

# MARKERLESS POINT CLOUD REGISTRATION WITH KEYPOINT-BASED 4-POINTS CONGRUENT SETS

P. W. Theiler, J. D. Wegner, K. Schindler

Institute of Geodesy and Photogrammetry, ETH Zürich, 8093 Zürich, Switzerland  
(pascal.theiler, jan.wegner, konrad.schindler)@geod.baug.ethz.ch

Commission III/2

**KEY WORDS:** point cloud registration, laser scanning, 3D feature extraction

## ABSTRACT:

This paper addresses the registration of LiDAR point clouds. More specifically, we present an automatic method for markerless registration of two such point clouds given in arbitrary local scan coordinates – i.e. without simplifying assumptions such as a common up-vector. Clearly, the critical step of the registration is to find a coarse initial alignment, to be refined with established local methods for fine registration, such as ICP or least-squares surface matching. The proposed approach builds on the *4-Points Congruent Sets* (4PCS) algorithm (Aiger et al., 2008), a popular registration tool in computer graphics, and extends it to better deal with the specific challenges of LiDAR data. The main limitations of the original 4PCS method in that context are (i) that it does not cope well with strongly varying point densities, such as they routinely occur in laser scans due to the constant angular sampling from different viewpoints; and (ii) that to remain efficient, huge LiDAR point clouds must be down-sampled so heavily that approximate point-to-point correspondence can no longer be guaranteed. To overcome these drawbacks we propose not to apply 4PCS to the original point cloud (respectively, a randomly or regularly subsampled version of it), but rather to represent the point clouds with sets of distinctive 3D keypoints, and run (a slightly modified) 4PCS on the keypoint sets. The resulting combination, termed *Keypoint-based 4-Points Congruent Sets* (K-4PCS), proves to be very reliable: with suitable parameter settings, tests in indoor as well as outdoor environments yield 100% success rates.

## 1 INTRODUCTION

Terrestrial laser scanning (TLS) has become a standard method to acquire 3D models in various applications, including manufacturing, surveying, archaeology and medicine. With ever-increasing scan resolutions and project sizes, there is a growing need to automate the processing of the scan data. TLS are line-of-sight instruments, thus multiple scans are typically needed to fully cover an object or region of interest. Consequently, one of the basic steps of data processing is the registration of these individual scans into a common coordinate frame. Since LiDAR point clouds have known metric scale, registration of a new scan to a point cloud amounts to finding a 6 DOF rigid-body transformation. The problem is commonly solved in two steps, a coarse initial alignment followed by a fine registration.

For **fine registration** many computational solutions exist, which are mostly variants of the standard principle to minimize the Euclidean distances between nearby points. The most popular approach is the *Iterative Closest Point* (ICP) algorithm Besl and McKay (1992) and its numerous variants (e.g. Bergevin et al., 1996; Bae and Lichteni, 2004; Minguez et al., 2006; Censi, 2008). ICP alternates between establishing point correspondence with nearest-neighbor search and updating the transformation parameters using those correspondences. For an overview of different variants see Rusinkiewicz and Levoy (2001). All optimal fine registration methods have in common that they perform local minimization of a non-convex error function and thus require a sufficiently accurate initial alignment in order to converge to the desired solution. For an analysis of the convergence properties see for example (Pottmann et al., 2006; Bae, 2009). In general the convergence basin is much too narrow to start from raw scan data in sensor coordinates.

**Coarse alignment** thus serves the purpose to establish a rough initial solution from which to start the fine registration. It is com-

monly done manually, or by placing targets in the scene which are geometrically well-defined and have high contrast, such that they can be detected automatically (e.g. Akca, 2003; Franaszek et al., 2009). While artificial targets afford high accuracy and robustness, they have obvious drawbacks: placing markers in the scene such that all scan overlaps are covered complicates field recording; the markers will inevitably occlude (small) parts of the scene; and in some cases the markers are not acceptable in the 3D models, such that they must be removed in post-processing. Automatic coarse alignment is thus desirable, but challenging for several reasons: First, laser scans routinely have at least millions of points, therefore computationally expensive methods are intractable. Second, the typical recording setup to minimize field work entails large baselines and limited overlap between scans. And third, the angular sampling implies a quadratic fall-off of the point density with distance, meaning that the same surface will have very different point densities in different scans.

Much of the research about markerless (coarse) point cloud alignment starts by extracting a sparse set of **features** from the raw point data, which is then followed by feature matching between scans. Popular features are planar surfaces (Dold and Brenner, 2006; Theiler and Schindler, 2012), 2D keypoints in the range or intensity images (Böhm and Becker, 2007; Kang et al., 2009), and 3D keypoints, usually found by generalizing 2D feature detectors to 3D (Flint et al., 2007; Allaire et al., 2008; Lo and Siebert, 2009). Note that keypoints found with 2D methods tend to be clustered around the scan position because in the near-field more geometric detail is captured.

**Matching** is done either with descriptors computed from keypoints' local neighborhoods (Böhm and Becker, 2007; Rusu et al., 2009), or with geometric invariants such as pairwise keypoint distances (Theiler and Schindler, 2012) or congruent triangles (Irani and Raghavan, 1996). 2D descriptors like SIFT are of limited use for laser scans: many studies confirm that they can

handle viewpoint differences only up to  $\approx 25$  degrees, which is often exceeded in laser scanning projects. 3D descriptors are hard to compute reliably if the point density is very uneven, as in TLS data.

Matching based on geometric constraints amounts to finding congruent point configurations in the two point clouds to establish correspondence, and is computationally demanding: the minimal geometric configuration to define the transformation are congruent point triplets in the two scans, and even when using randomization in both the sampling and the verification steps, searching for congruent triangles in clouds of  $n$  points has at least computational complexity  $\mathcal{O}(n^3 \log n)$  (Irani and Raghavan, 1996). Thus, even when using RANSAC-style sampling (Fischler and Bolles, 1981), a brute-force search for congruent point sets is only feasible if putative (albeit noisy and ambiguous) point-to-point correspondence has already been established, – which again requires descriptors. Such a combination of descriptor matching and sampling has been called *Sample Consensus Initial Alignment* (SAC-IA; Rusu et al., 2009). A main inspiration for the present work has been the finding of Aiger et al. (2008) that the computational cost of geometric matching can be reduced to  $\mathcal{O}(n^2)$  by exploiting the affine properties of 4 coplanar points. Their *4-points congruent sets* (4PCS) method efficiently aligns uniformly distributed point clouds – see section 2.2.

To summarize, the aim of the work presented here is a robust and efficient procedure for fully automatic, marker-free point-cloud alignment, which is accurate enough to serve as initialization for standard ICP. We propose a combination of the 4PCS method with ideas from 3D keypoint extraction, resulting in a method we call *Keypoint based 4-Points Congruent Sets* (K-4PCS). In particular, we propose to represent a LiDAR point cloud with a set of well-distributed, stable keypoints found with the 3D difference-of-Gaussians (DoG) detector. The resulting keypoint sets then serve as input for 4PCS. As will be shown in the experiments, 4PCS together with the compact keypoint representation make the method computationally efficient, while the use of repeatable, well-distributed keypoints ensure reliable and accurate alignment in spite of the compact representation, even in the presence of low overlaps.

## 2 COARSE POINT-CLOUD REGISTRATION

Figure 1 shows the workflow of the proposed K-4PCS method. Given two laser scanner point clouds, 3D DoG keypoints are extracted from both point clouds, after resampling them to regular point density for a more uniform coverage. This step is described in Sec. 2.1. The extracted keypoint sets serve as inputs for an adapted 4PCS correspondence search, described in Sec. 2.2. Matching results in a set of corresponding keypoints, from which the rigid-body transformation for alignment is estimated, and optionally refined with ICP.

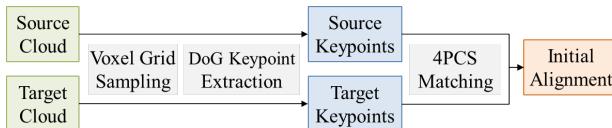


Figure 1: Workflow of the proposed K-4PCS algorithm

The framework is generic and can be combined with different keypoint detectors, the important property being that they should return true 3D features, not forward-projected 2D features, so as to ensure viewpoint invariance and an unbiased distribution. Here, we restrict ourselves to 3D-DoG features (local maxima of the 3D difference-of-Gaussian response), a natural choice in the

presence of large viewpoint changes, which routinely occur in TLS applications. In the following, both the 3D keypoint extraction and the adapted 4PCS constraint matching are explained in more detail.

### 2.1 DoG Keypoint Extraction

The difference-of-Gaussians keypoint detector in 2D was proposed by Lowe (1999), as part of his SIFT keypoint matching framework, and since has widely been used in image processing and machine vision. Scale invariant keypoints are detected by finding local maxima of the DoG response in  $(x, y, scale)$ -space. The DoG response is the difference between adjacent scales of the image pyramid having different amounts of Gaussian blur, and approximates the scale-normalized Laplacian, i.e. it fires at points where the image brightness has high curvature.<sup>1</sup>

Since laser scanner points come with a measured intensity of the reflected laser signal, one can readily apply the DoG method to them. Given the polar nature of laser scans it would be in principle possible to run standard 2D DoG on the corresponding panoramic image. However, we prefer to detect keypoints in the 3D point cloud, which yields a more uniform distribution over the depth range of the scan, and thus on average also over the overlap area of different scans. To evade the near-field bias due to the non-uniform point density, we resample the raw point cloud with a *Voxel Grid* filter: the scan volume is divided into a regular 3D voxel grid, and only one point per voxel is retained, computed as the centroid of the points inside a grid cell, see Fig. 2. Calculating the centroid – instead of the more commonly used voxel centre – improves the approximation accuracy w.r.t. the original point cloud, especially for voxels with irregular point distribution. As mentioned in Sec. 1, there have been numerous attempts to extend the SIFT approach of Lowe to 3D space. We point out that in our scheme DoG keypoints are used as input for geometrical constraint matching, thus no 3D descriptors are needed.

Just like in the 2D case, our keypoint extraction is based on the calculation of differences of Gaussians, only this time in 3D voxel space. I.e. a 3D scale-space of  $q$  levels is generated by convolution of the voxel grid  $I(x, y, z)$  with a 3D Gaussian kernel  $G(x, y, z, \sigma)$

$$L(x, y, z, k_i\sigma) = I(x, y, z) * G(x, y, z, k_i\sigma), \quad i = 1 \dots q \quad (1)$$

and the difference between neighbouring scales forms the DoG-response,

$$DoG(x, y, z, k_i\sigma) = L(x, y, z, k_{i+1}\sigma) - L(x, y, z, k_i\sigma). \quad (2)$$

Keypoints are extracted from the DoG-space by finding voxels which are local minima or maxima in  $(x, y, z, scale)$ -space and whose (absolute) DoG value exceeds a threshold. An example of extracted DoG keypoints is visualized in Fig. 2.

### 2.2 4-Points Congruent Sets

The *4-Points Congruent Sets* (4PCS) algorithm of Aiger et al. (2008) is a method to robustly align two partially overlapping point clouds with arbitrary orientations. To understand 4PCS, first consider a straight-forward random sampling strategy: given a randomly sampled triplet from  $I_1$ , we need to find a congruent triplet in  $I_2$ . From the two triangles the alignment can be computed and applied to the target point cloud to obtain  $I'_1$ , which is then tested for agreement with  $I_2$ . This triangle sampling is repeated until a transformation with large enough support is found.

<sup>1</sup>Note, sometimes DoG keypoints are also called SIFT keypoints; we prefer the name DoG, since SIFT is nowadays more commonly used for the gradient histogram descriptor devised by Lowe in the same work.

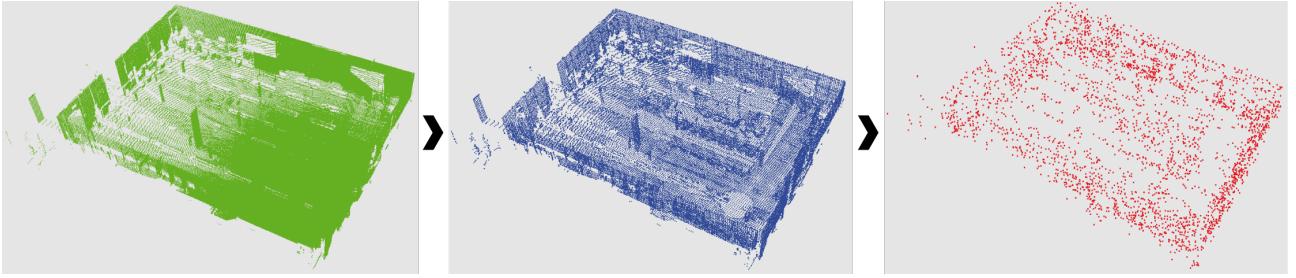


Figure 2: 3D keypoint extraction. The original scan point cloud (green) is resampled with a voxel grid filter, and the resulting uniform density cloud (blue) serves as input for the extraction of DoG keypoints (red).

Unfortunately such a procedure will be rather inefficient, having computational complexity at least  $\mathcal{O}(n^3 \log n)$  (Irani and Raghavan, 1996).

The concept behind 4PCS is to sample base sets  $B(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) \in I_1$  of 4 approximately *coplanar* points, rather than minimal sets of 3 points, and search for corresponding sets  $M(\mathbf{a}', \mathbf{b}', \mathbf{c}', \mathbf{d}') \in I_2$  (Fig. 3). Somewhat surprisingly, adding a fourth point *reduces* the runtime to  $\mathcal{O}(n^2)$  – see below. Random sampling is biased towards selecting distant points to ensure stable and reliable transformation parameters. In case of partially overlapping point clouds, the maximum distance between base points is calculated from the approximate overlap and diameter of the inputs. While the diameter can be automatically estimated, the approximated overlap needs to be set a priori.

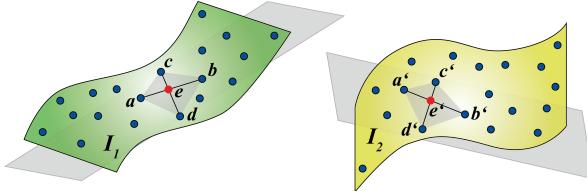


Figure 3: Principle of 4PCS, with  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in I_1$  the base point set and  $\mathbf{a}', \mathbf{b}', \mathbf{c}', \mathbf{d}' \in I_2$  the corresponding congruent set.

The insight behind 4PCS is that in a planar quadrangle of 4 points (the base set) one can compute the intersection point  $e$  of the diagonals (see Fig. 3). The resulting intersection ratios  $r_1$  and  $r_2$  of the two diagonals w.r.t. that point (eq. 3) are invariant under affine transformation (Huttenlocher, 1991).

$$\begin{aligned} r_1 &= \|\mathbf{a} - \mathbf{e}\| / \|\mathbf{a} - \mathbf{b}\| \\ r_2 &= \|\mathbf{c} - \mathbf{e}\| / \|\mathbf{c} - \mathbf{d}\| \end{aligned} \quad (3)$$

The task of finding a congruent point set in  $I_2$  can be carried out in  $\mathcal{O}(n^2)$ , with the following steps: First, find point pairs  $\{\mathbf{a}', \mathbf{b}'\}, \{\mathbf{c}', \mathbf{d}'\}$  that have (up to a tolerance  $\delta_1$ ) the same point-to-point distances  $\mathbf{x} = \mathbf{b}' - \mathbf{a}'$  and  $\mathbf{y} = \mathbf{d}' - \mathbf{c}'$  as the diagonals of  $B$  in  $I_1$ . Then, predict the two possible diagonal intersection points independently for each point pair:

$$\begin{aligned} \mathbf{e}_{ab} &= \mathbf{a}' + r_1 \cdot \mathbf{x} & ; \quad \mathbf{e}_{ba} &= \mathbf{b}' - r_2 \cdot \mathbf{x} \\ \mathbf{e}_{cd} &= \mathbf{c}' + r_1 \cdot \mathbf{y} & ; \quad \mathbf{e}_{dc} &= \mathbf{d}' - r_2 \cdot \mathbf{y}. \end{aligned} \quad (4)$$

Due to the affine invariant above, for a valid congruent set the predicted points must coincide (again up to a tolerance), i.e. we have to check four combinations:  $\|\mathbf{e}_{ab} - \mathbf{e}_{cd}\| < \delta_2$ ,  $\|\mathbf{e}_{ab} - \mathbf{e}_{dc}\| < \delta_2$ ,  $\|\mathbf{e}_{ba} - \mathbf{e}_{cd}\| < \delta_2$ , and  $\|\mathbf{e}_{ba} - \mathbf{e}_{dc}\| < \delta_2$ . After this test, verification proceeds as in standard RANSAC. A candidate transformation is computed for each of the  $k$  valid congruent sets, and the one with the highest number of inliers (according to a tolerance  $\delta_3$ ) is retained. This 4-point sampling and verification procedure is repeated  $L$  times, with  $L$  determined from the expected

number of point correspondences, given the approximate overlap between the data sets. It should be pointed out that while 4PCS has a number of thresholds (respectively tolerances)  $\delta_i$ , they are all estimated automatically, based on the mean density of the input data.

When using keypoints rather than raw point clouds as input, the data-driven estimation of these tolerances must be adapted to account for the properties of keypoint input. On the one hand, DoG keypoints are a lot sparser than the original point cloud, so the mean point density is low. On the other hand, the stable keypoints have a lot better repeatability (chance of being present in both input point clouds) than randomly sampled points. Thus, the chance of finding the same object point, rather than an arbitrary nearby point, is a lot higher than in point clouds generated by random downsampling to the same size. Estimating the tolerances from average densities or point-to-point distances in the keypoint cloud will therefore result in exceedingly high values, leading to grossly wrong alignments, and thus to a failure of the subsequent ICP. Using the original data will give too low tolerances, causing coarse alignment to fail. We found that the necessary tolerance for well-defined keypoints is independent of the density, and is a constant multiple of the minimum keypoint scale. We therefore calculate all tolerances from the (fixed) minimum scale used in the DoG detector.

### 2.3 Implementation

The described K-4PCS method was implemented in C++, making use of the open source *Point Cloud Library* (PCL, Rusu and Cousins, 2011). Empirically, the most important parameters are the minimum scale for keypoint extraction and the approximate point cloud overlap, whereas others which could potentially have a large influence (voxel grid spacing for filtering, spacing between DoG levels, minimum contrast threshold for DoG) do not greatly change the result when set in a reasonable range. Consequently, only these two parameters were varied in the evaluation, while all others were kept constant (numerical values are given in Tab. 1).

Minimum contrast	1% of brightness range
Number of octaves	5
Number of scale per octave	5
Minimum scale	variable
Voxel grid spacing	same as minimum scale
Point cloud overlap	variable

Table 1: Numerical parameter values used for K-4PCS.

For this project, both the standard 4PCS method and the proposed K-4PCS method were integrated into the PCL environment.

### 3 EVALUATION

To test the proposed K-4PCS method we have captured two datasets, each consisting of multiple scans with varying degrees of overlap. One set of scans covers an indoor scenario with predominantly planar walls (example shown in Fig. 2) and a high degree of symmetry, whereas the other set shows an outdoor scenario with an ancient arch as the object of interest, surrounded by paths, some vegetation and a distant building. The results are compared against those of the standard 4PCS as baseline.

To assess the *metric accuracy* and the *success rate* of our proposed method we compare it to ground truth generated by registering the scans with artificial sphere targets. The ground truth has residual registration errors  $< 5$  mm, and can thus be considered error-free when evaluating coarse alignments. The following error measures are used: to quantify metric accuracy, the differences between the true and estimated transformations are split into a *mean angular error* (MAE) using Euler angles, and a *mean translation error* (MTE). Additionally we also report the *root mean square error* (RMSE) between the point clouds transformed with the true and with the estimated alignment matrix. Note, since the RMSE is computed from two transformed versions of the same point cloud, it is based on true correspondences rather than nearest neighbors. Since coarse registration is not the final goal, a registration is considered successful if the RMSE is below an empirical threshold, chosen conservatively to ensure that the consecutive fine registration with ICP converges to the correct solution (see below).

#### 3.1 Baseline

As a state-of-the-art baseline for automatic point cloud registration we use the standard 4PCS method. It should be noted that a totally fair comparison is not straight-forward: since the method cannot deal with clouds of millions of points, as they are routinely recorded by TLS, the point clouds need to be down-sampled to a maximum size. The authors of the method resort to random subsampling – see Aiger et al. (2008) and associated source code. Unfortunately that strategy breaks down for the case of LiDAR scans: Due to the quadratically decreasing point density a large fraction of the points are concentrated near the respective scanner positions. When downsampling randomly one therefore quickly runs out of (even approximately) corresponding points in the lower-density regions of the scans, which account for most of the recorded scene and most of the overlap area. Empirically, agnostic downsampling to tractable sizes causes 4PCS to fail in almost all cases. Instead, we therefore downsample the point cloud with the same voxel grid filter also employed in our scheme, but using a wider grid spacing. The spacing is chosen such that the resulting point cloud has approximately the same number of points as our corresponding keypoint set (but distributed uniformly across the scan volume), and thus registration times become comparable. Running 4PCS with “intelligent” point cloud reduction to the same size works a lot better, and appears to be the most sensible direct comparison (i.e. in case of computation time).

#### 3.2 Indoor Application

A first evaluation was carried out using a set of 5 scans from a laboratory room. The scans have been acquired with the *Zoller+Fröhlich TLS Imager 5006i* and each have  $\approx 10$ M points, covering a field of view of  $360^\circ$  horizontally and  $150^\circ$  vertically. The size of the scanned room is about  $15 \times 10$  metres, and the distances between two scan positions range from 3 to 15 metres. Apart from standard office furniture (desks, chairs, whiteboards)

there are a number of cylindrical pillars. Although the environment is rather simple in terms of scene geometry, alignment is relatively challenging because of multiple rotational symmetries of the major structures, which all give rise to solutions with high overlap. We found that, due to the high degree of symmetry, ICP only converges to a correct solution reliably if the RMSE of the coarse alignment is below 0.5 m, hence that value is used as hard threshold for a successful registration. In order to test somewhat extreme cases with low overlap and/or significant density change and occlusions, two scans were deliberately captured from opposite corners of the room, as well as from different heights.

The approximate overlap was set to 80% for all scan pairs, reflecting the fact that all scans see the entire room, but with some degree of occlusion by furniture. To even out fluctuations due to random sampling, the alignment was repeated 10 times for each of the 10 scan pairs. In order to test the influence of different minimum scale  $\sigma$  (respectively number) of keypoints, all tests were run for three different settings  $\sigma = \{25\text{ mm}, 50\text{ mm}, 100\text{ mm}\}$ .

	$S = 0.7\text{M}$		$S = 2.5\text{M}$		$S = 10.0\text{M}$	
	KP	SR	KP	SR	KP	SR
$\sigma = 100\text{mm}$	1100	91%	1100	95%	1200	95%
$\sigma = 50\text{mm}$	2800	98%	2800	98%	3200	98%
$\sigma = 25\text{mm}$	5900	98%	6800	97%	8200	100%

Table 2: Mean number of extracted DoG keypoints (KP), success rate (SR) at different voxel grid sizes  $\sigma$  and different point cloud sizes  $S$ . Keypoint numbers slightly vary across different scans and are rounded to steps of 100.

Furthermore, we also tested whether the size/resolution of the original scan has an influence. To that end the original point cloud was downsampled by factors of 4 and 16 *before starting our registration scheme*, resulting in 3 different resolutions  $S = \{10.0\text{M}, 2.5\text{M}, 0.7\text{M}\}$ . Downsampling is applied to the raw range images, thus preserving the irregular point distribution of laser scans. Table 2 shows the number of extracted keypoints (KP) and the corresponding rate of successful registrations (SR) over all scan pairs. As expected, the minimum scale (respectively voxel grid spacing) has an impact on the number of extracted keypoints, and also on the success rate. The scan resolution on the other hand has little influence on the number of keypoints, but nevertheless higher point density yields a bit higher success rates.

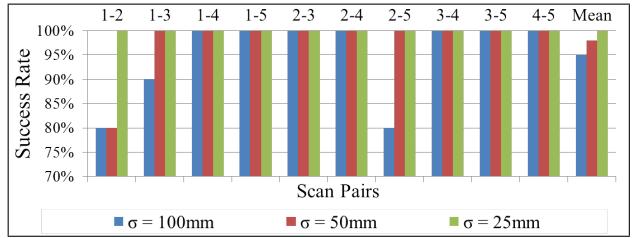


Figure 4: Success rates for varying minimum scale  $\sigma$ , using the full point cloud resolution of  $S = 10\text{M}$  points.

Overall, more than 90% of all alignments are successful even at the lowest resolution and scale, and that number improves to  $> 97\%$  for keypoint scales of at least 50 mm, and reaches 100% at the highest tested resolution and scale. Note that with decreasing minimum scale, the computation time increases (see Tab. 3). We point out that all incorrect alignments without exception are caused by the symmetry, i.e. they are  $180^\circ$  rotated around one of the room’s main axes. Also as expected, the extreme case of point clouds acquired from opposite corners, with only little overlap, has the lowest success rate (see Fig. 4), but is still registered correctly in 80% of all trials even at low resolution.

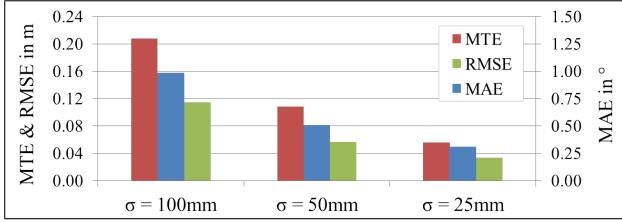


Figure 5: Accuracy of the resulting transformation for varying minimum scale  $\sigma$ , using the full point cloud resolution of  $S = 10\text{M}$  points. Results are similar at lower resolutions.

Figure 5 shows that all registration accuracies (MAE, MTE and *RMSE*) depend on the minimum DoG scale and thus on the voxel grid spacing. Decreasing the minimum scale results in more as well as more accurately localized keypoints, which naturally improves transformation parameters, too.

We compare success rate (*SR*), geometrical accuracy (*RMSE*) and computation time (*T*) of our proposed K-4PCS method with the 4PCS baseline (Tab. 3). As already mentioned in 3.1, scan point clouds were filtered to a more uniform distribution prior to applying the original 4PCS algorithm. Consequently, the *low*, *medium* and *high* values for  $\sigma$  are different from those used in K-4PCS.

	4PCS			K-4PCS		
	<i>SR</i>	<i>RMSE</i>	<i>T</i>	<i>SR</i>	<i>RMSE</i>	<i>T</i>
$\sigma_{high}$	4%	0.66m	18s	91%	0.15m	8s
$\sigma_{med}$	35%	0.46m	589s	98%	0.06m	41s
$\sigma_{low}$	62%	0.28m	9637s	98%	0.04m	102s

Table 3: Comparison between the 4PCS and the K-4PCS in terms of success rate (*SR*), accuracy (*RMSE*) and computation time (*T*), based on low-resolution ( $S = 0.7\text{M}$ ) indoor point cloud.

K-4PCS clearly outperforms 4PCS in terms of success rate, accuracy as well as computational time (Tab. 3, K-4PCS example shown in Fig. 6). The significantly lower success rates of 4PCS are caused by the coarse voxel grid filters required to yield tractable point clouds ( $\sigma_{high} = 900 \text{ mm}$ ,  $\sigma_{med} = 500 \text{ mm}$ ,  $\sigma_{low} = 350 \text{ mm}$ ). The resulting registrations are too inaccurate for ICP. A closer inspection reveals that most unsuccessful trials of 4PCS occur because it is often not able to correctly handle the high degree of symmetry of the room. K-4PCS shows a much higher robustness with respect to symmetries, which is owed to the DoG keypoints. Moreover, K-4PCS always achieves considerably higher geometrical registration accuracies (4–7 times lower *RMSE*), while at the same time significantly speeding up the computation (by factors of 2–95). Computation times of K-4PCS are easily acceptable even with low  $\sigma$ , whereas the runtimes of 4PCS quickly become too long for practical purposes.

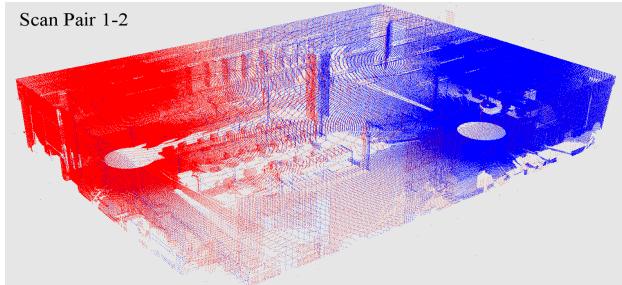


Figure 6: Example of a successful registration using K-4PCS,  $\sigma = 50\text{mm}$  and a low resolution cloud with  $S = 0.7\text{M}$  points.



Figure 7: Outdoor environment with Roman arch, paths, vegetation, and a distant building. Note the numerous artefacts caused by walking people.

### 3.3 Outdoor Application

We additionally tested K-4PCS to register scans acquired outdoors. The scene (Fig. 7) contains a Roman arch surrounded by vegetation (trees, lawns), paths, and a distant building. We took four scans with an approximate neighboring scan overlap of  $\approx 40\%$ , each taken from a different angle to the arch. What makes this dataset challenging are small scan overlaps and moving people, who visited the arch during data acquisition. As the environment (like most outdoor recordings) exhibits a less symmetric layout than the indoor scene, the *RMSE* threshold for successful registration could be raised to 1 m. Due to the recording setup, only three pairs overlap sufficiently for registration. Again, to even out the fluctuations of random sampling, each registration was repeated 10 times, resulting in a total of 30 trials.

Since the indoor tests suggest that scan resolution only has a minor influence on registration quality, we ran all outdoor tests with point clouds at a single (medium) resolution ( $S = 1.5\text{M}$  points). Tab. 4 shows the comparison of *SR*, *RMSE* with respect to groundtruth and *T* between K-4PCS and 4PCS for different minimum DoG scales respective voxel grid sizes. For K-4PCS, the minimum scales are equal to the ones in the indoor evaluation. To receive point clouds of equal size as input for 4PCS, the voxel grid size was set to  $\sigma_{high} = 1500 \text{ mm}$ ,  $\sigma_{med} = 1000 \text{ mm}$ ,  $\sigma_{low} = 700 \text{ mm}$ .

	4PCS			K-4PCS			
	<i>SR</i>	<i>RMSE</i>	<i>T</i>	<i>SR</i>	<i>RMSE</i>	<i>T</i>	<i>KP</i>
$\sigma_{high}$	10%	0.80m	652s	57%	0.52m	182s	2900
$\sigma_{med}$	30%	0.79m	9800s	100%	0.40m	663s	6000
$\sigma_{low}$	N.A.	N.A.	N.A.	83%	0.17m	2047s	10400

Table 4: Comparison between 4PCS and K-4PCS in terms of success rate (*SR*), accuracy (*RMSE*) and computation time (*T*), based on medium resolution ( $S = 1.5\text{M}$ ) outdoor point cloud.

K-4PCS generally outperforms 4PCS. It is faster, more accurate and achieves a much higher success rate (Tab. 4). Compared to the indoor scans computation time *T* of 4PCS as well as K-4PCS increases (cf. Tab. 3) due to the low scan overlap of only  $\approx 40\%$ . On one hand, the required number  $L$  of sampled 4-point bases increases with decreasing overlap; on the other hand the outdoor scans extend over a larger area, resulting in more DoG keypoints for K-4PCS. The computation time of 4PCS exceeds the practical limits already at a medium scale  $\sigma_{med}$ . Tests with 4PCS at the lowest scale did not converge in 72 hours.

The geometrical registration accuracy (*RMSE*) of K-4PCS increases with decreasing minimum DoG scale, as for indoor scenes. Note that the percentage of successful registrations using DoG keypoints at  $\sigma = 25 \text{ mm}$  is lower than at  $\sigma = 50 \text{ mm}$ . Most

likely, that effect is caused by the relatively low input point density, especially in the overlap areas. If that point density is below the minimum DoG scale, more keypoints are extracted in the non-overlapping parts, leading to a greater fraction of mismatches. A successful registration of two outdoor scans with K-4PCS is shown in Fig. 8.

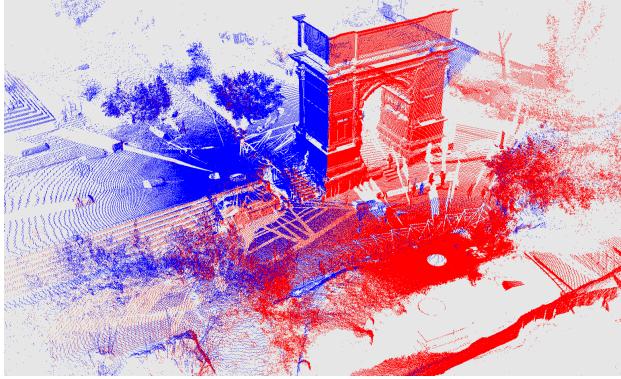


Figure 8: Registration of point clouds with low overlap in a complex outdoor environment with K-4PCS.

#### 4 CONCLUSIONS AND OUTLOOK

We have proposed an effective and efficient model for the coarse registration of point clouds acquired with terrestrial laserscanners based on DoG key points and 4PCS. Our approach significantly outperforms the baseline in terms of success rate, geometrical accuracy and computation time.

In the future we plan to conduct experiments with larger and more diverse indoor and outdoor datasets. For example, long hallways will result in long but narrow scans leading to additional challenges. More experiments with multiple scans and less overlap will also allow more insights into the robustness of K-4PCS and its potential for large practical projects (e.g. in large-scale construction). We point out that the K-4PCS method is not *per se* limited to terrestrial laser scans. We plan to adapt it to airborne LiDAR data as well as to point clouds from dense multiview image matching.

A limitation at present is the lack of robustness when the environment is nearly symmetric. Thus, we aim to further improve K-4PCS in such a way that those sets of congruent points are selected, which are most useful to break symmetries. Furthermore, recall that mismatches of outdoor scans were mainly due to overly low point density in the overlap area. To address this issue, we plan to investigate ways of enforcing a more even key-point distribution, regardless of scan point density.

#### References

- Aiger, D., Mitra, N. J. and Cohen-Or, D., 2008. 4-points congruent sets for robust pairwise surface registration. *ACM T. Graphics* 27(3), pp. 1–10.
- Akca, D., 2003. Full automatic registration of laser scanner point clouds. In: *Optical 3D Measurement Techniques VI*.
- Allaire, S., Kim, J., Breen, S., Jaffray, D. and Pekar, V., 2008. Full orientation invariance and improved feature selectivity of 3D SIFT with application to medical image analysis. In: *Workshops, IEEE Computer Vision and Pattern Recognition*.
- Bae, K.-H., 2009. Evaluation of the convergence region of an automated registration method for 3D laser scanner point clouds. *Sensors* 9(1), pp. 355–375.
- Bae, K.-H. and Lichten, D., 2004. Automated registration of unorganised point clouds from terrestrial laser scanners. In: *ISPRS Archives*, Vol. XXXV.
- Bergevin, R., Soucy, M., Gagnon, H. and Laurendeau, D., 1996. Towards a general multi-view registration technique. *IEEE T. Pattern Anal.* 18(5), pp. 540–547.
- Besl, P. J. and McKay, N. D., 1992. A method for registration of 3-D shapes. *IEEE T. Pattern Anal.* 14(2), pp. 239–256.
- Böhm, J. and Becker, S., 2007. Automatic marker-free registration of terrestrial laser scans using reflectance features. In: *Optical 3D Measurement Techniques VIII*.
- Censi, A., 2008. An ICP variant using a point-to-line metric. In: *Int'l Conf. on Robotics and Automation*.
- Dold, C. and Brenner, C., 2006. Registration of terrestrial laser scanning data using planar patches and image data. In: *ISPRS Archives*, Vol. XXXVI.
- Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24(6), pp. 381–395.
- Flint, A., Dick, A. and Hengel, A. v. d., 2007. Thrift: Local 3D structure recognition. In: *Digital Image Computing Techniques and Applications*.
- Franaszek, M., Cheok, G. and Witzgall, C., 2009. Fast automatic registration of range images from 3D imaging systems using sphere targets. *Automation in Construction*.
- Huttenlocher, D. P., 1991. Fast affine point matching: an output-sensitive method. In: *IEEE Computer Vision and Pattern Recognition*.
- Irani, S. and Raghavan, P., 1996. Combinatorial and experimental results for randomized point matching algorithms. In: *12th Annual Symposium on Computational Geometry*.
- Kang, Z., Li, J., Zhang, L., Thao, Q. and Zlatanova, S., 2009. Automatic Registration of Terrestrial Laser Scanning Point Clouds using Panoramic Reflectance Images. *Sensors* 9, pp. 2621–2646.
- Lo, T.-W. R. and Siebert, J. P., 2009. Local feature extraction and matching on range images: 2.5D SIFT. *Comput. Vis. Image Und.* 113(12), pp. 1235–1250.
- Lowe, D., 1999. Object recognition from local scale-invariant features. In: *IEEE Computer Vision and Pattern Recognition*.
- Minguez, J., Montesano, L. and Lamirault, F., 2006. Metric-based iterative closest point scan matching for sensor displacement estimation. *IEEE T. Robotics* 22(5), pp. 1047–1054.
- Pottmann, H., Huang, Q.-X., Yang, Y.-L. and Hu, S.-M., 2006. Geometry and convergence analysis of algorithms for registration of 3D shapes. *Int. J. Comput. Vision* 67, pp. 277–296.
- Rusinkiewicz, S. and Levoy, M., 2001. Efficient variants of the ICP algorithm. In: *3-D Digital Imaging and Modeling, IEEE*.
- Rusu, R. and Cousins, S., 2011. 3D is here: Point cloud library (PCL). In: *Int'l Conf. on Robotics and Automation*.
- Rusu, R., Blodow, N. and Beetz, M., 2009. Fast point feature histograms (FPFH) for 3D registration. In: *Int'l Conf. on Robotics and Automation*.
- Theiler, P. W. and Schindler, K., 2012. Automatic registration of terrestrial laser scanner point clouds using natural planar surfaces. In: *ISPRS Annals*, Vol. I/3.