

Object-Based Incremental Registration of Terrestrial Point Clouds in an Urban Environment

Xuming Ge^{a,b}, Han Hu^{b,*}

^a*Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Land and Resources of China*

^b*Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, China*

Abstract

Registration of terrestrial point clouds is essential for large-scale urban applications. The robustness, accuracy, and runtime are generally given the highest priority in the design of appropriate algorithms. Most approaches that target general scenarios can only fulfill some of these factors, that is, robustness and accuracy come at the cost of increased runtime and vice versa. This paper proposes an object-based incremental registration strategy that accomplishes all of these objectives without the need for artificial targets, aiming at a specific scenario, the urban environment. The key is to decompose the degrees of freedom for the SE(3) transformation to three separate but closely related steps, considering that scanners are generally leveled in urban scenes: (1) 2D transformation with matches from line primitives, (2) vertical offset compensation by robust least-squares optimization, and (3) full SE(3) least-squares refinement using uniformly selected local patches. The robustness is prioritized in the whole pipeline, as structured first by a primitive-based registration and two least-squares optimizations with robust estimations that do not require specific keypoints. An object-based strategy for terrestrial point clouds is used to increase the reliability of the first step by the line primitives, which significantly reduces the search space without affecting the recall ratio. The least-squares optimization contributes to achieve a global optimum for the accurate registration. The three coupling steps are also more efficient than segregated coarse-to-fine registration. Experimental evaluations for point clouds acquired in both a metropolis and in old-style cities reveal that the proposed methods are superior to or *on par* with the state-of-the-art in robustness, accuracy, and runtime. In addition, the methods are also agnostic to the primitives adopted.

Keywords: Point Clouds, Registration, Terrestrial laser scanner, Urban scenes, Line segments, Planar segments

¹ 1. Introduction

Point clouds are standard datasets in the production of digital terrain models, 3D reconstructions, and land use maps (Vosselman et al., 2017; Ge et al., 2019b). An individual scan obtained by a terrestrial laser scanner (TLS) is constrained by the limited field-of-view; therefore, to generate a complete 3D model, all related individual scans should be registered into the same coordinate system by estimating and assigning an *SE*(3) transformation to each point cloud. The registration strategies can be classified into coarse and fine registration in terms of either the initial information required or the goal of the processing (Ge and Wunderlich, 2016).

*Corresponding Author: han.hu@swjtu.edu.cn

9 Although recent studies (Cai et al., 2019; Ge and Wu, 2019; Xu et al., 2019) have discussed the
10 topic of end-to-end solutions of point cloud registration from various perspectives, several core
11 issues require further investigation.

12 1) *The volume of data.* Most registration methods operate at the point level (Ge, 2017; Dong
13 et al., 2018; Cai et al., 2019) due to the ability to generalize various scenarios. Nevertheless, point-
14 based methods must confront a formidable challenge in handling two massive datasets with more
15 than 10 million points, or at least thousands of keypoints. The vast search space necessitates a
16 prohibitively high running time, and various approaches have quite diverse runtime performances
17 with several orders of magnitudes of differences.

18 2) *Distinctiveness of the descriptors.* Most registration methods that use keypoints also
19 require the extraction of high dimensional descriptors for feature matching; for instance, the fast
20 point feature histogram (FPFH) is probably the most widely adopted descriptor for point clouds
21 (Rusu et al., 2009). The descriptors, in general, use the local shape information around the key
22 points; however, in an urban environment dominated by the planar structures, the distinctiveness
23 of the 3D feature descriptors is significantly reduced (Cheng et al., 2018a). More representative
24 features are thus required.

25 3) *Segregated steps.* Most methods address part of the registration problem, that is, either
26 coarse registration or fine registration. Two different strategies are responsible for the segregated
27 steps, and no information is shared between the two steps. However, from a practical perspective,
28 an end-to-end solution is preferred that can provide accurate results with no knowledge of the
29 initial orientation in a consistent framework. In addition, the segregated steps also require the
30 independent solution of two SE(3) transformations, which commonly leads to greater complexity
31 and longer runtimes.

32 To surmount these problems for an efficient, robust and accurate solution in urban envi-
33 ronment, an object-based incremental strategy is proposed as an end-to-end solution for the
34 registration of terrestrial point clouds. Inspired by adoption of the 4-degrees-of-freedom (DoFs)
35 approach to reduce the complexities of the problem, the 6 DoFs in the SE(3) transformation are
36 decomposed into three coupled subproblems: 1) 2D transformation (3 DoFs), 2) vertical offset (1
37 DoF), and 3) global SE(3) refinement (6 DoFs). It turns out that the “3 + 1 → 6” strategy not
38 only reduces the complexities of the problem, but also accelerates convergence. By leveraging
39 the leveling information of the state-of-the-art scanners, we can accelerate the estimation of the
40 2D transformation through much less but equally representative primitives by abstraction of the
41 point clouds into objects that have semantic meaning. In addition, auxiliary information are
42 shared among the three sub-problems, such as the density and index structures, rather than the
43 segregated coarse-to-fine strategy.

44 In summary, the contributions of this study are threefold: (1) a unified framework for end-
45 to-end registration that avoids the incompatibility caused by independent settlements of the
46 coarse-to-fine pipeline; (2) an incremental registration strategy that decomposes the DoFs into
47 three subproblems that are solved efficiently and robustly; and (3) an object-based approach that
48 significantly reduces the primitives without loss of the recall ratio. The reminder of this paper
49 is organized as follows. Section 2 briefly introduces related works, and Section 3 gives a detailed
50 description of the proposed approach. In Section 4, we report the experiments and analyses, and
51 the conclusions are presented in Section 5.

52 2. Related works

53 The registration of point clouds has been extensively studied. Registration strategies can
54 be divided into two categories according to the geometric primitives exploited: *i.e.* point-based
55 methods and primitive-based (*i.e.*, lines, planes) methods.

56 *Point-based approaches.* Point-based approaches are more prevalent due to their feasibility for a
57 range of scenes. For coarse registration, most point-based methods typically include three consecutive
58 steps: (1) extraction of point features, (2) generation of correspondence candidates, and (3)
59 estimation of the transformations. For the first step, keypoints should be extracted from the raw
60 point clouds by certain feature detectors and descriptors. For the detectors, Theiler et al. (2014)
61 used DoG (Difference of Gaussian) key points, and Yang et al. (2016) represented the scanning
62 scenes with the semantic key points. Suitable feature descriptors should then be used to describe
63 the spatial shape around each keypoint, which is generally represented by a high dimensional
64 vector. For example, Holz et al. (2015) and Weber et al. (2015) used FPFH to describe the key
65 points; Lei et al. (2017) designed a fast multiscale descriptor to encode information; Dong et al.
66 (2017) exploited the binary shape context descriptor for feature information, which was also ex-
67 tended to multiview registration (Dong et al., 2018); and Ge and Wu (2019) applied contextual
68 information to generate descriptors. Second, the putative correspondence pools are established
69 from the obtained keypoints. Spatial geometric consistency is one of the critical constraints for
70 establishment of the correspondences, such as the 4PCS-like approaches, which depend mainly
71 on Euclidean distances and scale invariances (Theiler et al., 2014; Ge, 2017). SIFT-like (Lowe,
72 2004) descriptor searching is also a viable option (Zhou et al., 2016; Dong et al., 2018). Finally,
73 robust model-estimation strategies are leveraged to calculate the rigid transformation parameters
74 between the two scans. Random Sample Consensus (RANSAC) (Bolles and Fischler, 1981) is one
75 of the most widely used paradigms, but the efficiency and convergence of the RANSAC-based
76 strategy in registration problems depend entirely upon the outlier ratio (Cai et al., 2019), and
77 the number of iterations (or, equivalently, the runtimes) of the validation procedure is difficult
78 to determine (Ge et al., 2019b). Many modifications other than the classical RANSAC-based
79 approaches have been proposed to improve the efficiencies and robustness of the model estima-
80 tion step. For example, Zhou et al. (2016) assigned a greater weight to the correspondences that
81 were more likely to be a positive match in a robust least-squares manner, which turned out to
82 be one of the most efficient approaches.

83 *Primitive-based approaches.* To overcome the limitations in the point-based approaches, primitive-
84 based approaches generally exploit higher-level geometric primitives, such as lines and planes,
85 for correspondence matching. Stamos and Leordeanu (2003) presented a registration method
86 based on lines extracted from adjacent intersecting planes. Theiler et al. (2012) extracted vir-
87 tual tie points from linear features and registered two scans based on those tie points. Cheng
88 et al. (2018b) used building outlines to align terrestrial and airborne point clouds, but they still
89 resorted to the point-based 4PCS of Aiger et al. (2008) in the initialization. Dold and Brenner
90 (2006) used planar patches to establish correspondences and in the registration of geometric
91 information, and other attributes, such as areas, lengths, boundaries, and intensities, could be
92 derived and used to prune the outliers. Khoshelham (2016) proposed a closed-form solution for
93 point-to-plane correspondences to register two point clouds, that was independent of the initial
94 approximation and required no iteration. Xiao et al. (2013) and Xu et al. (2017) also used pla-
95 ninar patches to register pairwise scans in the urban scene, but their methods attempted to select
96 triples of planes from each scan for the candidate pools of correspondences and for determination
97 of the transformation parameters. Without additional constraints, the number of matching can-
98 didates was in the cubic order of the plane primitives. The large search space not only increased
99 the runtime but also reduced the successful registration rate. In consideration of these issues,
100 Xu et al. (2019) incorporated a 4PCS-like strategy to accelerate the process; the authors used
101 supervoxels (Zhu et al., 2017; Papon et al., 2013) instead of planar primitives to increase the
102 generalization ability, but this approach still degraded to the point-based strategy in registration
103 and therefore displays the limitations of point-based methods.

104 The proposed approach is also inspired by the work of Yang et al. (2015), which uses the
 105 building outlines to coregister aerial and terrestrial point clouds; despite the use of 2D line primitives
 106 for the horizontal alignment, Yang et al. (2015) requires the line primitives are complete
 107 for both the aerial and terrestrial datasets and several TLS scans are pre-registered to fulfill
 108 the above requirement. On the contrary, the proposed methods can handle partial overlapping
 109 primitives with robust primitive searching strategy.

110 3. Object-based incremental registration of point clouds

111 3.1. Problem formulation and overview of methods

112 Given two point clouds $\mathcal{S} = \{\mathbf{p}_i\}$ and $\mathcal{S}' = \{\mathbf{p}'_i\}$ that are independently captured from
 113 different locations, with partially overlapping regions Ω , the purpose of point cloud registration
 114 is to estimate an SE(3) transformation in a rigid manner, to align \mathcal{S} with \mathcal{S}' . For convenience, we
 115 also record the normal vectors \mathbf{n} and \mathbf{n}' for the source and target point clouds, respectively. The
 116 normal vectors are estimated using local patch defined by the K nearest neighbor (KNN) with
 117 $k = 12$. The SE(3) transformation consists of the rotation matrix $\mathbf{R} \in \mathbb{R}^{3 \times 3}$ and the translation
 118 vector $\mathbf{T} \in \mathbb{R}^3$. The rotation matrix \mathbf{R} has three DoFs and is parametrized in the angle-axis
 119 representation as $\mathbf{r} \in \mathbb{R}^3$ (Hartley and Zisserman, 2003). In the ideal case, we can generate
 120 sufficient correspondences $(\mathbf{p}, \mathbf{p}')$ in Ω and optimize for the parameters \mathbf{R} and \mathbf{T} as,

$$\min_{\mathbf{R}, \mathbf{T}} \sum_i \|\mathbf{p}'_i - (\mathbf{R}\mathbf{p}_i + \mathbf{T})\|^2. \quad (1)$$

121 Because no exact correspondence is available in the point clouds, the above optimization
 122 should be conducted using an iterative procedure. The termination criterion is generally defined
 123 on the distance between the correspondences,

$$\eta(\mathbf{p}'_i, \mathbf{p}_i^j) < \epsilon, \forall i = 1, 2, \dots, n, \quad (2)$$

124 where \mathbf{p}_i^j indicates the nearest neighbor for \mathbf{p}'_i after the j -th iteration. The function $\eta(\cdot)$ quantifies
 125 the differences with a certain upper bound. To improve the performance, most studies
 126 concentrate on two aspects: (1) the quality of correspondences and (2) the efficiency of the
 127 iterative strategy, which are also the aims of this paper.

128 In urban applications, the TLS is generally leveled, and most state-of-the-art scanners have
 129 the function of leveling compensation for slight tilting during installation. Therefore, the rotation
 130 degrades to only a single DoF as the rotation angle θ along the vertical axis $(0, 0, 1)^T$, i.e.
 131 $\mathbf{r}^* = (0, 0, \theta)$. The objective function in Equation (1) can then be reduced to the following form,
 132

$$\min_{\mathbf{R}^*, \mathbf{T}} \sum_i \|\mathbf{p}'_i - (\mathbf{R}^* \mathbf{p}_i + \mathbf{T})\|^2. \quad (3)$$

133 In addition, we can further simplify the problem by projecting the points to the horizontal
 134 plane. Therefore, we can enforce the z -dimension of the 3D points $(\mathbf{p}'_i, \mathbf{p}_i)$ and translation vector
 135 \mathbf{T} to 0 as $(\mathbf{p}'_i^*, \mathbf{p}_i^*)$ and \mathbf{T}^* , respectively. The 2D registration can be more concise as below,

$$\min_{\mathbf{R}^*, \mathbf{T}^*} \sum_i \|\mathbf{p}'_i^* - (\mathbf{R}^* \mathbf{p}_i^* + \mathbf{T}^*)\|^2. \quad (4)$$

136 Based on the problem formulations in Equations (1), (3) and (4), we design an incremental
 137 registration strategy that consists of three connected steps as shown in Figure 1. First, we align
 138 two scans in the 2D space (Subsection 3.2) by the line primitives. Second, we recover the offset
 139 in the vertical dimension by searching consistent correspondences in the local cylinder regions
 140 and solving a robust least-squares problem (Subsection 3.3). Third, we refine the registration
 141 with regard to all 6 DoFs by the uniformly selected local-oriented patches (Subsection 3.4).

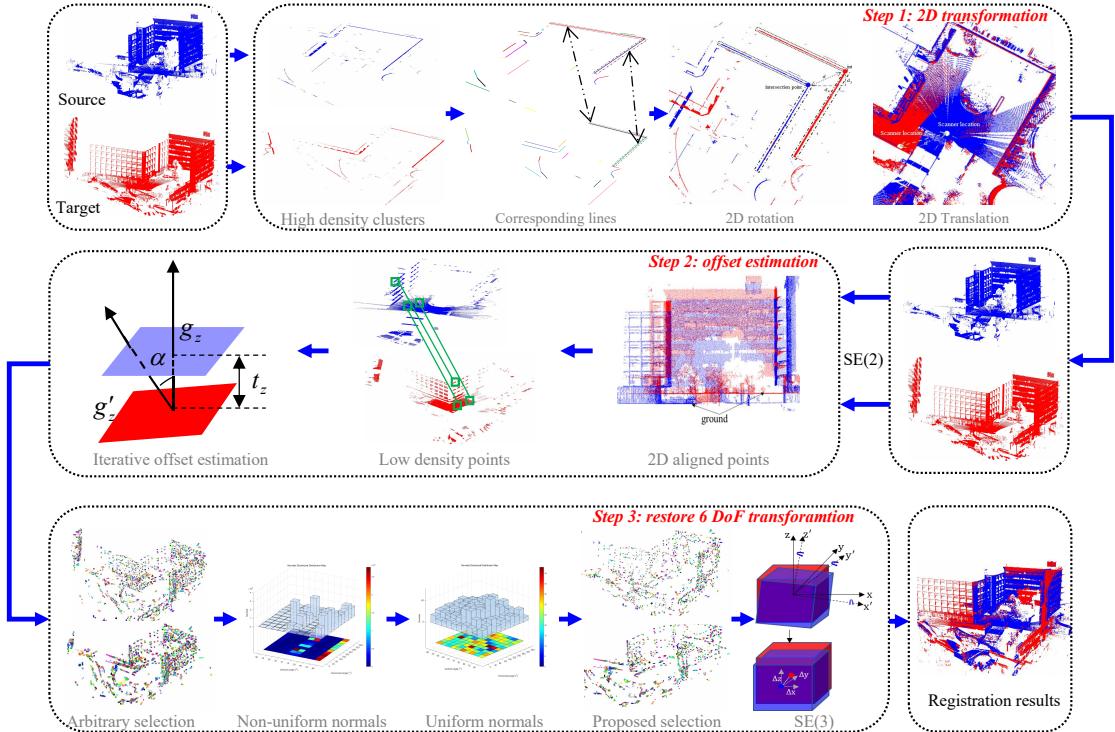


Figure 1: Overview of the proposed object-based incremental registration of TLS point clouds in an urban environment, which consists of three steps.

142 3.2. Line primitives for registration in a 2D space

143 3.2.1. Extraction of the line primitives

144 Similar to the work by Yang et al. (2015), we also used line primitives for the 2D alignment of
 145 the point clouds. For general purpose applications, the line primitives are commonly extracted
 146 from the intersections of adjacent planes (Al-Durgham et al., 2013; Poreba and Goulette, 2015;
 147 Date et al., 2018). But in the urban environment, after the 3D points are projected onto 2D
 148 horizontal planes, the facades of the buildings have remarkably greater point density than the
 149 regions of vegetation and other natural objects (Figure 2). Due to reliability issues, the points
 150 for building regions are preferred. The differences in the point densities contribute to the identifi-
 151 cation and extraction of the line primitives. The strategy of extracting line features in the 2D
 152 space is more resilient to the quality of the point clouds and arguably more efficient due to less
 153 parameters and data involved.

154 To remove outliers and simplify the scenes, we first extract gridded points from the point
 155 clouds. Specifically, the average spacing ρ of the point clouds is first estimated, and the points
 156 are gridded in 2D space with the spacing of 0.2ρ . If more than one point falls into the same grid,
 157 the one near the center is preserved, and the number n of points in the same grid is recorded.
 158 As described above, because the purpose is to exploit building facades and other static artificial
 159 objects for registration, the line primitives only use the gridded points with $n \geq 10$. The average
 160 height of each grid is also recorded, which can be used to distinguish adjacent line segments from
 161 the same facades clustered from different floors. In summary, each gridded point is represented
 162 as $\mathbf{g} = (g_p, g_z, g_n)$ for the coordinate, the average height, and the point density, respectively. By
 163 setting this threshold, most undesired structures in the scenes, such as vegetation and dynamic



Figure 2: Differences in point densities for point clouds in an urban environment on the horizontal plane. (a) Perspective view and (b) top view with two enlarged areas.

164 objects, can be successfully separated. In addition, general purpose outlier filters are also useful
 165 and could be used in parallel with the proposed screening approach, such as the statistical outlier
 166 removal (Rusu and Cousins, 2011). Figure 3 demonstrates the preserved points after projection
 167 and outlier removal.

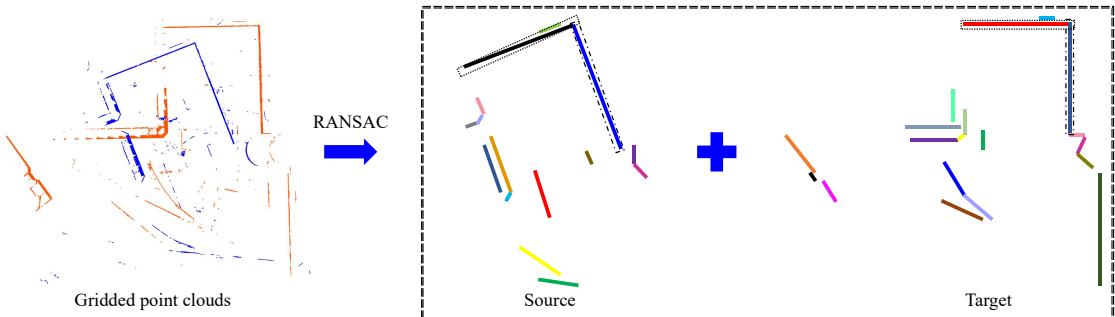


Figure 3: Extracted line primitives from the projected point clouds.

168 We use the efficient RANSAC approach (Schnabel et al., 2007) to detect all line shapes from
 169 the 2D point clouds and to refine the lines by least-squares fitting. The approach by Schnabel
 170 et al. (2007) accounts for both fitting error and spatial locality; therefore, discontinued lines
 171 are clustered into different line segments. To ensure continuity and handle the problems caused
 172 by noise and occlusion, only large segments are preserved by the threshold of 1 m. Although
 173 the targeted scenes are an urban environment that features structured objects, the number of
 174 extracted line segments should be low, which is the desired property for representativeness. Only
 175 a few line segments are used to represent the whole structure (Figure 3b), compared to more
 176 than 10 million points and thousands of keypoints (Ge, 2017; Ge and Wu, 2019). The heights of
 177 the line segments are taken from the average of the occupied grids.

178 3.2.2. 2D registration using the line primitives

179 As demonstrated in previous works (Yang et al., 2016; Sumi et al., 2018), reducing the DoFs
 180 will accelerates process of the searching of primitives, even for point-based features. The following
 181 explores further the 2D registration using line features, which are more robust and efficient.

182 Although the number of line segments is small and a single pair of line match suffices for
 183 estimation of the rotation angle θ in Equation (4), it is still a non-trivial work to robustly recover

184 the 2D transformation. Unlike the robust least-squares optimization (Zhou et al., 2016), we
 185 approach the RANSAC paradigm (Bolles and Fischler, 1981) due to the large number of outliers.
 186 Instead of exhausting all possible configurations of line segments, inspired by the 4PCS (Aiger
 187 et al., 2008), we first establish a pool of candidates before the RANSAC searching. Specifically,
 188 we denote the extracted line segments as $\mathcal{L} = \{l_i\}$ and $\mathcal{L}' = \{l'_i\}$ from point clouds S and
 189 S' , respectively. For the candidates, we randomly choose a pair of line segments instead of
 190 only a single segment from \mathcal{L} , i.e. $c_{ij} = (l_i, l_j)$. The number of iterations for RANSAC, is
 191 therefore bounded by the number of combinations of possible pairs; we also limit the procedure
 192 by the running time (60 s is used) (Ge, 2017). In the target set \mathcal{L}' , if the corresponding pair
 193 $c'_{pq} = \{l'_p, l'_q\}$ also selects randomly, the time complexities quickly explode; therefore, we apply
 194 several geometric rules in the choices for c'_{pq} .

195 i. *Length rule.* The lengths of the corresponding line segments should be bounded by the
 196 threshold ϵ_1 , which is fixed at $\epsilon_1 = \frac{1}{3}\|l_{max}\|$ with l_{max} for the longer segment,

$$\|l_i\| - \|l'_p\| < \epsilon_1 \text{ and } \|l_j\| - \|l'_q\| < \epsilon_1. \quad (5)$$

197 ii. *Angle rule.* The intersection angles for pairs c_{ij} and c'_{pq} should be similar, and the difference
 198 is limited by the threshold δ ($\delta = 2^\circ$ in this paper),

$$|\langle l_i, l_j \rangle - \langle l'_p, l'_q \rangle| < \delta. \quad (6)$$

199 iii. *Height rule.* The height differences of the correspondences of the line segments should be less
 200 than a given threshold ϵ_2 ($\epsilon_2 = 5 m$ in this paper) to prevent adjacent lines from belonging
 201 to the same facades on different floors,

$$|h_{l_i} - h_{l'_p}| < \epsilon_2 \text{ and } |h_{l_j} - h_{l'_q}| < \epsilon_2. \quad (7)$$

202 iv. *Intersection rule.* After transformation using the rotation angle θ , the corresponding line
 203 segments in the sets \mathcal{L} and \mathcal{L}' should be parallel. The distance between the intersected
 204 points after rotation, that is, the points v_{ij} and v'_{pq} intersected by the line candidates c_{ij}
 205 and c'_{pq} , is approximately equal to the distance between the two stations (Figure 4), which
 206 should reside within a reasonable range ϵ_3 ($\epsilon_3 = 20 m$ in this paper),

$$\|v_{ij} - v'_{pq}\| < \epsilon_3. \quad (8)$$

207 We allow one line to be searched with multiple candidates in the RANSAC procedure for
 208 better recall ratio; however, in practice, the constraints above and the robust scoring scheme will
 209 handle the one-to-many correspondences appropriately. In addition to restricting the selection
 210 of the matching pair c'_{pq} , the four rules also guide the procedure for the efficient identification of
 211 possible candidates, as illustrated in Algorithms 1 and 2. If no candidate is available for pair c_{ij} ,
 212 in the cases in which either l_i or l_j is not in the overlapping region Ω , the algorithm continues for
 213 another random pair. Otherwise, we must score the candidates and select the best one. Using
 214 the correspondence of c_{ij} and c'_{pq} , we have a closed-form solution for the 2D transformation
 215 defined in Equation (4). All line segments in \mathcal{L} are then warped to the coordinate system of \mathcal{L}' .

216 Possible matched-line correspondences are determined by the nearest distances of the lines
 217 (Schneider and Eberly, 2018). In addition, the correspondences with length and angle differences
 218 greater than ϵ_1 and δ , respectively, make no contribution to the score of the RANSAC validation.
 219 The score of this configuration is computed from the truncated and normalized residuals of angle
 220 δ_k and distance μ_k for all matched-line correspondences (Poreba and Goulette, 2015; Ge and
 221 Wu, 2019),

$$s = \sum_k \alpha w(\delta_k) + (1 - \alpha) w(\mu_k), \quad (9)$$

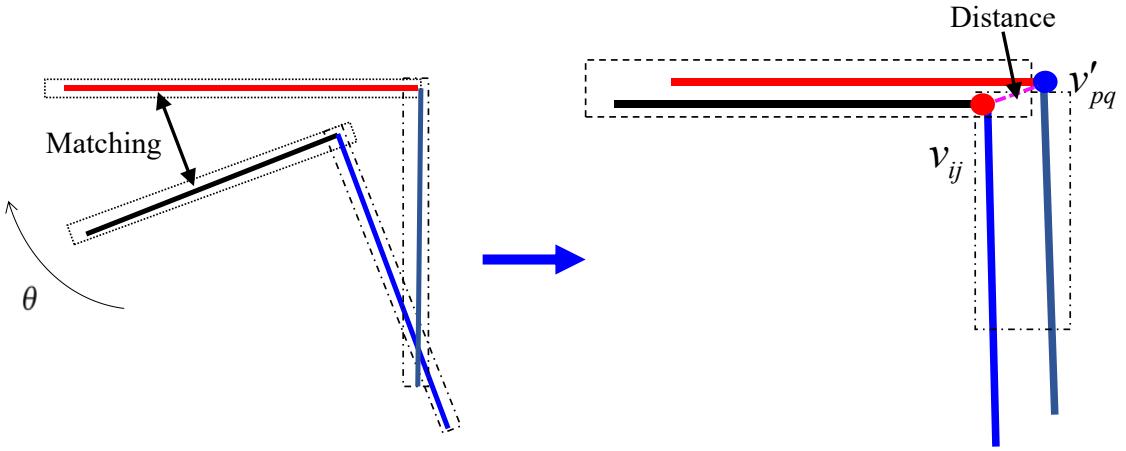


Figure 4: Distance rule for the selection and pruning of matching candidates.

222 where α is the weight factor ($\alpha = 0.3$ is used empirically) and $w(\cdot)$ is a single peak robust
 223 estimator (Figure 5) as detailed in (Ge and Wu, 2019). In summary, the estimation of the 2D
 224 transformation is illustrated in Algorithm 2.

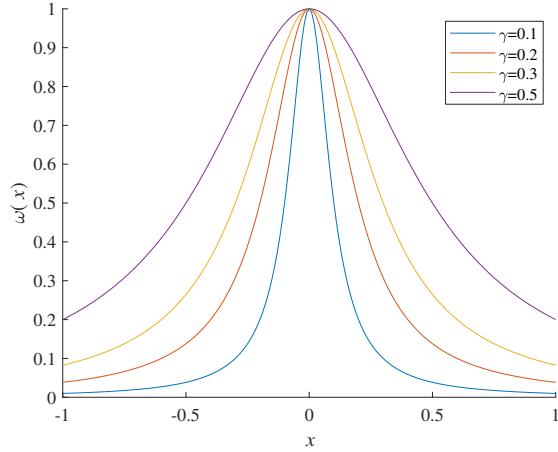


Figure 5: Single-peak estimator for the evaluation of the RANSAC score, with respect to different γ .

225 *3.3. Robust estimation of the height offset*

226 After 2D transformation, it is assumed that structures such as facades align well in the
 227 horizontal direction; however, discrepancies remain in the z -dimension due to the different heights
 228 at which the stations are located, as shown in Figure 6. During the previous step, we clustered
 229 the point clouds into the regions of high and low point density. The clusters with low point
 230 density generally present a horizontal layout, such as the ground. Therefore, we use the cluster
 231 with low point density to estimate the height offset.

232 For both points with a low density \mathcal{S}_L and \mathcal{S}'_L extracted from \mathcal{S} and \mathcal{S}' , respectively, the height
 233 offset estimation is conducted in a robust least-squares manner with iterative outlier removal.
 234 The correspondences between \mathcal{S}_L and \mathcal{S}'_L are defined in the regular grids and searched in the

Algorithm 1 Efficient searching for candidate pairs.

```

procedure SEARCHCANDIDATE( $l_i, l_j, \mathcal{L}'$ )
2:    $\mathcal{C}' \leftarrow \emptyset$ 
3:   for  $l'_p$  in  $\mathcal{L}'$  do
4:     if  $l'_p$  disobeys neither Equation (5) nor (7) then
5:       continue;
6:     end if
7:      $\theta \leftarrow \langle l_i, l'_p \rangle$ 
8:     for  $l'_q$  in  $\mathcal{L}'$  do
9:       if  $(l'_p, l'_q)$  disobey Equations (5) through (8) then
10:        continue;
11:      end if
12:       $\mathcal{C}' \leftarrow (l'_p, l'_q)$ 
13:    end for
14:  end for
15:  return  $\mathcal{C}'$ 
16: end procedure

```

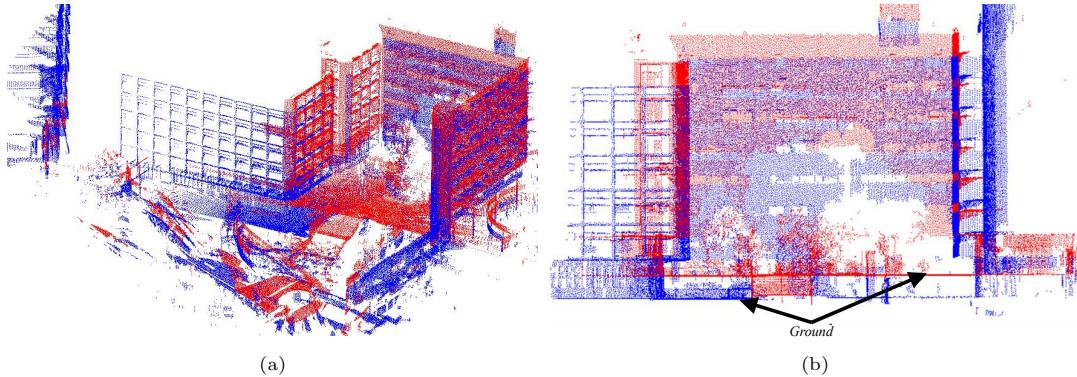


Figure 6: Point clouds after the 2D transformation of the two point clouds. (a) Perspective view and (b) front view. The red and blue colors indicate different scans.

235 cylindrical regions with a fixed radius of 5ρ , denoted as (g_i, g'_i) . In addition, we restrict the
236 corresponding grids to have similar normal vectors, i.e. $\mathbf{n} \cdot \mathbf{n}' > 0.99$ (approximately 8°). The
237 robust least-squares estimation is formulated briefly as,

$$\min_{t_z} l(g'_z - (t_z + g_z))$$

$$l(x) = \begin{cases} x^2/2, & |x| \leq a, \\ a|x| - a^2/2, & |x| > a \end{cases} \quad (10)$$

238 where we use the robust Huber loss function $l(\cdot)$ (Hartley and Zisserman, 2003; Hu et al., 2016) in
239 the estimation, and the hyper-parameter $a = 2\sigma$ is used. In addition, we also iteratively remove
240 the inevitable outliers using the 68–95–99.7 law at 3σ for the residuals. The initial value of the
241 height offset \bar{t}_z is the average of the height differences of all corresponding grids (g, g') . The
242 robust loss function and least-squares solver are both implemented in the Ceres Solver (Agarwal
243 et al., 2010).

Algorithm 2 2D registration using line primitives.

```

procedure LINE2DREGISTRATION(  $\mathcal{L}$ ,  $\mathcal{L}'$  )
2:    $N \leftarrow \#$ Iterations
     $s \leftarrow 0$ 
4:    $\mathcal{C} \leftarrow \emptyset$                                  $\triangleright$  The set of existing candidates
    for  $i = 0, \dots, N$  do
6:      $c_{ij} \leftarrow (l_i, l_j)$                        $\triangleright$  A random pair in  $\mathcal{L}$ 
      if  $c_{ij} \cap \mathcal{C} \neq \emptyset$  then
8:       continue;
      end if
10:     $\mathcal{C} \leftarrow c_{ij}$ 
11:     $\mathcal{C}' \leftarrow \text{SearchCandidate}(l_i, l_j, \mathcal{L}')$ 
12:    for  $c'_{pq}$  in  $\mathcal{C}'$  do
13:       $(\mathbf{R}^*, \mathbf{T}^*) \leftarrow \text{Aligning } c_{ij} \text{ and } c'_{pq}$ .
14:       $s_2 \leftarrow \text{Compute the score according to Equation (9)}$ 
         $s \leftarrow \max(s, s_2)$ 
16:    end for
    end for
18:    return  $(\mathbf{R}^*, \mathbf{T}^*)$  with the maximum score  $s$ 
end procedure

```

244 *3.4. Local oriented patch for refinement in 3D space*

245 The solution with regard to 4 DoFs, as in Equation (3), still has two limitations: 1) the 2D
 246 transformation is estimated in a RANSAC paradigm, which only uses two line correspondences
 247 and is not the global optimal, and 2) perfect leveling cannot be satisfied. Therefore, we restore
 248 the solution from 4 DoFs to the full 6 DoFs. Although a standard ICP or its variants may
 249 suffice for the task, an independent implementation is not efficient to solve the problem, and
 250 point-based approaches are prone to be trapped to a local minimum, which may be biased in
 251 a structured urban environment. To fulfill this requirement, we consider a point as an oriented
 252 patch (\mathbf{s}, \mathbf{n}) to make use of both the positional and directional information. The local patch is
 253 a small sufficient plane that can be obtained by a point with its K (=12 in our cases) nearest
 254 neighbor. Because the two cues have different orders of magnitude, the optimization is separated
 255 into two parts and iterated similar to ICP,

$$\begin{aligned} & \min_{\mathbf{R}} \sum_i |\mathbf{n}'_i \cdot \mathbf{R} \mathbf{n}_i|^2 \\ & \min_{\mathbf{T}} \sum_i \|\mathbf{p}'_i - (\mathbf{R} \mathbf{p}_i + \mathbf{T})\|^2. \end{aligned} \tag{11}$$

256 Unlike the standard ICP, the proposed methods do not rely upon the point-based Eulerian
 257 distances. The local patches are used instead, as shown in Figure 7. Two strategies are exploited
 258 to select the correspondences: (1) the directional distance constraint and (2) uniform distribution
 259 ([Rusinkiewicz and Levoy, 2001](#)). For the first, we search for the nearest neighbor in the extruded
 260 cylinder defined by the local patch and restrict the valid correspondences to consistent normal
 261 directions (approximately 8°). For the latter, we prefer uniformly distributed normal vectors,
 262 especially for balanced rotation estimation in Equation (11). As shown in Figure 8, projection
 263 of the normal vectors of the points onto the Gaussian sphere ([Voselman et al., 2017](#)) shows that
 264 the distribution is quite nonuniform. This can also be noticed in the histograms by discretizing

265 the horizontal and vertical angles into certain bins; the histograms have several peaks. To obtain
 266 a uniform distribution of normal vectors, the maximum number of selected local patches for each
 267 bin is truncated at $n_b = n_s/100$, with n_s as the expected total number of patches ($n_s = 5000$ is
 268 used).

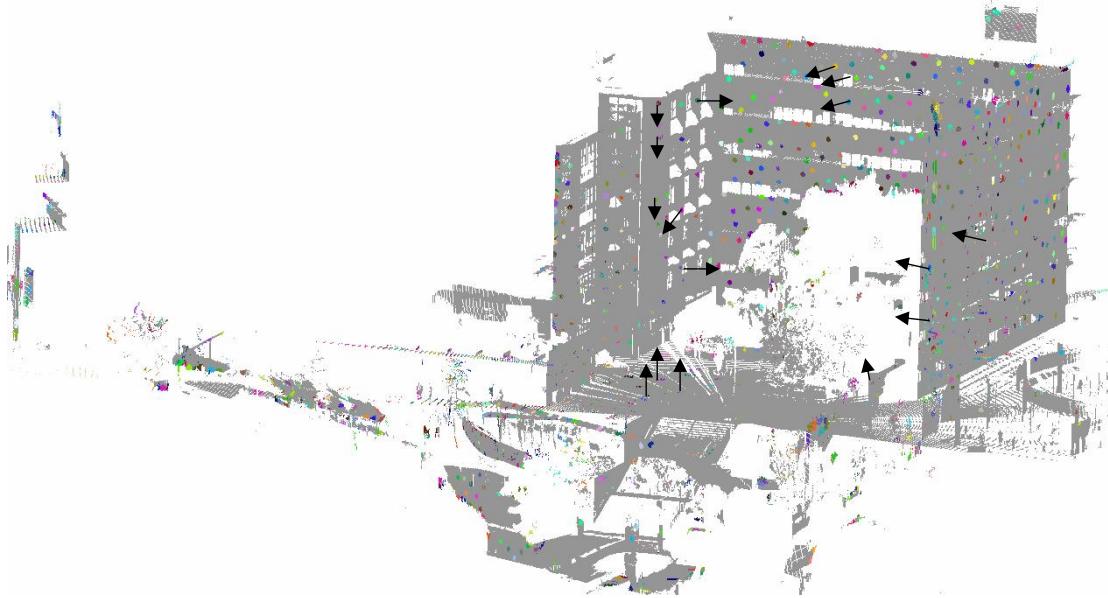


Figure 7: Local patch is an oriented planar disk that represents KNN neighbors centered around a point.

269 4. Experimental evaluations

270 4.1. Datasets and evaluation metric descriptions

271 The performance of the proposed object-based incremental registration (OBIR) is evaluated
 272 with five challenging datasets. Datasets 1 and 2 were captured by the authors using Leica BLK
 273 360 and Faro 3D scanners, respectively. Dataset 1 includes two scans of the university campus,
 274 and dataset 2 includes four scans of the streets of urban Hong Kong. Datasets 3 through 5 are
 275 the benchmark point clouds delivered by the Semantic3D (Hackel et al., 2017), including point
 276 clouds of the market square in Feldkirch, the cathedral in St. Gallen, and the church in Bildstein.
 277 Table 1 shows more details of the datasets.

Table 1: Summary of the experimental datasets

| Dataset | Description | Abbr. | #Scans | #Points(10^6) | Dimensions (m) |
|---------|----------------------------|-------|--------|-------------------|-----------------------------|
| 1 | University Campus | CP | 2 | 27.1 | $107 \times 86 \times 47$ |
| 2 | Urban Streets of Hong Kong | US | 4 | 29.2 | $350 \times 477 \times 107$ |
| 3 | Market Square in Feldkirch | MS | 3 | 22.6 | $253 \times 407 \times 56$ |
| 4 | Cathedral in St. Gallen | CG | 2 | 28.2 | $287 \times 278 \times 73$ |
| 5 | Church in Bildstein | CB | 2 | 29.7 | $230 \times 246 \times 65$ |

278 To compare the proposed OBIR with the state-of-the-art approaches in both registration
 279 accuracy and computational runtime, we evaluate six other methods or configurations in our

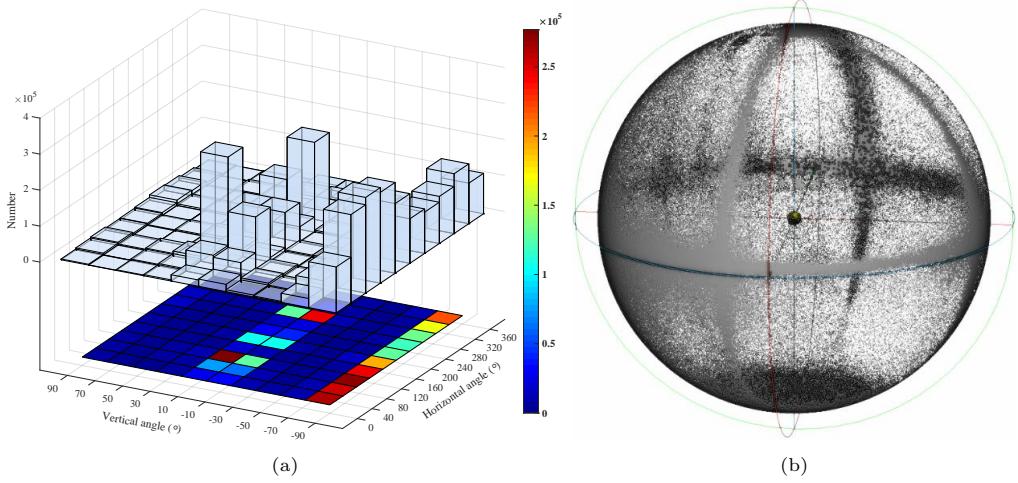


Figure 8: Nonuniform selection of local patches. (a) and (b) are histograms of the orientations of the normal vectors and the projection of the normal vectors onto the Gaussian sphere for unnormalized selection of the local oriented patches, respectively.

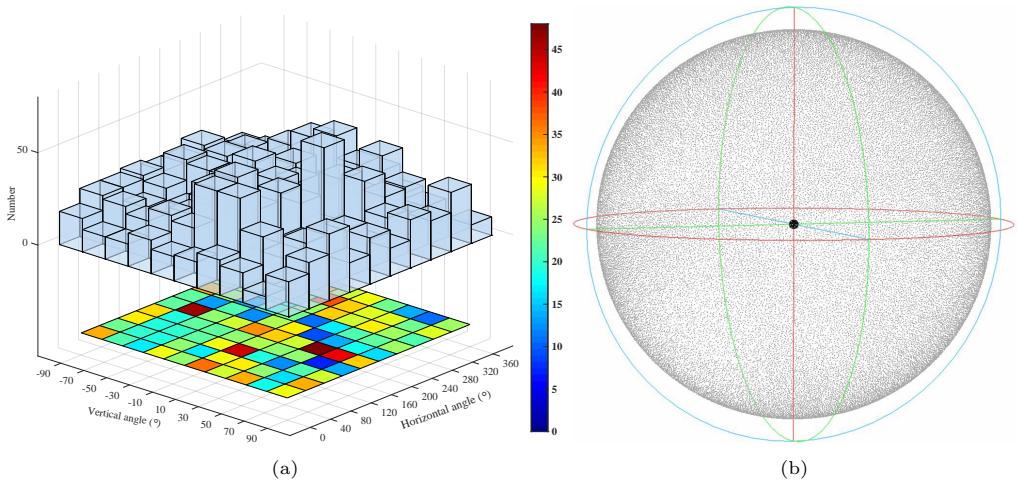


Figure 9: Uniform selection of local patches. (a) and (b) are histograms of the orientations of the normal vectors and the projection of the normal vectors onto the Gaussian sphere for the normalized selection of the local oriented patches, respectively.

280 experiments, including (1) the branch-and-bound (BnB) algorithm, (2) fast match pruning BnB
 281 (FMP-BnB) (Cai et al., 2019), (3) fast global registration (FGR) (Zhou et al., 2016), (4) the
 282 game-theory approach (GTA) (Albarelli et al., 2015), and (5) the Keypoint 4PCS (K4PCS)
 283 (Theiler et al., 2014) and (6) K4PCS augmented by the Generalized ICP (G-ICP) (Segal et al.,
 284 2009). We use the same keypoint detector and descriptor when necessary, such as for the ISS
 285 keypoints (Zhong, 2009) and the FPFH descriptors (Rusu et al., 2009).

286 As in many previous studies (Dong et al., 2018; Cai et al., 2019; Xu et al., 2019; Ge et al.,
 287 2019a), we introduce three metrics to evaluate the OBIR in terms of the rotation error $e_{\mathbf{R}}$, the
 288 translation error $e_{\mathbf{T}}$, and the runtimes,

$$e_{\mathbf{R}} = \arccos \frac{\text{tr}(\bar{\mathbf{R}} - \mathbf{R}) - 1}{2}, \quad (12)$$

$$e_{\mathbf{T}} = \|\bar{\mathbf{T}} - \mathbf{T}\|$$

289 where $\bar{\mathbf{R}}$ and $\bar{\mathbf{T}}$ represent the ground truth in the evaluation, which are generated either via
 290 artificial spherical targets (for CP and US) using an industrial proved solution¹ or via interactive
 291 registration (in public benchmark datasets) followed by the optimized ICP². Both methods,
 292 technically, can reach accuracies at millimeter-level and therefore are sufficient for the evaluation
 293 of the markerless registrations. For the runtimes, we directly record the CPU runtime for each
 294 method monitored on the same hardware. For the grid sizes, 0.2ρ is used.

295 4.2. Qualitative evaluations

296 For qualitative evaluation of the OBIR, we demonstrate the overall and enlarged registration
 297 results in Figures 10 and 11, respectively; for the enlarged demonstrations, if more than two
 298 scans are involved for a dataset, only the first two scans are shown for brevity. In addition,
 299 for datasets with multiple scans, such as the US and MS, the strategy of sequential pairwise
 300 registration is used. First, all pairwise registrations converge satisfactorily without any human
 301 intervention. Although only a sequential strategy is used, the scans that are not adjacent show
 302 no obvious inconsistencies, as seen in Figure 10. Turning to the enlarged views, the facades are
 303 aligned quite well, as shown in the top view of Figure 11a and in the left part of Figure 11c. More
 304 interesting results are illustrated in the small structures in the enlarged view, such as the trees
 305 in the US dataset, the adjacent facades in the crossroad (MS), the domes and dormers (CG),
 306 and the crosses atop the churches (CB).

307 4.3. Quantitative evaluations

308 To evaluate the performances of pairwise registration among methods, we separate the
 309 datasets with multiple scans into several sequential pairwise registrations, such as US:1-2 and
 310 MS:2-3. For the campus dataset (CP), we intentionally create another pair with a large initial
 311 orientation difference (denoted as CP*) in the evaluations. With regard to the methods, we add
 312 a variant of K4PCS (Theiler et al., 2014) by augmenting it with the G-ICP (Segal et al., 2009)
 313 to allow a fair comparison, because K4PCS is deemed a coarse registration method whereas the
 314 others are all considered to be end-to-end solutions. Therefore, we have nine pairs of datasets
 315 and seven methods or strategies in the comparisons.

316 With regard to the overall performance, Figures 12 through 14 display comparisons of the
 317 OBIR and the other methods in terms of rotation errors, translation errors, and runtimes. The

¹RecapPro: <https://www.autodesk.com/products/recap>

²Implemented in CloudCompare: <https://www.danielgm.net/cc/>

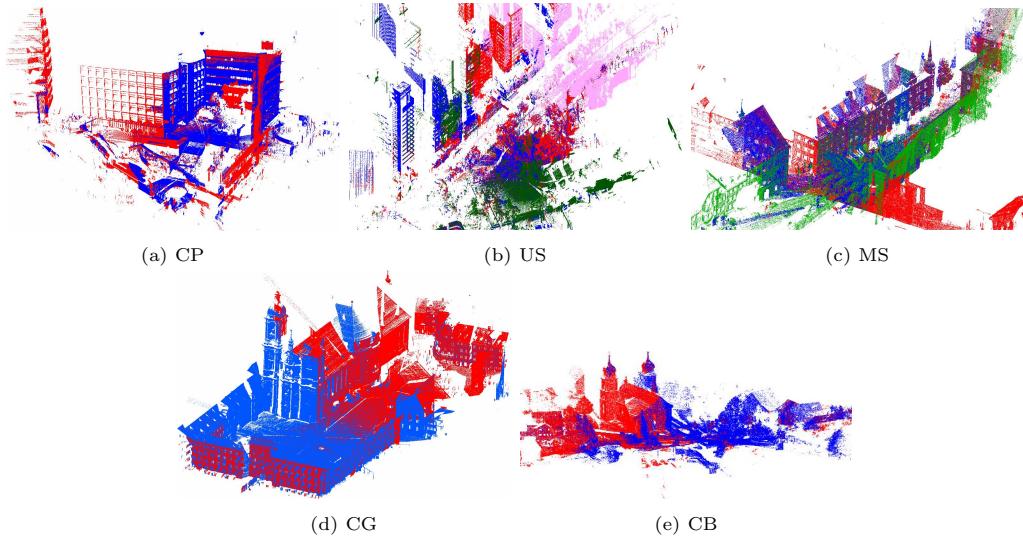


Figure 10: Scans registered by the proposed method. Different scans in the same dataset are distinguished by color.

best and second-best solutions are highlighted with red and blue cells, respectively, in the comparisons. OBIR generally satisfies the requirements of fine registration in terms of registration accuracy. In most cases, the rotation error and translation error of OBIR fall below 0.1° and 0.1 m , respectively, and in nearly all cases (i.e., with only one exception), OBIR achieves the best or second-best performances with regard to both errors. Furthermore, OBIR delivers the best overall performance in terms of the average and median results.

Turning to comparisons with a specific method, it could be noted that the performance of OBIR is *on par* with or even surpasses the K4PCS+G-ICP method. The G-ICP is an ad-hoc fine registration method, which is a more powerful and drop-in replacement for the standard ICP; in contrast, the OBIR is an end-to-end solution. In addition, G-ICP is severely dependent upon the initial values. For example, in the CB dataset, K4PCS cannot provide sufficient initial values, which leads directly to the poor performance of G-ICP (Figures 12 and 13).

FGR is also a viable end-to-end solution. In theory, FGR does not require good initializations compared with the fine registration methods such as G-ICP and ICP, but in practice, it fails quite often when it is not well initialized. Although the point clouds are leveled, the large difference in tertiary orientation may still lead to the nonconvergence of FGR, which was also noted by Cai et al. (2019), likely because FGR does not involve an iterative strategy and because a single optimization in a least-squares manner still depends upon good initial values.

The two variants for the branch-and-bound strategy, the BnB and FMP-BnB, have quite similar and stable performances. However, although the BnB approaches, especially the FMP-BnB, were claimed to be end-to-end solutions with global optimal, the achieved accuracies are inferior to the other approaches when converged. The rotation errors all exceed 1° , and in five of the nine cases the translation errors exceeds 1 m because the approach considers only 4 DoFs even in the refinement, and practical cases may not meet this requirement strictly.

Another important evaluation metric for the registration of point clouds is the runtime. As shown in Figure 14, OBIR is the fastest in five of the nine datasets and has the best overall performance. In addition, the performance is quite stable, never exceeding 30 s in any experiment. FGR also shows a quite impressive runtime performance, which is expected due to the extensive

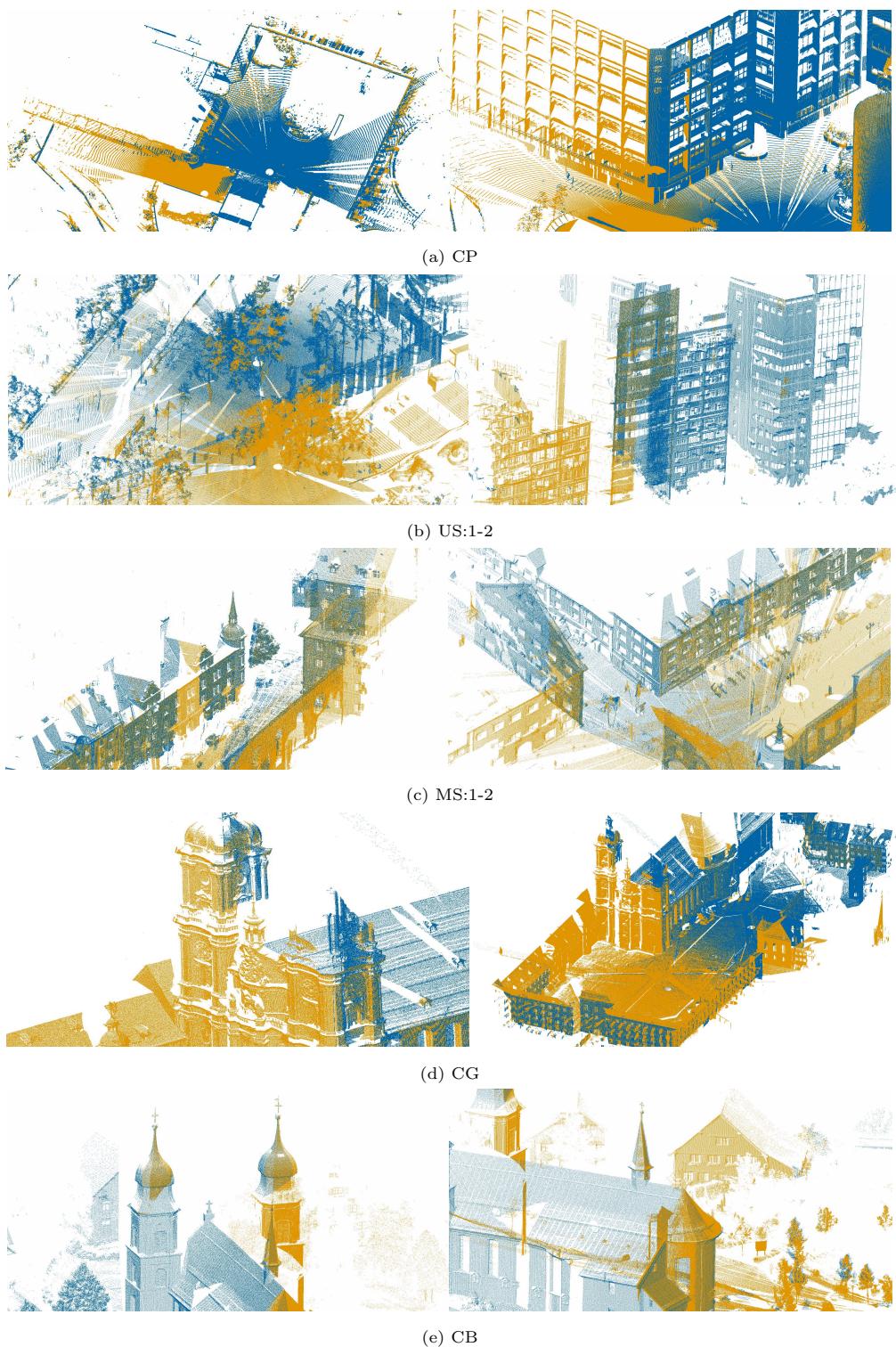


Figure 11: Enlarged view of the pairwise results for the five datasets. Blue and yellow points indicate source and target scans, respectively.

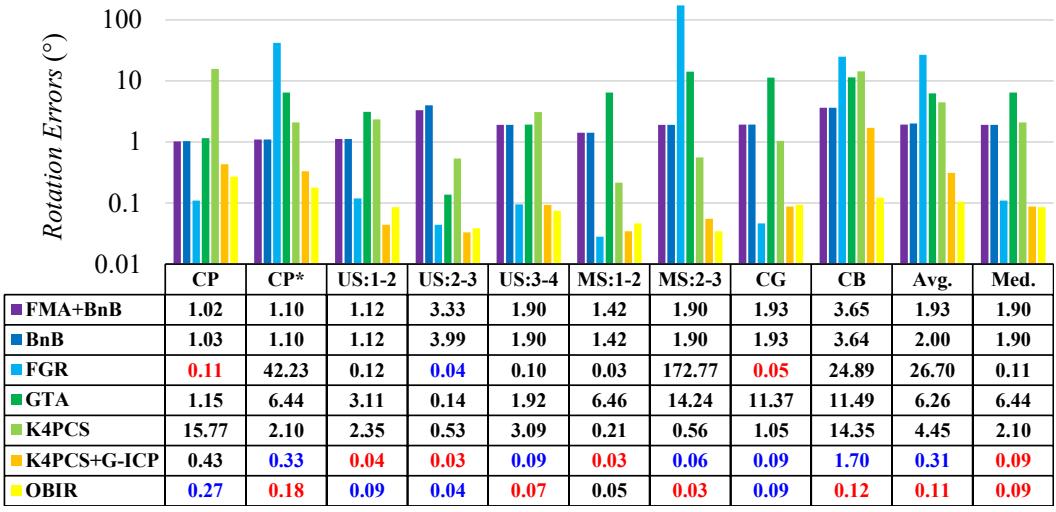


Figure 12: Comparison of the rotation errors in the nine pairwise registrations. Red and blue cells denote the best and second best results, respectively.

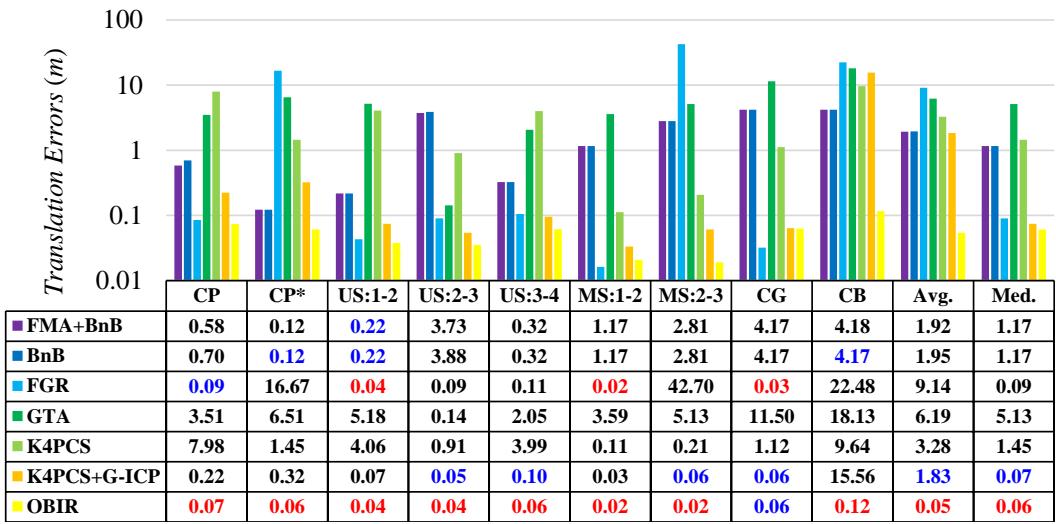


Figure 13: Comparison of the translation errors in the nine pairwise registrations. Red and blue cells denote the best and second best results, respectively.

346 work done on pruning false matches and the sophisticated optimization without iterative pro-
 347 cedures. However, the efficiency of FGR comes at the cost of decreased robustness, that is, it
 348 does not even converge in a third of the experiments, and when not converged, the runtimes of
 349 FGR are relatively longer. FMP-BnB also shows very competitive runtimes and is the second-
 350 best with regard to the average results due to the simplification of the reduced 4-DoF strategy.
 351 This strategy is also shared by the proposed method, and we exploit the above merit further
 352 by abstracting the urban scenes with much fewer, but more representative, objects. The reason
 353 that we use fewer primitives is also addressed by comparison with the K4PCS approach, which
 354 uses thousands of keypoints. Even if the 4PCS strategy significantly reduces the search space,
 355 the candidate pool is still inevitably larger due to the number of keypoints involved. Therefore,
 356 even without refinement by the G-ICP, K4PCS already requires the greatest runtime.

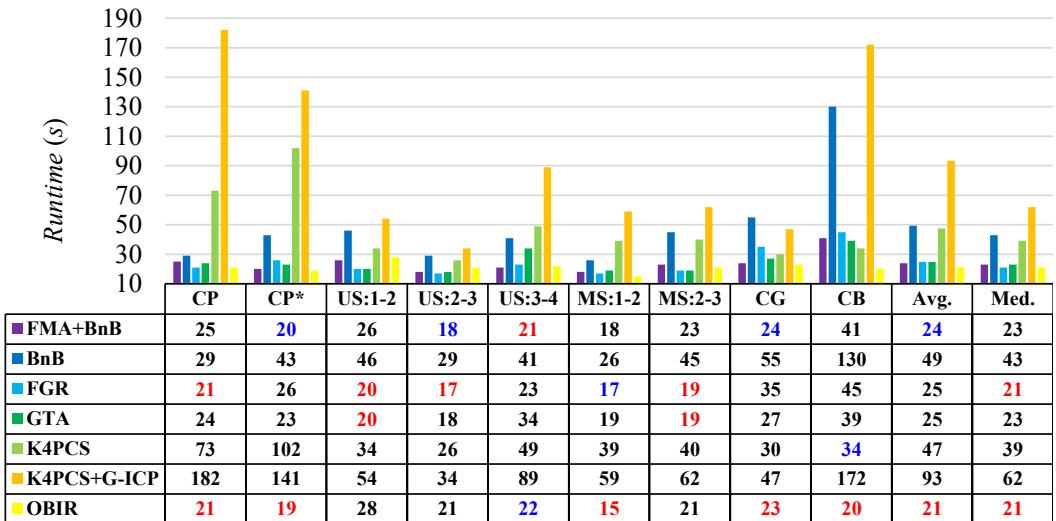


Figure 14: Comparison of the runtimes for the nine pairwise registrations. Red and blue cells denote the best and second best results, respectively.

357 4.4. Evaluations of different primitives

358 In the evaluations above, the other methods all use the same type of keypoints, and OBIR
 359 uses the line primitives detected from the efficient RANSAC approach (Schnabel et al., 2007). To
 360 compare the robustness and generalizability with regard to different primitives, we also conduct
 361 experiments using other types of keypoint detectors. For the point-based approaches, besides
 362 the baseline method of ISS, the SIFT3D, Harris3D, and NARF methods (Hana et al., 2018) are
 363 also evaluated. For the line primitives, we also test OBIR combined with the Hough transform
 364 (Yao et al., 2010; Rusu and Cousins, 2011), the line segment detector (LSD) (Von Gioi et al.,
 365 2008), and the region growing strategy (RG) (Rusu and Cousins, 2011). The CP and MS:1-2
 366 datasets are used for this experiment, and the sensitivity of various keypoint detectors are tuned
 367 to make the number of keypoints at the same order of magnitude.

368 Table 2 shows the number of extracted keypoints and the line primitives for the different
 369 methods. The number of line primitives is considerably smaller than the number of keypoint de-
 370 tectors. The reduced number of primitives, however, does not decrease the OBIR’s performance.

371 Another objective of the evaluation is to test the reliability with regard to various methods of
 372 primitive extraction. Figures 15 through 17 demonstrate the rotation errors, translation errors,

373 and runtimes for each method and primitive combinations. It can be immediately noted that for
 374 the rotation and translation errors, the OBIR remains stable among the exploited line primitives.
 375 In contrast, all of the point-based approaches have obvious oscillations for the four popular
 376 keypoint detectors, likely due to the inherent repeatability issue for the keypoint detectors. With
 377 regard to the runtime stability, it may seem at first glance that the performance is not stable,
 378 especially for the combination of OBIR and RG; however, the differences in the runtimes cannot
 379 be attributed to the registration pipeline because the runtimes also account for the extraction of
 380 the line primitives and, in fact, the time for registration is quite stable.

Table 2: Summary of the number of keypoints extracted by different detectors and the number of line segments extracted by OBIR.

| Dataset / Scan | #Points | | | | #Lines | | | |
|----------------|---------|--------|----------|------|--------|-------|-----|----|
| | ISS | SIFT3D | Harris3D | NARF | RANSAC | Hough | LSD | RG |
| CP / S1 | 3742 | 3659 | 2072 | 2698 | 21 | 28 | 31 | 31 |
| CP / S2 | 2780 | 2964 | 2768 | 2772 | 24 | 33 | 40 | 32 |
| MS / S1 | 2319 | 4320 | 2538 | 2898 | 22 | 31 | 31 | 25 |
| MS / S2 | 2952 | 2939 | 1726 | 2765 | 19 | 25 | 34 | 19 |

4.5. Discussion and limitations

Based on these evaluations, we now discuss some characteristics of the proposed OBIR and
 note possible limitations that must be surmounted.

384 1) *Robustness through an incremental strategy.* Recent studies (Cai et al., 2019) have con-
 385 firmed that reducing the number of DoFs from six to four significantly increases the robustness.
 386 We go further in this direction by incrementally increasing the DoFs to three, four, and six. This
 387 strategy not only simplifies the problem, but also, surprisingly, reduces the runtime, because
 388 each subproblem is easier to handle. The incremental strategy delivers more stable performances
 389 at different initial setups, i.e. broader convergence region (Bae, 2009), than other strategies.

390 2) *Representativeness of the line primitives.* Except for the 2D transformation using line
 391 primitives, the following steps are all performed in a (iterative) least-squares manner. Although
 392 previous studies have also investigated line or plane primitives (Theiler et al., 2012), the under-
 393 lying registration is still point-based. In this paper, we demonstrate that the number of lines is
 394 only a small fraction of the keypoints; as long as the lines are repeatable among various scans, the
 395 reduced amount significantly narrows the search space for feature correspondences. In addition,
 396 due to the representativeness of line primitives, the OBIR is agnostic to the methods used for
 397 line extraction.

398 3) *Limitations.* Although urban environments rarely lack suitable structures for line primi-
 399 tives, the proposed methods are not applicable to natural scenes, in which no vertical structures
 400 exist; in addition, two unparalleled line primitives are required. The proposed method, similar
 401 to (Cai et al., 2019), also premise the approximate leveling of the scanners, which imposes slight
 402 constraints on the data collections.

5. Conclusions

404 This paper proposes an object-based incremental registration (OBIR) of point clouds collected
 405 in an urban environment. The two pillars of the proposed method are the incrementally increased
 406 DoFs and the representativeness of the line primitives. Both benefit in the reduction of the

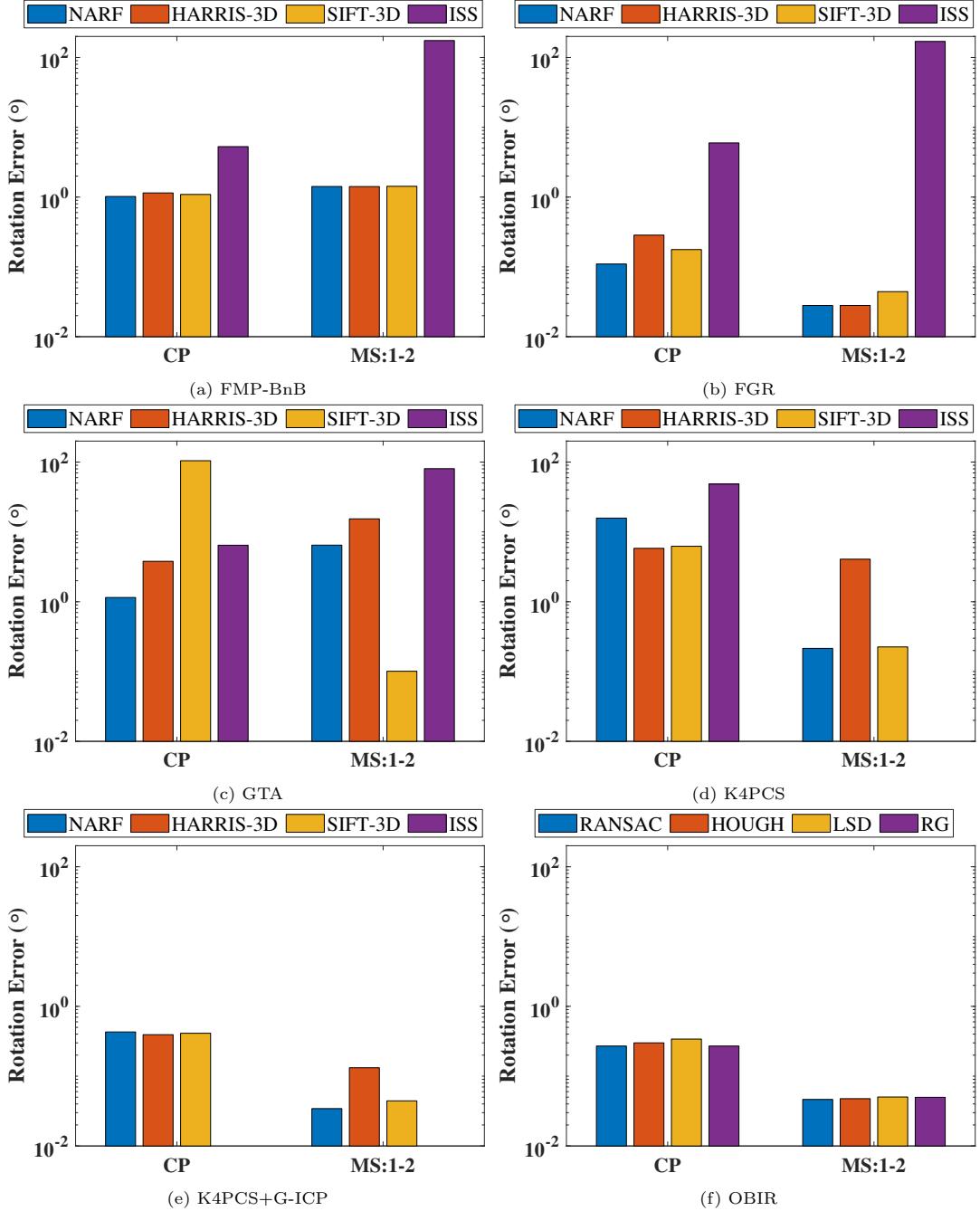


Figure 15: Rotation errors for different keypoints or line configurations. Missing columns result from failures of convergence for specific configurations.

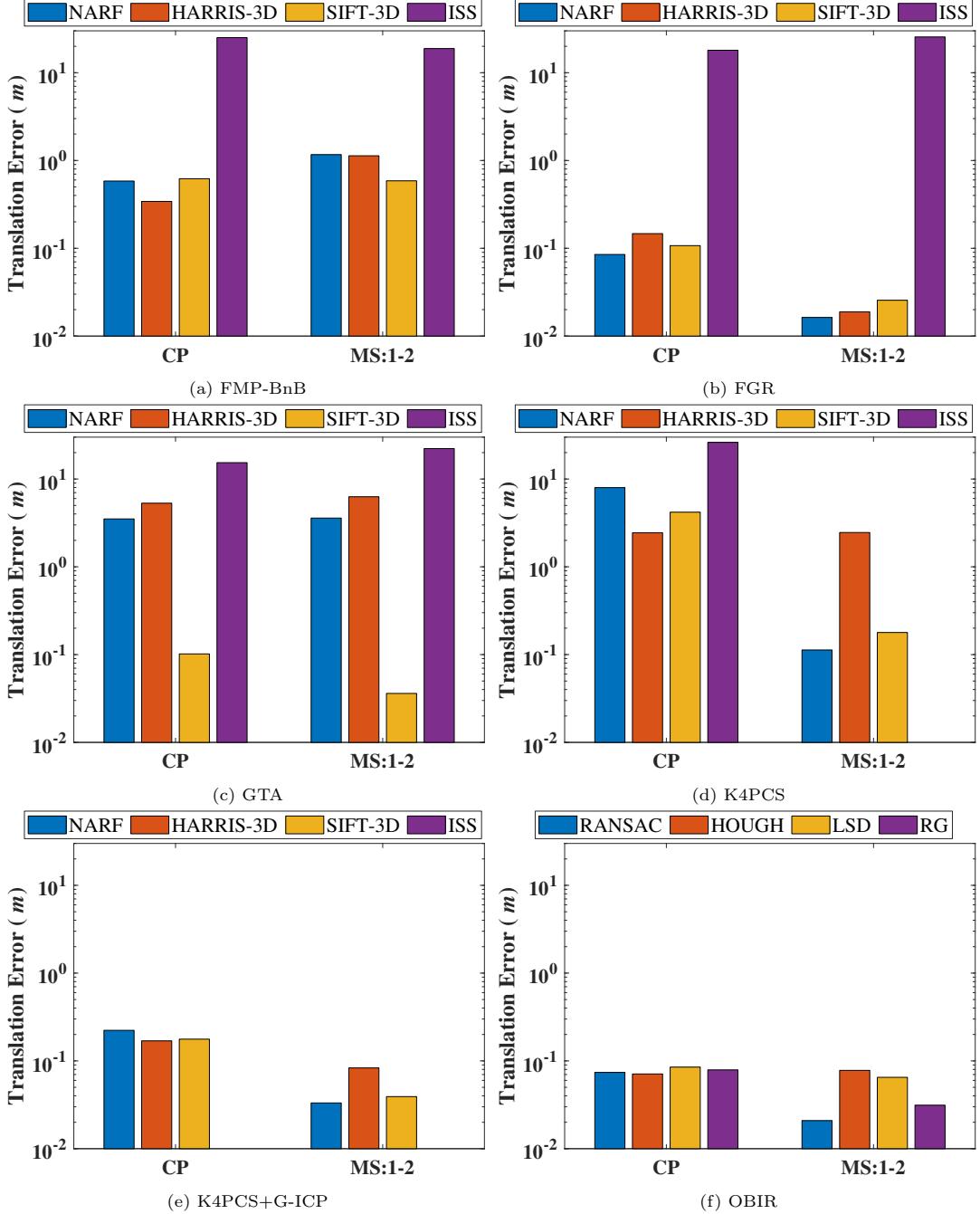


Figure 16: Translation errors for different keypoints or line configurations. Missing columns result from failures of convergences for specific configurations.

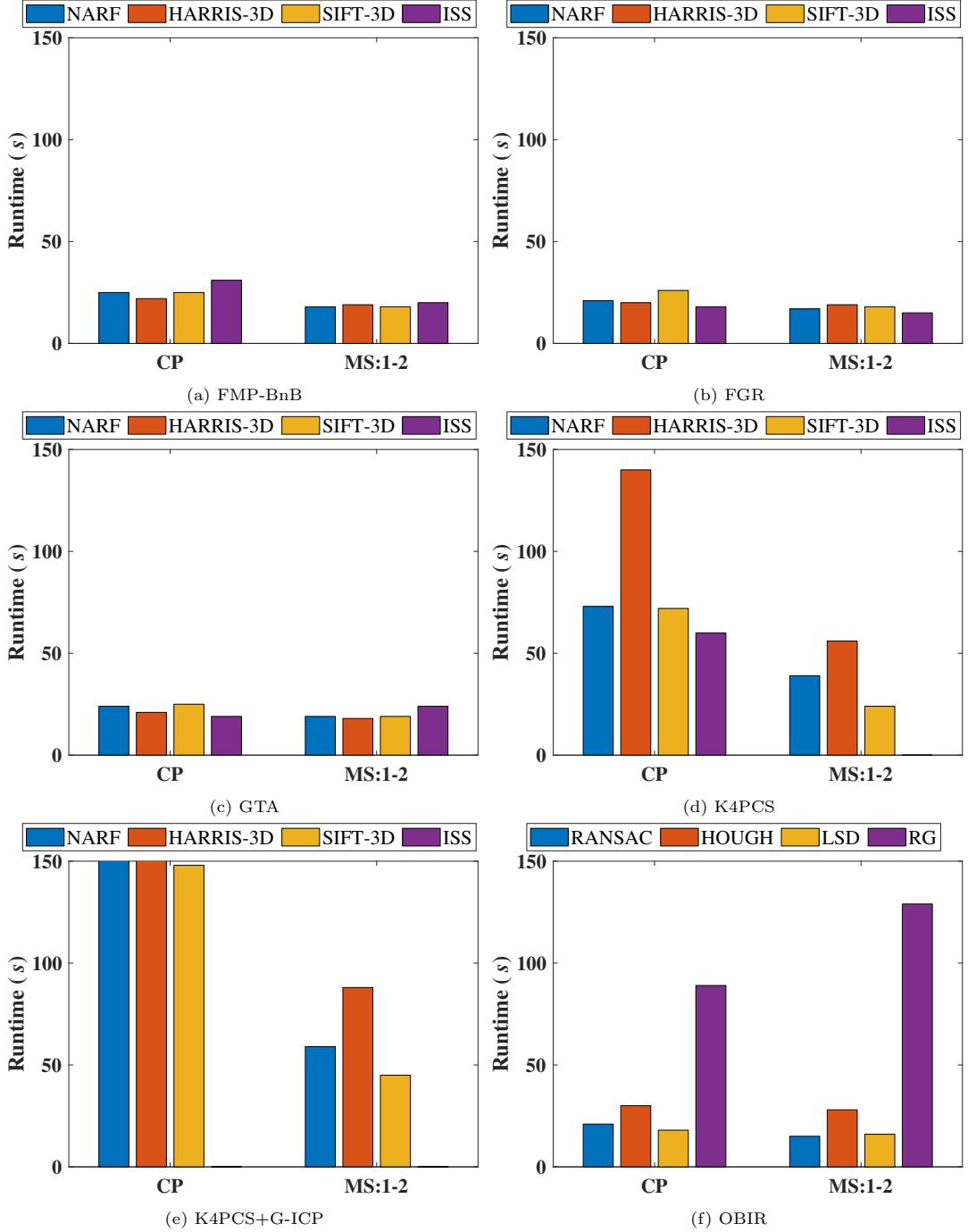


Figure 17: Runtimes for various keypoints or line configurations. Missing columns result from failures of convergences for specific configurations. To allow a fair comparison, the time taken to extract the keypoints/lines is also counted.

407 search space of the registration: the first reduces the unknowns, and the second reduces the
408 candidate pools of the correspondences. Only the first step, the 2D transformation, involves
409 global correspondence searching, for which the robustness is improved by the line primitives,
410 and the following two steps are local searching with least-squares optimization, for which robust
411 estimation and uniformly selected patches are used to increase the reliability. In an urban
412 environment, the proposed methods are superior to or *on par* with the state-of-the-art methods
413 with regard to both accuracy and runtime.

414 As noted above, the line primitives may be problematic in other scenarios, but the object-
415 based strategy can be exploited further in the future. For example, other structured or point-
416 like objects would also be useful, such as individual canopies and street furniture. In addition,
417 registration is not limited to only a single type of primitives; hybrid structures and configurations
418 can be considered, such as point-line configurations rather than the line-line structure used in
419 this study.

420 References

- 421 Agarwal, S., Mierle, K., Others, 2010. Ceres solver. <http://ceres-solver.org>.
- 422 Aiger, D., Mitra, N.J., Cohen-Or, D., 2008. 4-points congruent sets for robust pairwise surface
423 registration, in: ACM Transactions on Graphics (TOG), Acm. p. 85.
- 424 Al-Durgham, K., Habib, A., Kwak, E., 2013. Ransac approach for automated registration of
425 terrestrial laser scans using linear features. International Archives of the Photogrammetry,
426 Remote Sensing and Spatial Information Sciences 2, 13–18.
- 427 Albarelli, A., Rodolà, E., Torsello, A., 2015. Fast and accurate surface alignment through an
428 isometry-enforcing game. Pattern Recognition 48, 2209–2226.
- 429 Bae, K.H., 2009. Evaluation of the convergence region of an automated registration method for
430 3d laser scanner point clouds. Sensors 9, 355–375.
- 431 Bolles, R.C., Fischler, M.A., 1981. A ransac-based approach to model fitting and its application
432 to finding cylinders in range data., in: IJCAI, pp. 637–643.
- 433 Cai, Z., Chin, T.J., Bustos, A.P., Schindler, K., 2019. Practical optimal registration of terrestrial
434 lidar scan pairs. ISPRS Journal of Photogrammetry and Remote Sensing 147, 118–131.
- 435 Cheng, L., Chen, S., Liu, X., Xu, H., Wu, Y., Li, M., Chen, Y., 2018a. Registration of laser
436 scanning point clouds: A review. Sensors 18, 1641.
- 437 Cheng, X., Cheng, X., Li, Q., Ma, L., 2018b. Automatic registration of terrestrial and airborne
438 point clouds using building outline features. IEEE Journal of Selected Topics in Applied Earth
439 Observations and Remote Sensing 11, 628–638.
- 440 Date, H., Yokoyama, T., Kanai, S., Hada, Y., Nakao, M., Sugawara, T., 2018. Efficient regis-
441 tration of laser-scanned point clouds of bridges using linear features. International Journal of
442 Automation Technology 12, 328–338.
- 443 Dold, C., Brenner, C., 2006. Registration of terrestrial laser scanning data using planar patches
444 and image data. International Archives of the Photogrammetry, Remote Sensing and Spatial
445 Information Sciences-ISPRS Archives 36 (2006) 36, 78–83.

- 446 Dong, Z., Yang, B., Liang, F., Huang, R., Scherer, S., 2018. Hierarchical registration of unordered
 447 tls point clouds based on binary shape context descriptor. ISPRS Journal of Photogrammetry
 448 and Remote Sensing 144, 61–79.
- 449 Dong, Z., Yang, B., Liu, Y., Liang, F., Li, B., Zang, Y., 2017. A novel binary shape context
 450 for 3d local surface description. ISPRS Journal of Photogrammetry and Remote Sensing 130,
 451 431–452.
- 452 Ge, X., 2017. Automatic markerless registration of point clouds with semantic-keypoint-based
 453 4-points congruent sets. ISPRS Journal of Photogrammetry and Remote Sensing 130, 344–357.
- 454 Ge, X., Hu, H., Wu, B., 2019a. Image-guided registration of unordered terrestrial laser scanning
 455 point clouds for urban scenes. IEEE Transactions on Geoscience and Remote Sensing doi:[10.1109/TGRS.2019.2925805](#).
- 456 Ge, X., Wu, B., 2019. Contextual global registration of point clouds in urban scenes. Photogram-
 457 metric Engineering & Remote Sensing 85, 559–571.
- 458 Ge, X., Wu, B., Li, Y., Hu, H., 2019b. A multi-primitive-based hierarchical optimal approach
 459 for semantic labeling of als point clouds. Remote Sensing 11, 1243.
- 460 Ge, X., Wunderlich, T., 2016. Surface-based matching of 3d point clouds with variable coordi-
 461 nates in source and target system. ISPRS Journal of Photogrammetry and Remote Sensing
 462 111, 1–12.
- 463 Hackel, T., Savinov, N., Ladicky, L., Wegner, J.D., Schindler, K., Pollefeys, M., 2017. SEMAN-
 464 TIC3D.NET: A new large-scale point cloud classification benchmark, in: ISPRS Annals of the
 465 Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 91–98.
- 466 Hana, X.F., Jin, J.S., Xie, J., Wang, M.J., Jiang, W., 2018. A comprehensive review of 3d point
 467 cloud descriptors. arXiv preprint arXiv:1802.02297 .
- 468 Hartley, R., Zisserman, A., 2003. Multiple view geometry in computer vision. Cambridge uni-
 469 versity press.
- 470 Holz, D., Ichim, A.E., Tombari, F., Rusu, R.B., Behnke, S., 2015. Registration with the point
 471 cloud library: A modular framework for aligning in 3-d. IEEE Robotics & Automation Mag-
 472 azine 22, 110–124.
- 473 Hu, H., Ding, Y., Zhu, Q., Wu, B., Xie, L., Chen, M., 2016. Stable least-squares matching
 474 for oblique images using bound constrained optimization and a robust loss function. ISPRS
 475 Journal of Photogrammetry and Remote Sensing 118, 53–67.
- 476 Khoshelham, K., 2016. Closed-form solutions for estimating a rigid motion from plane correspon-
 477 dences extracted from point clouds. ISPRS Journal of Photogrammetry and Remote Sensing
 478 114, 78–91.
- 479 Lei, H., Jiang, G., Quan, L., 2017. Fast descriptors and correspondence propagation for robust
 480 global point cloud registration. IEEE Transactions on Image Processing 26, 3614–3623.
- 481 Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. International journal
 482 of computer vision 60, 91–110.

- 484 Papon, J., Abramov, A., Schoeler, M., Worgotter, F., 2013. Voxel cloud connectivity
485 segmentation-supervoxels for point clouds, in: Proceedings of the IEEE Conference on Com-
486 puter Vision and Pattern Recognition, pp. 2027–2034.
- 487 Poreba, M., Goulette, F., 2015. A robust linear feature-based procedure for automated regis-
488 tration of point clouds. Sensors 15, 1435–1457.
- 489 Rusinkiewicz, S., Levoy, M., 2001. Efficient variants of the icp algorithm., in: 3DIM, pp. 145–152.
- 490 Rusu, R.B., Blodow, N., Beetz, M., 2009. Fast point feature histograms (fpfh) for 3d registration,
491 in: 2009 IEEE International Conference on Robotics and Automation, IEEE. pp. 3212–3217.
- 492 Rusu, R.B., Cousins, S., 2011. 3d is here: Point cloud library (pcl), in: 2011 IEEE International
493 Conference on Robotics and Automation, IEEE. pp. 1–4.
- 494 Schnabel, R., Wahl, R., Klein, R., 2007. Efficient ransac for point-cloud shape detection, in:
495 Computer Graphics Forum, Wiley Online Library. pp. 214–226.
- 496 Schneider, P., Eberly, D.H., 2018. Geometric tools for computer graphics. Elsevier.
- 497 Segal, A., Haehnel, D., Thrun, S., 2009. Generalized-icp., in: Robotics: Science and Systems,
498 Seattle, WA. p. 435.
- 499 Stamos, I., Leordeanu, M., 2003. Automated feature-based range registration of urban scenes
500 of large scale, in: 2003 IEEE Computer Society Conference on Computer Vision and Pattern
501 Recognition, IEEE. pp. II–Ii.
- 502 Sumi, T., Date, H., Kanai, S., 2018. Multiple tls point cloud registration based on point projection
503 images. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial
504 Information Sciences XLII-2, 1083–1090.
- 505 Theiler, P., Schindler, K., et al., 2012. Automatic registration of terrestrial laser scanner point
506 clouds using natural planar surfaces. ISPRS Annals of Photogrammetry, Remote Sensing and
507 Spatial Information Sciences 3, 173–178.
- 508 Theiler, P.W., Wegner, J.D., Schindler, K., 2014. Keypoint-based 4-points congruent sets—
509 automated marker-less registration of laser scans. ISPRS Journal of Photogrammetry and
510 Remote Sensing 96, 149–163.
- 511 Von Gioi, R.G., Jakubowicz, J., Morel, J.M., Randall, G., 2008. Lsd: A fast line segment
512 detector with a false detection control. IEEE Transactions on Pattern Analysis and Machine
513 Intelligence 32, 722–732.
- 514 Vosselman, G., Coenen, M., Rottensteiner, F., 2017. Contextual segment-based classification
515 of airborne laser scanner data. ISPRS Journal of Photogrammetry and Remote Sensing 128,
516 354–371.
- 517 Weber, T., Hänsch, R., Hellwich, O., 2015. Automatic registration of unordered point clouds
518 acquired by kinect sensors using an overlap heuristic. ISPRS Journal of Photogrammetry and
519 Remote Sensing 102, 96–109.
- 520 Xiao, J., Adler, B., Zhang, J., Zhang, H., 2013. Planar segment based three-dimensional point
521 cloud registration in outdoor environments. Journal of Field Robotics 30, 552–582.

- 522 Xu, Y., Boerner, R., Yao, W., Hoegner, L., Stilla, U., 2017. Automated coarse registration
523 of point clouds in 3d urban scenes using voxel based plane constraint. ISPRS Annals of
524 Photogrammetry, Remote Sensing & Spatial Information Sciences 4.
- 525 Xu, Y., Boerner, R., Yao, W., Hoegner, L., Stilla, U., 2019. Pairwise coarse registration of
526 point clouds in urban scenes using voxel-based 4-planes congruent sets. ISPRS Journal of
527 Photogrammetry and Remote Sensing 151, 106–123.
- 528 Yang, B., Dong, Z., Liang, F., Liu, Y., 2016. Automatic registration of large-scale urban scene
529 point clouds based on semantic feature points. ISPRS Journal of Photogrammetry and Remote
530 Sensing 113, 43–58.
- 531 Yang, B., Zang, Y., Dong, Z., Huang, R., 2015. An automated method to register airborne
532 and terrestrial laser scanning point clouds. ISPRS Journal of Photogrammetry and Remote
533 Sensing 109, 62–76.
- 534 Yao, A., Gall, J., Van Gool, L., 2010. A hough transform-based voting framework for action
535 recognition, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern
536 Recognition, IEEE. pp. 2061–2068.
- 537 Zhong, Y., 2009. Intrinsic shape signatures: A shape descriptor for 3d object recognition, in:
538 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops,
539 IEEE. pp. 689–696.
- 540 Zhou, Q.Y., Park, J., Koltun, V., 2016. Fast global registration, in: European Conference on
541 Computer Vision, Springer. pp. 766–782.
- 542 Zhu, Q., Li, Y., Hu, H., Wu, B., 2017. Robust point cloud classification based on multi-level
543 semantic relationships for urban scenes. ISPRS Journal of Photogrammetry and Remote
544 Sensing 129, 86–102.