the metric accuracy shown in Tab. 5, MAE, MTE as well as $\mathrm{RMSE_M}$ are worse than in data set I. Probable reason for this effect, is the lower overlap between neighboring scans, leading to less stable base samples and consequently to lower accuracies. Although the average accuracy is 3 to 4 times worse than in data set I, the alignment is still good enough to be correctly refined by ICP. The accuracy of $\approx 1\,cm$ after refinement is again close to the expected measurement accuracy of the scanning device.

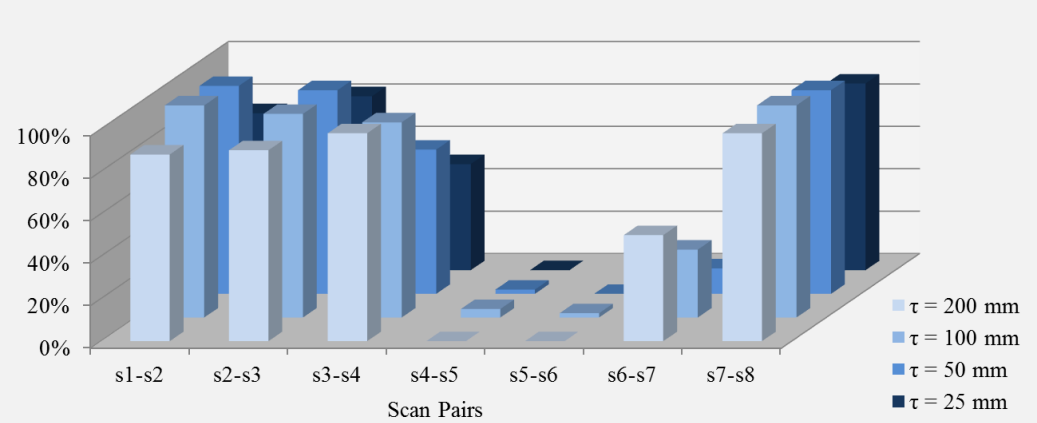

Figure 7: Success rate $\dot{\Upsilon}$ of the different scan pairs of data set II based on four different voxel sizes τ and DoG keypoints.

	$\dot{\Upsilon}$ in %	**MAE** in °	**MTE** in mm	$\mathbf{RMSE_M}$ in mm	$\mathbf{RMSE_{ICP}}$ in mm
s1-s2	98.0	0.8 ± 0.8	331 ± 334	389 ± 316	7 ± 0
s2-s3	96.0	0.8 ± 1.0	377 ± 283	418 ± 282	11 ± 0
s3-s4	68.0	1.9 ± 2.1	926 ± 662	1122 ± 689	8 ± 0
s4-s5	2.0	2.1 ± 1.6	673 ± 0	647 ± 0	5 ± 0
s5-s6	0.0	—	—	—	—
s6-s7	12.0	3.1 ± 3.5	866 ± 456	1340 ± 687	11 ± 0
s7-s8	96.0	1.2 ± 1.4	587 ± 533	616 ± 408	13 ± 0
Mean	**53.1**	$\mathbf{1.2 \pm 1.3}$	$\mathbf{537 \pm 435}$	$\mathbf{621 \pm 410}$	$\mathbf{10 \pm 0}$

Table 5: Metric accuracy represented by the *Mean Angular Error* (MAE), *Mean Translation Error* (MTE) and *Root Mean Square Error* between true correspondences before ($\mathrm{RMSE_M}$) and after ICP refinement ($\mathrm{RMSE_{ICP}}$). Results for data set II using $\tau = 50\,mm$ and DoG keypoints.

6.2. Outdoor applications

The applicability of *K-4CPS* to outdoor TLS projects is evaluated on two challenging data sets. Data set III (Fig. 8) was acquired in an urban area, data set IV in a forest (Fig. 9). Similar to the experiments on indoor scans, suitable ranges for all parameters were defined according to preliminary tests.

Data set III consists of four high-resolution scans with more than 20 million points per acquisition. The scans cover a Roman arch and its surroundings including paths, vegetation, small parks, and a building (Fig. 8)

and were acquired using a *Zoller+Fröhlich TLS Imager 5006i* (cf. data set I). What makes this data set challenging are (i) low overlap of adjacent scan pairs ($\approx 40\%$), (ii) vegetation and (iii) numerous artifacts caused by moving people who where visiting the arch during scan acquisition. Only four scan pairs have reasonable overlap for matching.

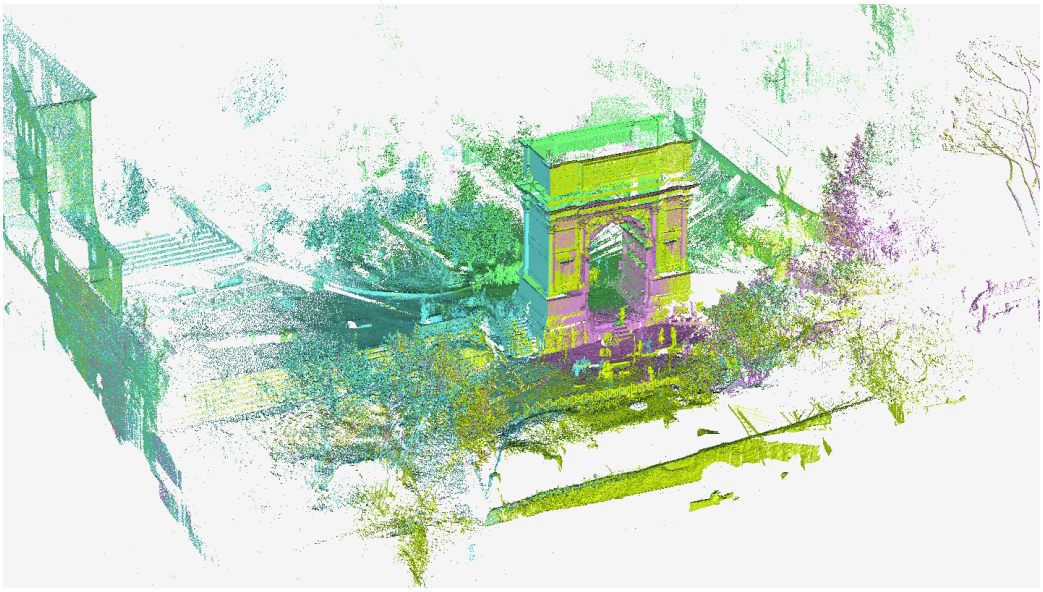

Figure 8: Data set III with four scans distributed around a Roman arch. Note that walking people cause artifacts and adjacent scan pairs have low overlap. For reasons of visualization data are downsampled to 25% of the original scan resolution.

In a first experiment, the average success rate, matching runtime and number of extracted keypoints averaged over all four scan pairs are analyzed. Parameter settings are: for DoG $R^G_{min} = 0.01$ and for Harris $R^H_{min} = 0.001$, representing a good trade-off between quality and number of keypoints; estimated overlap 0.4. Tests are run with three different voxel sizes $\tau = \{0.2\,m, 0.1\,m, 0.05\,m\}$; results are shown in Tab. 6. Recall that indoor experiments achieve best results if the voxel size approximately corresponds to the mean point density in overlap areas. Data set III has a rather low mean point density due to the small overlap areas and we thus did not test voxel sizes below $0.05\,m$.

With $\tau = 50\,mm$ for DoG keypoints and $\tau = \{50\,mm, 100\,mm\}$ for Harris keypoints we achieve high success rates of $\dot{\Upsilon} > 90\%$. Harris keypoints clearly work better on this data set, because a lack of distinctive texture in the

	DoG			Harris		
$\boldsymbol{\tau}$ in mm	200	100	50	200	100	50
$\dot{\boldsymbol{\Upsilon}}$ in %	23.0	56.0	92.0	85.0	100.0	96.5
$\boldsymbol{T}$ in s	12	64	387	40	192	887
$\boldsymbol{k}$	1300	3600	11000	1900	6500	17700

Table 6: Results of data set III: mean success rate $\dot{\Upsilon}$, matching time T and number of extracted keypoints k over all scan pairs and trials. All tests were run with four different voxel sizes τ and are based on either DoG *or* Harris keypoints.

scan intensities leads to fewer and less reliable DoG keypoints. Runtimes are higher compared to both indoor data sets due to the low overlap between scan pairs, which requires a higher iteration number L, but still remain feasible for many practical applications.

	$\dot{\boldsymbol{\Upsilon}}$	**MAE**	**MTE**	$\mathbf{RMSE_M}$	$\mathbf{RMSE_{ICP}}$
	in %	in °	in mm	in mm	in mm
s1-s2	74.0	2.4 ± 2.3	1275 ± 767	1498 ± 804	22 ± 5
s1-s3	96.0	1.2 ± 1.4	743 ± 481	904 ± 529	15 ± 0
s2-s3	98.0	0.9 ± 0.7	478 ± 418	619 ± 404	18 ± 2
s4-s3	100.0	0.5 ± 0.5	303 ± 169	347 ± 156	12 ± 3
Mean	**92.0**	$\mathbf{1.2 \pm 1.2}$	$\mathbf{660 \pm 437}$	$\mathbf{796 \pm 450}$	$\mathbf{16 \pm 2}$

Table 7: Metric accuracy represented by the mean angular error (MAE), mean translation error (MTE) and the root mean square error between true correspondences before ($\mathrm{RMSE_M}$) and after ICP refinement ($\mathrm{RMSE_{ICP}}$). Results are for data set III, $\tau = 50\,mm$ and DoG keypoints.

Generally Tab. 7 shows that the metric accuracy is again significantly lower compared to data set II in the indoor applications, due to even smaller overlap between adjacent scans. The mean angular accuracy is $\approx 1°$ while the translation error is around $0.7\,m$. The refinement step reduces the RMSE by a factor of 60 to $\approx 1.6\,cm$ and again reaches the expected scanner accuracy. Note, that scan pair s1-s2 has significantly worse accuracies, probably caused by the even lower overlap between this two scans, and the associated smaller base length between the correspondences. The smaller overlap also explains the lower success rate of $\dot{\Upsilon} = 74\%$ for this scan pair.

An additional experiment with data set III addresses the sensitivity to the estimated overlap (with varying values of $\{0.3, 0.4, 0.5\}$), using DoG keypoints and setting $\tau = 50\,mm$. The tests uncover that $\dot{\Upsilon}$ is more sensitive to

the estimated overlap compared to indoor data set I. An estimated overlap of 0.3 gives clearly better results with success rates of 100%, but at the cost of much larger runtimes ($> 1'000\,s$).

To evaluate *K-4PCS* under extreme conditions we perform experiments on data set IV of a forested area (Fig. 9). It consists of two groups of three scans. All scans in a group overlap with $\approx 50\%$ and mainly contain bushes and trees. The data has been acquired with the a *Faro Focus3D* laser scanner (cf. data set II).

Figure 9: Data set IV of a forested area. Unstructured trees and bushes make this data set challenging. The illustration only displays 1% of the points of each scan.

Again, we run the same tests. Estimated overlap is fixed to 0.5 for all scan pairs. The large amount of vegetation with strongly varying reflectance and geometry heavily increases the number of keypoints. To keep the point count at a reasonable level the minimum response is set to $R^G_{min} = 0.02$ and $R^H_{min} = 0.002$.

Tab. 8 shows that results on this challenging data set are $\approx 30\%$ points below those of data set III. Fig. 10 reveals that automated coarse alignment for two scan pairs (out of six in total) almost completely fails. If we put those two aside and only look at average success rates over all trials of the four remaining pairs, we achieve $\dot{\Upsilon} = 96\%$ for DoG and $\dot{\Upsilon} = 99\%$ for

	DoG			Harris		
$\boldsymbol{\tau}$ in mm	200	100	50	200	100	50
$\dot{\boldsymbol{\Upsilon}}$ in %	32.7	45.3	64.7	66.0	66.3	68.0
$\boldsymbol{T}$ in s	8	34	116	25	44	100
$\boldsymbol{k}$	1100	3500	9200	1600	4000	8700

Table 8: Results of data set IV: mean success rate $\dot{\Upsilon}$, matching time T and number of extracted keypoints k over all scan pairs and trials. All tests were run with four different voxel sizes τ and are based on either DoG or Harris keypoints.

Harris keypoints, respectively. Similar to indoor data set II of the tube-like workshop, failure cases are caused by repetitive structures, which, somewhat surprisingly, also appear in the forest setting.

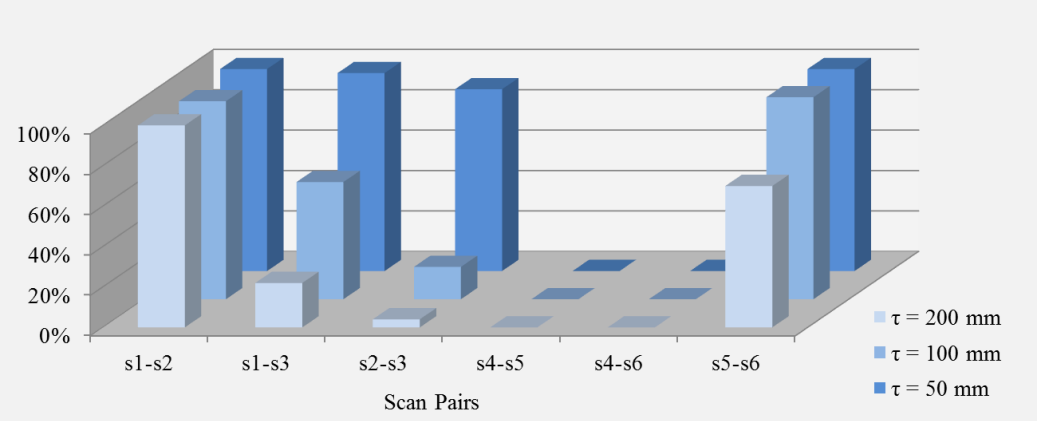

Figure 10: Success rate $\dot{\Upsilon}$ of the different scan pairs of data set IV based on three different voxel sizes τ and DoG keypoints.

The achieved metric accuracy is visualized in Tab. 9. The values are again based on $\tau = 50\,mm$ and DoG keypoints. On the one hand, the angular accuracy is worse compared to previous tests with other data sets. The reason for this effect is probably the limited extent of the point clouds in lateral direction (w.r.t. the main displacement of the scanner, see Fig. 9). This in combination with the low overlap of $\approx 50\%$ limits the base length of candidate keypoints and consequently reduces angular accuracy. Standard ICP brings the accuracy down to $< 1\,cm$.

After comprehensive experiments with two indoor (I, II) and two outdoor data sets (III, IV), the major bottleneck that remains apparently is how to

	$\dot{\Upsilon}$	**MAE**	**MTE**	$\mathbf{RMSE_M}$	$\mathbf{RMSE_{ICP}}$
	in %	in °	in mm	in mm	in mm
s1-s2	100.0	2.5 ± 2.2	797 ± 367	924 ± 367	4 ± 0
s1-s3	98.0	3.8 ± 3.0	1102 ± 408	1463 ± 569	5 ± 0
s2-s3	90.0	1.8 ± 1.5	582 ± 268	718 ± 321	6 ± 0
s4-s5	0.0	—	—	—	—
s4-s6	0.0	—	—	—	—
s5-s6	100.0	1.6 ± 1.4	469 ± 198	523 ± 207	2 ± 0
Mean	**64.7**	$\mathbf{2.5 \pm 2.1}$	$\mathbf{740 \pm 311}$	$\mathbf{909 \pm 366}$	$\mathbf{4 \pm 0}$

Table 9: Metric accuracy represented by the mean angular error (MAE), mean translation error (MTE) and the root mean square error between true correspondences before ($\mathrm{RMSE_M}$) and after ICP refinement ($\mathrm{RMSE_{ICP}}$). Here, the results of the tests based on data set IV, $\tau = 50\,mm$ and DoG keypoints are shown.

deal with symmetries and repetitive structures. To get a deeper understanding of this potential failure scenario we perform additional experiments on simulated scans.

7. Evaluation on synthetic data

Experimental results on both indoor and outdoor scans reveal that the large majority of failure cases is caused by translationally or rotationally symmetric scene content. To investigate this issue in detail under controlled conditions, we turn to synthetic data sets. To that end we have designed a LiDAR point cloud simulator for simple indoor scenes (7.1) and have performed multiple tests (7.3).

7.1. Simulator

This algorithm simulates TLS point clouds taken indoors. A simple symmetric room is setp up with six major planes $Pl_j, j = 1 \ldots 6$ representing ceiling, floor, and walls. The dependency of *K-4PCS* on objects that break symmetry is analyzed by introducing an additional seventh plane Pl_7 which cuts off a part of the symmetric room (Fig. 11). To address different levels of symmetry, the normal $\mathbf{n_7}$ and the orthogonal distance o_7 (with respect to the room origin) of the seventh plane are varied.

The geometrical construction of a virtual scan is carried out as follows. The virtual scanner is placed inside the room, with its orientation set randomly and its inclination is limited to vary only up to a few degrees around

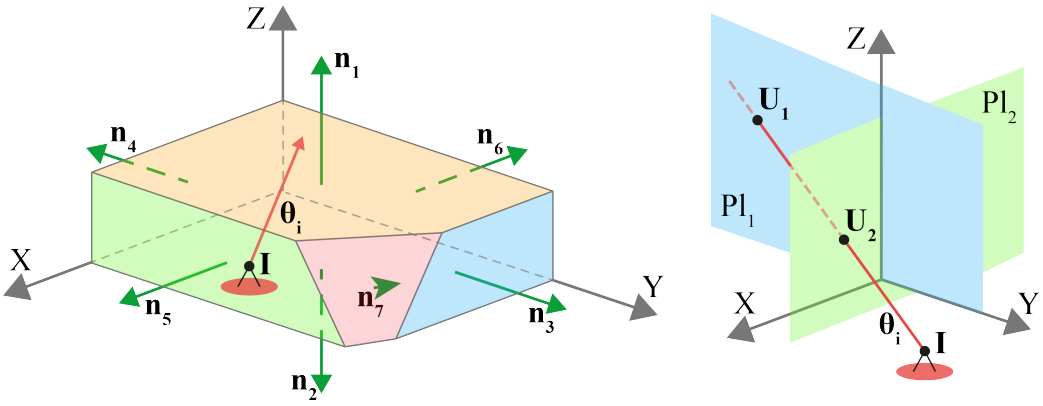

Figure 11: Principle of the TLS simulator. Left: geometrical setup of the synthetic indoor room with six major planes and symmetry-breaking seventh plane (red). Right: point cloud generation by searching for closest intersection points $\mathbf{U}$ between the seven room planes Pl and the measurement beams $\theta_\mathbf{i}$ sent from a scan station $\mathbf{I}$.

zero to simulate a approximately leveled instrument. After setting the angular resolution of the simulated scanner, a direction vector θ_i for each laser beam i is calculated. Then starting from the instrument position I, all beams are intersected with each plane Pl. The length of the beam is obviously $\|\mathbf{U} - \mathbf{I}\|$ and can directly be calculated from:

$$d_{i,j} = \frac{o_j - \mathbf{n_j} \cdot \mathbf{I}}{\theta_\mathbf{i} \cdot \mathbf{n_j}} \tag{11}$$

Since LiDAR is a line-of-sight instruments, the intersection point with the shortest distance d_i is kept for every beam θ_i. To simulate the range measurement accuracy of a LiDAR instrument, Gaussian noise with $\sigma = 5\,mm$ is added to the distance.

Working with DoG keypoints additionally requires a gray value. For real scans this is the intensity of the received measurement signal recorded by the LiDAR instrument. We assign intensities to the synthetic point cloud via small patches from real scan intensity images of indoor walls, floors, and ceilings. A texture image per plane is generated by randomly assembling small patches until the required image size (given by the room dimension) is reached. Image resolution is set to $5\,mm$ to ensure that neighboring LiDAR points get intensities information from different pixels. Examples of texture images are shown in Fig. 12.

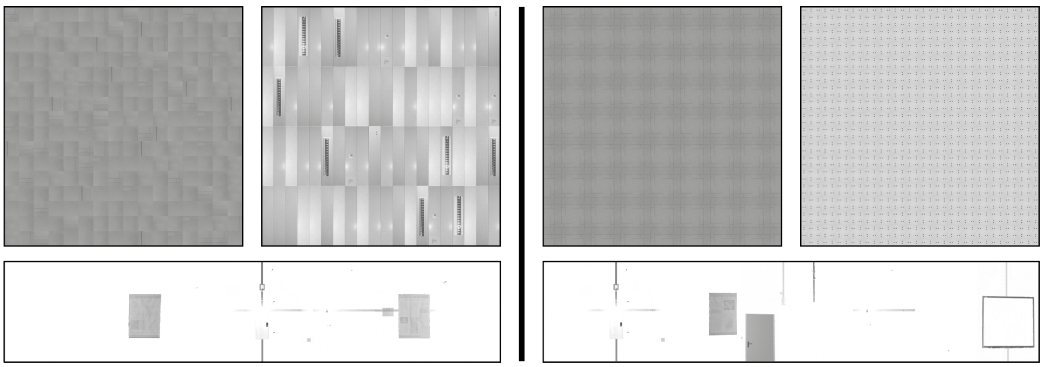

Figure 12: Generated texture images (floor, ceiling and a wall example) to map intensity values onto the point cloud. Left: Texture images of setup 1 to simulate a geometrically symmetrical room with asymmetric textured planes. Right: Texture images of setup 2, which simulates a geometrically *and* radiometrically symmetrical room. Note that texture images of walls do not need to be symmetric, because rotation-symmetry is achieved by mapping the same texture onto all walls.

When looking up the synthetic intensities, original texture values are modified based on the incidence angle of the beam on the wall, to mimick the behavior of a real scanner. We apply an exponential damping function to decrease intensity with smaller incidence angles.

7.2. Synthetic data sets

We generate two different collections of synthetic data sets (i.e., setups) to evaluate the influence of symmetric scene content on registration performance with respect to *geometry* and to *radiometry*. Both setups are based on a quadratic room with a side length of $20\,m$ and a height of $4\,m$. We first address geometric symmetry in setup 1. Radiometric (texture) variation is created by randomly assembling different small patches to texture images. This is repeated for each room plane so that all planes are differently textured. The room's geometric layout is thus 90° rotationally symmetric with respect to its vertical axis, as opposed to its radiometric layout, which is not symmetric. To evaluate the influence of concurrent radiometric and geometric symmetry, setup 2 has rotationally symmetric texture images for ceiling and floor. All walls are textured with the same image (which in itself does not have to be symmetric). The room of setup 2 is 90° rotationally symmetric (with respect to the vertical axis of the room) in terms of both geometry and radiometry. Examples of the applied texture images are shown in Fig. 12. In

Fig. 13 an example of a room of setup 1 is shown left whereas a room of setup 2 is shown right. Note that both examples already contain the seventh plane that cuts off the corner (pointing towards the reader) to break symmetry.

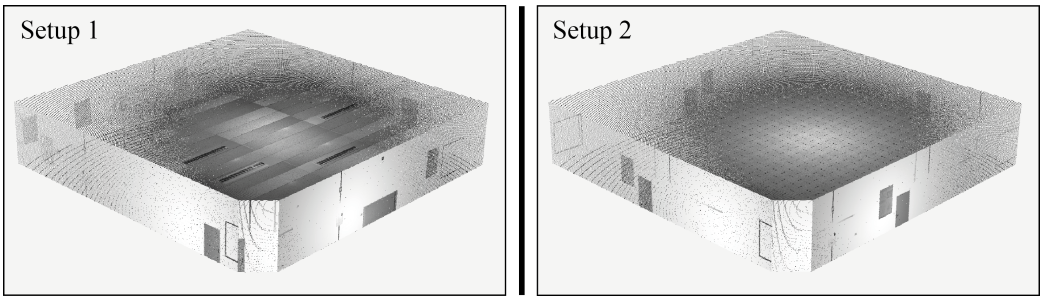

Figure 13: Examples of the simulated point clouds with a symmetry-breaking plane at the room corner in front. Left: setup 1 with geometrical symmetric room but asymmetric radiometry (texture). Right: exemple of setup 2 with repetitive texture patterns on the walls leading to a 90° rotational symmetry in geometry *and* radiometry.

Six data sets are simulated for both setups 1 and 2. More precisely, we test six different settings of the symmetry-breaking plane that can be divided into two test scenarios A and B, each including a zero case (Fig. 14) without the seventh plane. Scenario A simply moves the cutting plane into the room in three steps (orthogonal distances to the room corner of $1\,m$, $2\,m$, and $3\,m$) whereas the angles to the walls on both sides are 45° (left in Fig. 14). In B a vertical plane cuts off the same room corner as in A but with three different angles of 11.25°, 22.5°, and 45° (right in Fig. 14), while the intersection line between the cutting plane and the reference wall is kept.

All rooms (six for setup 1 and six for setup 2) are virtually scanned from five positions, such that the room is well covered. The five scan positions are the same over all experiments, so as to make the different runs directly comparable.

7.3. Results

As the simulated environment lacks geometric clutter like chairs or tables, the use of the 3D Harris detector is not reasonable. Hence, all tests with simulated data are performed with 3D DoG keypoints. In the following, we will first show results of setup 1. It addresses the challenge of geometrical symmetry, while the radiometry (LiDAR intensity texture) is randomly assigned. Thereafter, we describe setup 2 with radiometric and geometrical symmetries.

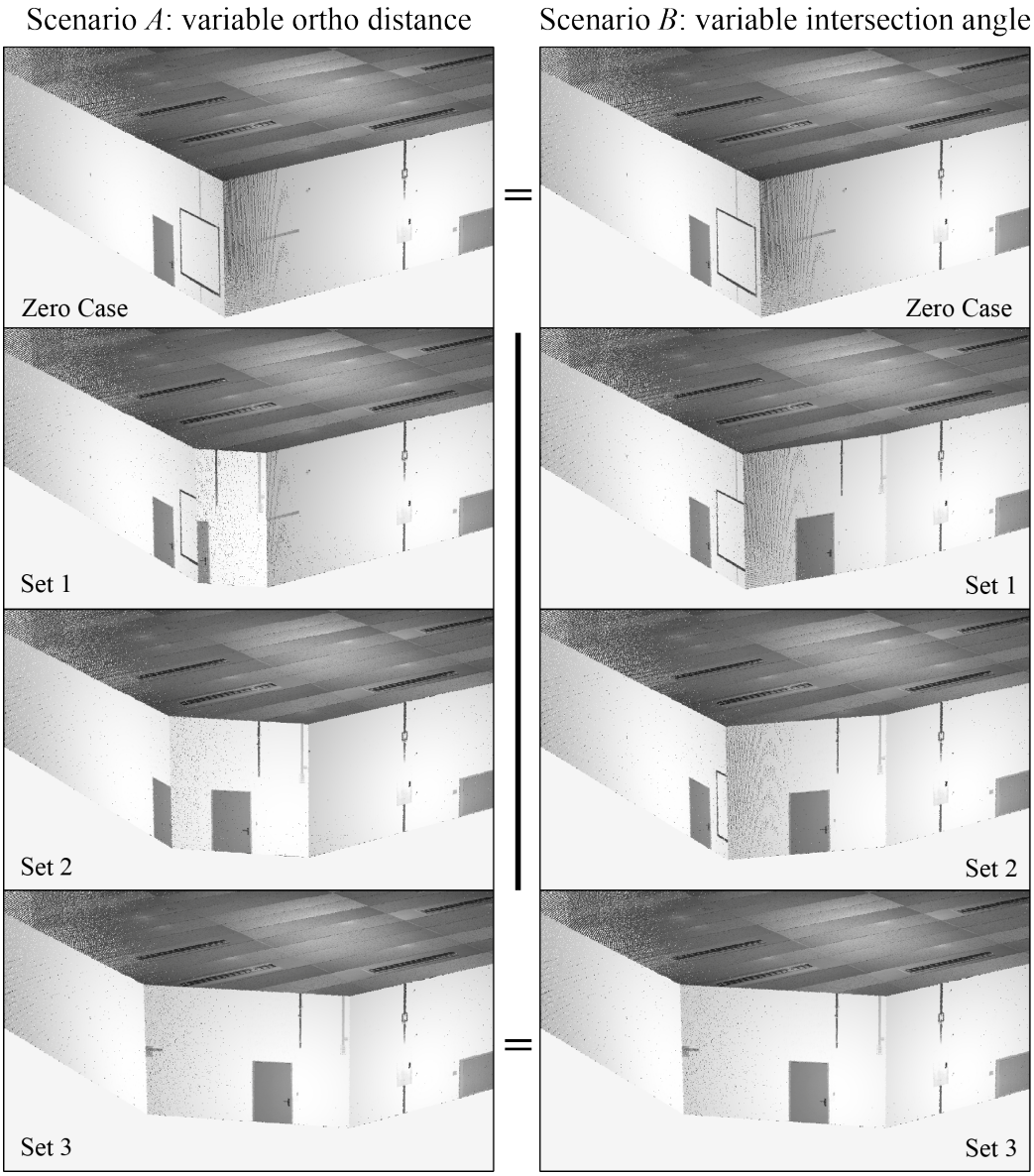

Figure 14: Each of the two setups consists of 6 data sets with variable properties of the symmetry-breaking additional plane. In test scenario A (left) the orthogonal distance is increased by steps of $1\,m$ whereas in test scenario B (right) the intersection angle with a room wall is varied. The images show setup 1, without radiometric symmetry.

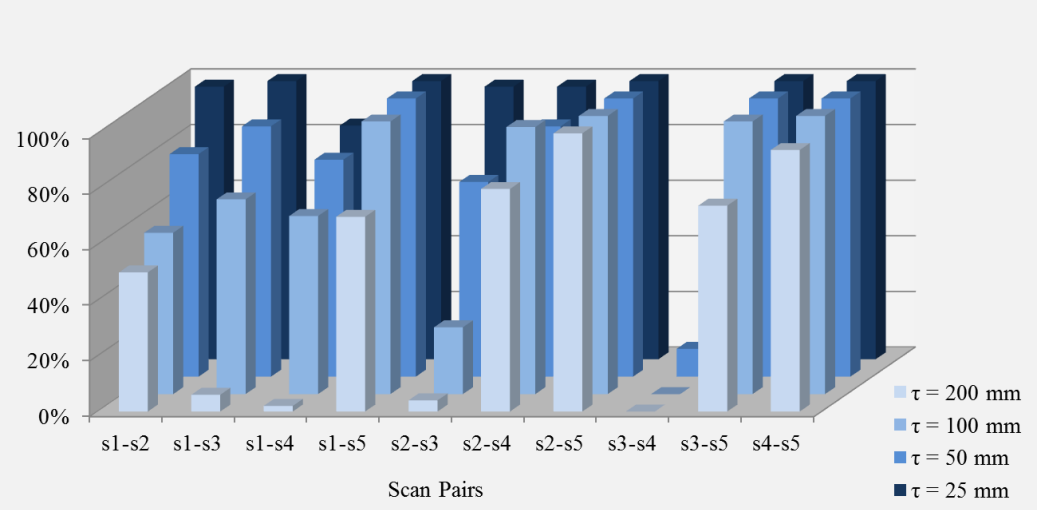

Figure 15: Success rates $\dot{\Upsilon}$ (vertical axis) of the zero case of setup 1 per scan pair (horizontal axis), as a function of voxel size τ (depth/color).

To evaluate the influence of the symmetry-breaking plane, we first analyze success rate ($\dot{\Upsilon}$) of the zero case without any additional plane. Fig. 15 clearly shows an increasing $\dot{\Upsilon}$ with decreasing voxel grid size τ, for all scan pairs. This dependency has also been observed in real TLS data sets (see 6). Obviously, τ strongly influences the amount and repeatability of keypoints and consequently the success rate $\dot{\Upsilon}$. Scans s3-s4, placed on opposite room corners and close ($< 5\,m$) to the walls, stand out with very low $\dot{\Upsilon}$, especially at higher τ. Recall that *K-4PCS* is somewhat biased towards solutions with small translation. In case of opposite room corners, where the large majority of keypoints is detected in the scanner's near field, the method tends to return 180° rotated solutions.

The influence of the degree of symmetry for setup 1 on the success rate is shown in Fig. 16 for both scenarios A and B. The zero case already delivers relatively high success rates for setup 1 (with only geometrical symmetry and random texture). Breaking geometrical symmetry by moving the additional plane into the room further improves results significantly. It is evident that the use of DoG keypoints, which rely on texture, stabilizes the algorithm. However, as soon as texture becomes repetitive and symmetric, the picture drastically changes, as can be seen in Figs. 17 and 18. In setup 2 with symmetric geometry and texture, the results severely degrade. Registration

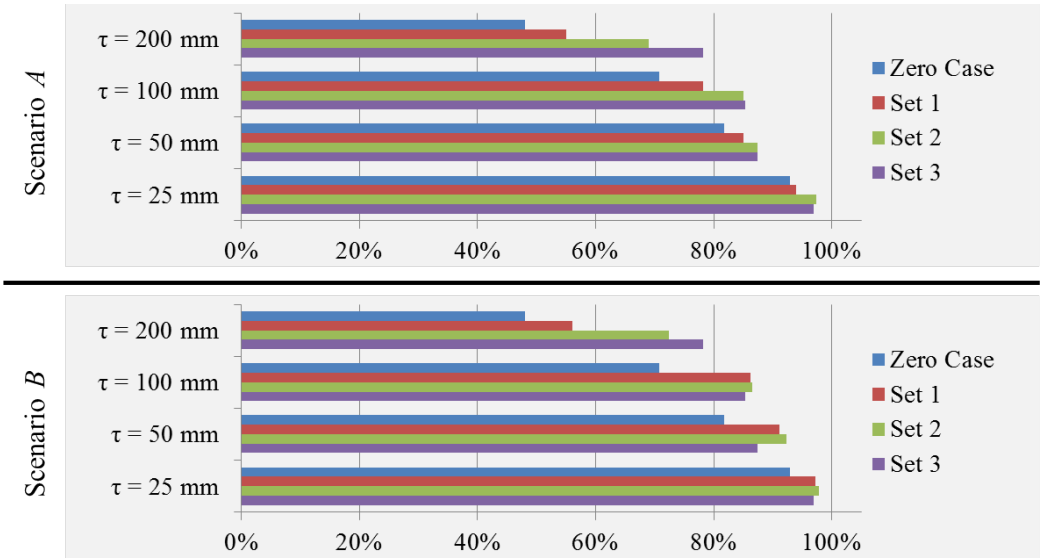

Figure 16: Mean success rates $\dot{\Upsilon}$ for scenarios A (top) and B (bottom) of setup 1, grouped with respect to voxel grid size τ. Blue bars indicate the zero case without any additional plane.

for the zero case without any additional plane that breaks symmetry almost completely fails for most scan pairs (Fig. 17). However, the more the plane is moved into the room and breaks the symmetry, the higher the success rate (Fig. 18). As expected complete or near-complete symmetry causes matching failures. As asymmetry is introduced matching success gradually increases, but from the experiment we conclude that a significant amount of asymmetry is needed to reach high success rates.

8. Conclusions and outlook

We have presented and analyzed the *Keypoint-based 4-Points Congruent Sets* (*K-4PCS*) method for coarse registration of TLS point clouds. The strategy, to represent original raw scans as sparse clouds of 3D keypoints (DoG, Harris) and register these with an adapted version of Aiger's 4PCS approach yields high success rates at reasonable computation times, as long as the scanned scene is not too repetitive or symmetric. Experimental results on multiple indoor and outdoor test data sets have shown sufficient geometrical registration accuracy for a subsequent ICP refinement.

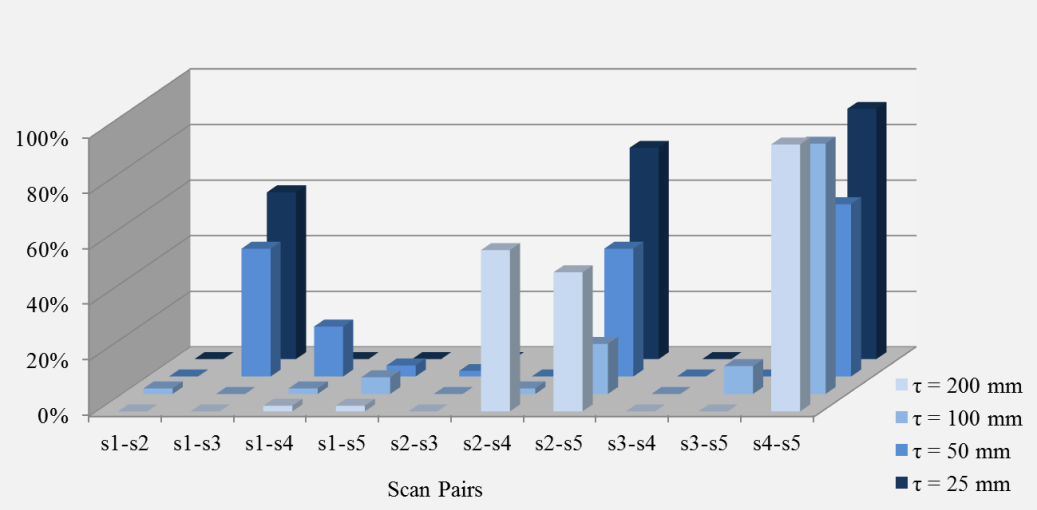

Figure 17: Success rates $\dot{\Upsilon}$ (vertical axis) of the zero case of setup 2 per scan pair (horizontal axis), as a function of voxel size τ (depth/color).

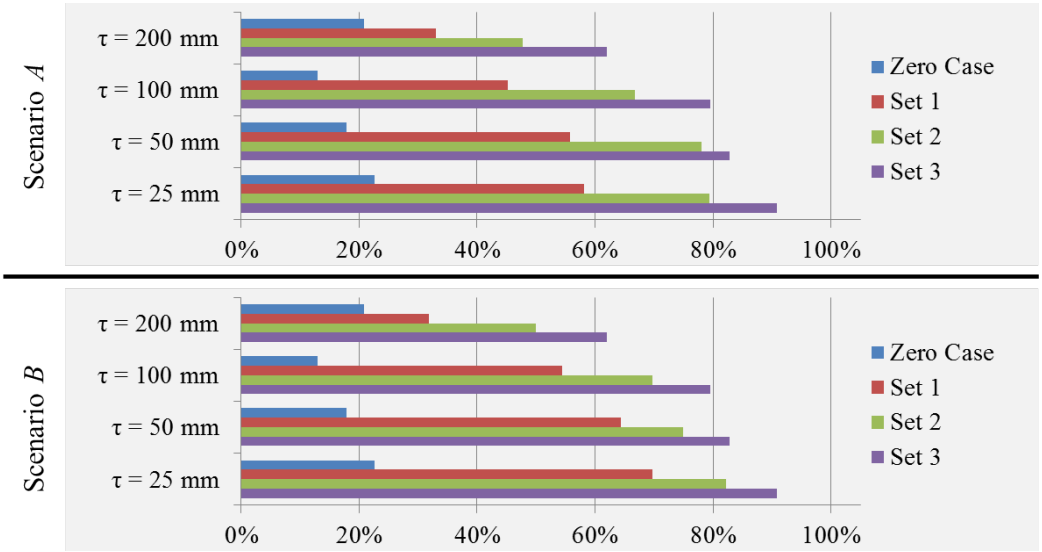

Figure 18: Mean success rates $\dot{\Upsilon}$ for scenarios A (top) and B (bottom) of setup 2 grouped with respect to voxel grid size τ. Blue barsindicate the zero case without any additional plane.

Two main challenges remain for future work: (i) to further bring down processing times, and (ii) to improve robustness against scenes with transla-

tional or rotational symmetries.

The straight-forward parallelization over independent random samples already yields realistic runtimes in most cases. Still, low overlap combined with the necessity to use a large number of keypoints lead to processing time $> 10\,min$. One possibility to further reduce computation times is to adopt a multi-scale approach, in order to solve the bulk of the work with fewer points and iteratively refine the registration with more data. A further idea to be explored is to cluster candidate solutions based on the transformation parameters, with the hope that a small number of promising candidate transformations will emerge quicker and fewer samples are wasted.

Experiments with real as well as synthetic scans demonstrate that robustness to repetitive and symmetric structures must still be improved. While extreme cases do not have a unique solution, and are difficult even for humans, we still lose cases that do contain the necessary structure. One possible direction is to augment the keypoints with descriptors, so as to overcome geometric ambiguity with local appearance information. A further idea is to include prior information, such as expected or approximate distances between scans.

Finally, on a more conceptual note, it would be interesting to find better measures for the goodness of the solution than the number of matching points. Our experiments show that there are cases where a larger overlap can be achieved with a wrong solution, i.e., the (widely employed) objective function does not yet adequately model all aspects of the registration problem.

References

Aiger, D., Mitra, N.J., Cohen-Or, D., 2008. 4-points congruent sets for robust pairwise surface registration. ACM Transactions on Graphics 27 (3), 1–10.

Akca, D., 2003. Full automatic registration of laser scanner point clouds, in: Proc. Optical 3D Measurement Techniques VI, pp. 330–337.

Allaire, S., Kim, J., Breen, S., Jaffray, D., Pekar, V., 2008. Full orientation invariance and improved feature selectivity of 3D SIFT with application to medical image analysis, in: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 1–8.

Bae, K.H., 2009. Evaluation of the convergence region of an automated registration method for 3D laser scanner point clouds. Sensors 9 (1), 355–375.

Bae, K.H., Lichti, D., 2004. Automated registration of unorganised point clouds from terrestrial laser scanners. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 35 (Part B5), 222–227.

Bergevin, R., Soucy, M., Gagnon, H., Laurendeau, D., 1996. Towards a general multi-view registration technique. IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (5), 540–547.

Besl, P., McKay, N.D., 1992. A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (2), 239–256.

Böhm, J., Becker, S., 2007. Automatic marker-free registration of terrestrial laser scans using reflectance features, in: Proc. Optical 3D Measurement Techniques VIII, pp. 338–344.

Brenner, C., Dold, C., 2007. Automatic relative orientation of terrestrial laser scans using planar structures and angle constraints. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 3/W52), 84–89.

Brenner, C., Dold, C., Ripperda, N., 2008. Coarse orientation of terrestrial laser scans in urban environments. ISPRS Journal of Photogrammetry and Remote Sensing 63 (1), 4–18.

Censi, A., 2008. An ICP variant using a point-to-line metric, in: Proc. IEEE International Conference on Robotics and Automation, pp. 19–25.

Chen, Y., Medioni, G., 1992. Object modelling by registration of multiple range images. Image and Vision Computing 10 (3), 145–155.

Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the Association for Computing Machinery 24 (6), 381–395.

Flint, A., Dick, A., van den Hangel, A., 2007. Thrift: Local 3D structure recognition, in: Proc. 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications, pp. 182–188.

Flitton, G., Breckon, T., Megherbi Bouallagu, N., 2010. Object recognition using 3D SIFT in complex CT volumes, in: Proc. British Machine Vision Conference, pp. 11.1–11.12.

Franaszek, M., Cheok, G., Witzgall, C., 2009. Fast automatic registration of range images from 3d imaging systems using sphere targets. Automation in Construction 18 (3), 265–274.

Harris, C., Stephens, M., 1988. A combined corner and edge detector, in: Proc. 4th Alvey Vision Conference, pp. 147–151.

Huttenlocher, D., 1991. Fast affine point matching: an output-sensitive method, in: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR '91, pp. 263–268.

Irani, S., Raghavan, P., 1996. Combinatorial and experimental results for randomized point matching algorithms, in: Proc. 12th Annual Symposium on Computational Geometry, pp. 68–77.

Kang, Z., Li, J., Zhang, L., Thao, Q., Zlatanova, S., 2009. Automatic registration of terrestrial laser scanning point clouds using panoramic reflectance images. Sensors 9 (4), 2621–2646.

Lo, T.W.R., Siebert, J.P., 2009. Local feature extraction and matching on range images: 2.5D SIFT. Computer Vision and Image Understanding 113 (12), 1235–1250.

Lowe, D., 1999. Object recognition from local scale-invariant features, in: Proc. 7th IEEE International Conference on Computer Vision, pp. 1150–1157.

Minguez, J., Montesano, L., Lamiraux, F., 2006. Metric-based iterative closest point scan matching for sensor displacement estimation. IEEE Transactions on Robotics 22 (5), 1047–1054.

Novák, D., Schindler, K., 2013. Approximate registration of point clouds with large scale differences. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 2 (5/W2), 211–216.

Pottmann, H., Huang, Q.X., Yang, Y.L., Hu, S.M., 2006. Geometry and convergence analysis of algorithms for registration of 3D shapes. International Journal of Computer Vision 67 (3), 277–296.

Rusu, R., Blodow, N., Beetz, M., 2009. Fast point feature histograms (FPFH) for 3D registration, in: Proc. IEEE International Conference on Robotics and Automation, pp. 3212–3217.

Rusu, R., Cousins, S., 2011. 3D is here: Point cloud library (PCL), in: Proc. IEEE International Conference on Robotics and Automation, pp. 1–4.

Theiler, P.W., Schindler, K., 2012. Automatic registration of terrestrial laser scanner point clouds using natural planar surfaces. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 1 (3), 173–178.

Theiler, P.W., Wegner, J.D., Schindler, K., 2013. Markerless point cloud registration with keypoint-based 4-points congruent sets. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 2 (5/W2), 283–288.

Yang, J., Li, H., Jia, Y., 2013. Go-ICP: solving 3D registration efficiently and globally optimally, in: Proc. IEEE International Conference on Computer Vision, pp. 1457–1464.

Zeisl, B., Köser, K., Pollefeys, M., 2013. Automatic registration of RGB-D scans via salient directions, in: Proc. IEEE International Conference on Computer Vision, pp. 2808–2815.

Zhang, Z., 1994. Iterative point matching for registration of free-form curves and surfaces. International Journal of Computer Vision 13 (2), 119–152.