The International Journal of
Robotics Research
0(0) 1–7
© The Author(s) 2012
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0278364912458814
ijr.sagepub.com

Challenging data sets for point cloud registration algorithms

François Pomerleau, Ming Liu, Francis Colas and Roland Siegwart

Abstract

The number of registration solutions in the literature has bloomed recently. The iterative closest point, for example, could be considered as the backbone of many laser-based localization and mapping systems. Although they are widely used, it is a common challenge to compare registration solutions on a fair base. The main limitation is to overcome the lack of accurate ground truth in current data sets, which usually cover environments only over a small range of organization levels. In computer vision, the Stanford 3D Scanning Repository pushed forward point cloud registration algorithms and object modeling fields by providing high-quality scanned objects with precise localization. We aim to provide similar high-caliber working material to the robotic and computer vision communities but with sceneries instead of objects. We propose eight point cloud sequences acquired in locations covering the environment diversity that modern robots are susceptible to encounter, ranging from inside an apartment to a woodland area. The core of the data sets consists of 3D laser point clouds for which supporting data (Gravity, Magnetic North and GPS) are given for each pose. A special effort has been made to ensure global positioning of the scanner within mm-range precision, independent of environmental conditions. This will allow for the development of improved registration algorithms when mapping challenging environments, such as those found in real-world situations.[1]

Keywords

Field robots, field and service robotics, search and rescue robots, range sensing, sensing and perception, computer vision

1. Motivation

Urban environment navigation has received much attention in recent years and has triggered the creation of large-scale data sets, several km long (Smith et al., 2009; Huang et al., 2010; Pandey et al., 2011). Even though these data sets are undeniably very useful, other platforms, like the ones used for Search and Rescue missions, encounter a broader range of environments in which the robustness of localization needs to be assessed. Many environments that are likely to be faced are composed of complex structures, and some of them have particular problematic features such as a forest with dense foliage (see Figure 1) that shades GPS signals. On the registration side, the planarity of the environment was taken for granted in early implementations (Chen and Medioni, 1991) and until recent versions of scan-matching algorithms (Pathak et al., 2010). Clearly, there is a need for semi-structured and unstructured data sets to challenge this planar hypothesis and to validate the robustness of registration solutions in a variety of environments that are encountered in the real world. Recently, Peynot et al. (2010) presented data sets that highlight various situations, but the focus was on atmospheric conditions (airborne dust, smoke and rain). We continue in the same direction but for land-based studies by proposing data sets that cover a larger spectrum of environmental structures, so registration solutions can be further evaluated.

In this data-oriented paper, we present 8 sequences of around 35 point clouds each. The sequences were selected to challenge point cloud registration algorithms with respect to: semi-structured and unstructured environments, rapid variation of scanning volumes, repetitive elements, and finally, dynamic elements. Given that we targeted global positioning evaluations, special attention was given to the methodology used to record ground truth poses with a consistent protocol for all sequences.

2. Ground truth localization

The notion of *ground truth* is highly dependent on the intention of use and can hardly be absolute. The error of the

Autonomous Systems Laboratory, ETH Zürich, Switzerland

Corresponding author:
François Pomerleau, Autonomous Systems Lab, ETH Zürich, Tannenstrasse 3, 8092 Zurich, Switzerland.
Email: francois.pomerleau@mavt.ethz.ch

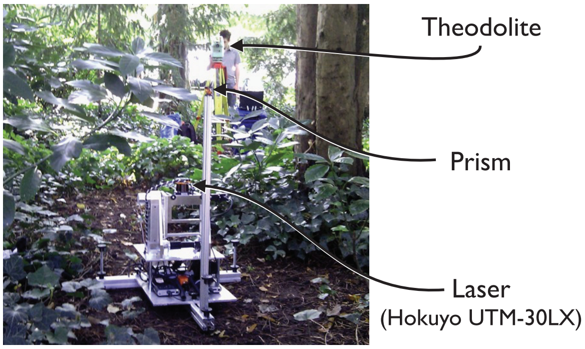


Fig. 1. The scanner in a targeted unstructured environment with dense foliage cover.

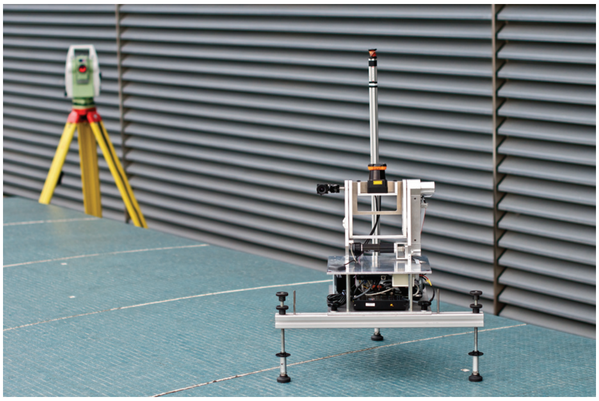


Fig. 2. Tilting scanner with the prism mounted at p_0.

reference should be significantly lower than the expected outcome of the algorithm to achieve a fair comparison.

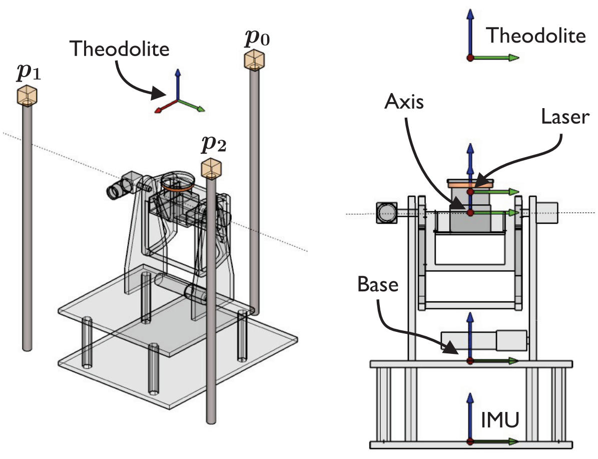
Precise global positioning can be reached using an arm that is fixed on a base holding a scanner, but this solution offers a limited motion range. On the other hand, GPS and DGPS systems can accommodate a large range of motion but are limited to outdoor locations displaying clear-sky conditions. The precision of such systems can be highly variable (i.e. depending on foliage coverage, satellite alignment and number, multi-paths, etc.), which also limits the evaluation of registration precision. Optical motion capture systems, like the one proposed by Vicon, have recently appeared as a precise way to track sensor poses (Pomerleau et al., 2011). These systems offer mm precision at 100 Hz, but cannot be installed outdoors or in highly cluttered environments. Instead of using fixed sensors and mobile markers, Tong and Barfoot (2011) proposed a methodology to directly reuse laser reflectivity readings combined with some reflective beacons. This is a convenient way to ensure ground truth localization in open space, but would lead to the installation of multiple landmarks in a highly occluded environment, like a forest. Finally, the Jet Propulsion Laboratory used a theodolite to track specialized prisms fixed on a mobile platform to validate visual odometry performance (Maimone et al., 2007). The precision reported was less than 2 mm in position, and less than 0.2° in attitude. In addition to the precision, the system reduces infrastructure installation and ensures a fixed precision over all recorded sequences, independent of environmental location and conditions, which is why we have applied this technique to our data sets.

3. Materials and methodology

The data sets were recorded with a partially custom-made rotating scanner used in conjunction with a theodolite, as depicted in Figure 2. The main sensor of the scanner is a laser rangefinder (Hokuyo UTM-30LX) mounted on a tilting device. The sensor has a compact size (87 × 60 × 60 mm) and covers a field of view of 270° with a reading at every 0.25°. A comparative study, realized by Wong et al. (2011), concluded that the Hokuyo UTM-30LX has comparable precision and accuracy to the SICK LMS-200. The precise control of the motor was ensured by a Maxon Motor EPOS controller. The control system put in place used a dual regulation loop based on two encoders. One encoder was located directly on the motor shaft to provide control stability while the second was located at the end of the transmission chain. The encoders had 2,000 and 48,000 ticks per revolution, respectively, the precision difference coping with the gear reduction employed. The latter encoder gave us a resolution of 0.00013 rad on the tilting axis. This setup removes the uncertainty from gear backlash and transmission strap deformation, which was estimated to be around 5 deg on a formal prototype. Supporting data (Gravity, Magnetic North and GPS) was provided by a consumer-grade GPS-aided IMU, Xsens MTi-G.

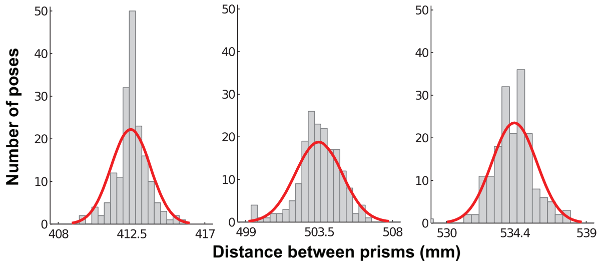
The theodolite used was the Total Station (TS15) from Leica Geosystems. As it only measures one position at a time, and three measurements are necessary to retrieve the complete pose (translation and orientation), a specialized reflective prism was mounted on a pole, which could be secured at three different locations on the scanner, namely p_0, p_1 and p_2 [see Figure 3(a)]. A steel guide ensured that the pole was positioned at the same location on the scanner every time. The pole was higher than the scanner to reduce visual occlusion from the theodolite.

Most of the recording process was done manually. The scanner was moved from one location to another by an operator. Extra precautions were taken to ensure that the scanner stayed in place while scanning (usually 20 s) and while the ground truth pose was measured (under 2 min). On hard floors, rubber feet were used whereas on soft grounds, metal spikes were used. The inertia of the platform also guaranteed good stability while recording a scan. In some cases, like in a compartmented area such as an apartment, a single line of sight cannot track all poses. For those situations, we changed the theodolite pose and then used the last scanner pose as a fixed beacon to globally relocalize the theodolite. We carefully planned those relocalizations to minimize their number, so that for all the sequences, we never had to



(a) (b)

Fig. 3. Configuration of the scanner. The dashed line corresponds to the rotation axis. (a) Perspective view with the positions of the three prisms used to reconstruct the global pose. (b) Reference frame notation.



(a) $\mu_{12} = 412.5$ mm $\sigma_{12} = 1.2$ mm (b) $\mu_{01} = 534.4$ mm $\sigma_{01} = 1.4$ mm (c) $\mu_{02} = 503.5$ mm $\sigma_{02} = 1.4$ mm

Fig. 4. Histograms of the distances between prisms (mm) measured by the theodolite. (a) Distances between p_1 and p_2. (b) Distances between p_0 and p_2. (c) Distances between p_0 and p_1.

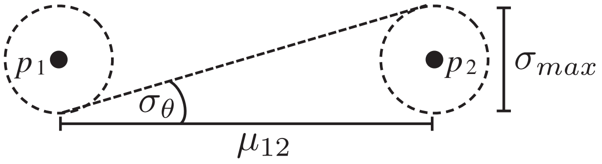


Fig. 5. The worst-case orientation error given the position error σ_max and the smallest expected distance μ_{12} between the prisms p_1 and p_2.

relocalize more than twice. We acknowledge that the overall system is costly and time-consuming (e.g. 3 h for 35 scans), but we firmly believe that this methodology is necessary to ensure that high-quality data sets are available for further research.

All sensor data were logged on the same computer so the data are timestamped based on the same clock. Supporting data were recorded at a different frequency to the laser, and they were segmented per 3D scan pose. On the accompanying website, we also present a post-processed version of the supporting data, in which the average values per 3D scan pose can be used.

3.1. Noise evaluation

In this section, sources of noise, from the global pose down to a measured laser point, are overviewed. The theodolite used has an accuracy proposed by the manufacturer of 1 mm per km. Given the fact that we do not have access to an additional and more precise sensor to validate the ground truth, we evaluated the distances between each prism (d_{12}, d_{02} and d_{01}) over 181 scanner poses measured in different conditions. Figure 4 presents the resulting histograms. The maximum standard deviation (σ_max) of the three distances is 1.4 mm. Since it is the distance between two points, we can assume that σ of one point is ∼1.0 mm. This error includes the noise of the theodolite and some human manipulation errors while moving the prism from one position to another.

In the field, we used these inter-prism statistics to cancel spurious pose measurements before taking the 3D scan. The translation component of the global pose was obtained using the mean of the three prism positions, which would lead to a total global position error, σ_t, of $\frac{1.0}{\sqrt{3}} = 0.58$ mm under the assumption of isotropic Gaussian noise. For the rotational components, we used the smallest estimated distance between prisms ($\mu_{12} = 412.5$ mm) and when using basic geometry from Figure 5, we can estimate an angular error (σ_θ) of 0.003 rad. These errors apply if the theodolite is kept at the same place during a complete data set recording, which was not the case for three of the data sets. Table 1 shows that, at most, the theodolite was moved two times. After a simple error propagation, we can approximate σ_t to be under 1.8 mm and σ_θ to be under 0.006 rad, which is consistent with the level of precision reported by Maimone et al. (2007). As for the link between the *Theodolite* and the *Base*, the transformation was computed using a global optimization technique explained in more detail on the supplementary website. To evaluate residual errors, we used a different data set to the one used for the optimization.

As for the transformations from *Laser* to *Base*, most of them were taken from the construction plans and were machined with a precision of under a mm in cm-thick aluminum plates. Since encoders work in relative position, a homing procedure needs to be applied to reset the count of the encoders. The offset between the homing position and the position of the rotating frame that is parallel to the base is directly added by the low-level controller (EPOS). This offset was measured using two off-the-shelf laser pointers, typically used for public presentations, fixed on the tilting *Axis* and on the *Base*. The two laser points were projected onto a wall at a distance of 8 m. The angle was adjusted to ensure that the distances between the projected points and the laser pointers were the same. We roughly estimated the homing error σ_h to be under 0.001 rad.

Finally, the Hokuyo UTM-30LX is a time-of-flight sensor with a minimum range of 0.1 m and a maximal range

Table 1. Characteristics of the point clouds for each data set.

Sequence name	No. Scans	No. Points per scan	No. Relocalizations	Pose volume ($x \times y \times z$)	Scene volume ($x \times y \times z$)
ETH Hauptgebaude	36	191,000	0	24×2×0.50m	62×65×18m
Apartment	45	365,000	2	5×5×0.06m	17×10×3m
Stairs	31	191,000	0	10×3×2.50m	21×111×27m
Gazebo Summer	32	170,000	1	5×4×0.07m	35×45×16m
Gazebo Winter	32	153,000	1	4×5×0.09m	72×70×19m
Mountain Plain	31	102,000	0	18×6×2.70m	36×40×8m
Wood Summer	37	182,000	0	10×15×0.50m	30×53×20m
Wood Autumn	32	178,000	0	6×12×0.50m	36×60×22m

of 30 m. The specifications of the sensor propose a range accuracy, σ_r, varying from 0.01 to 0.03 m depending on the distance and reflectivity of the object. Values for the transformations between the different frames depicted in Figure 3(b) are listed in Table 2 with their estimated precision. We used a right-handed coordinate system with the x-axis pointing forward, y-axis on the left and z-axis upward. All the transformations are given the notation $T_{X \leftarrow Y}$, which can be read as: a transformation T that can express a point, originally in the Y coordinate frame, in an X coordinate frame. The translation vector t is represented as $[t_x, t_y, t_z]$ and the rotation vector q is represented as a quaternion $[q_x, q_y, q_z, q_w]$, where q_w is the real part of the quaternion.

As a general observation, very small angular misalignments can have a large impact on the point location at large distances, especially for highly slanted surfaces. For example, we had to manually tune the orientation of the laser to the tilting axis by a third of a degree to ensure that a single point cloud joins properly after a rotation of 180°. This slight offset might be due to tolerances in the construction or related to the divergence of the laser beam. Although the global pose of the scanner is on the order of mm, it is most likely that the uncertainty of the reflected points in the environment is way larger when the beams have a diameter of several cm at a few m distance. This uncertainty is inherent to the sensors and occurs in most robotic systems. Further evaluations should be considered to give more precise error bounds.

4. Overview of the data sets

The aim of the proposed data sets is to provide unregistered point clouds for researchers who are seeking to evaluate their registration solutions on a common base. The point clouds are provided in *Base* frame, which can be compared against the measured global poses. We also provide globally consistent point clouds for researchers doing environmental modeling. Before presenting the specific sequences, we first introduce the nomenclature used to characterize the different sequences. The abbreviations defined below are reused in Table 3, which also presents an overview of the eight sequences recorded.

The organization of the environment is characterized as follows:

Structured (S): The environment can mainly be explained using geometric primitives (e.g. offices or buildings).

Unstructured (US): The environment mainly involves more complex structures (e.g. a dense forest or a very untidy room).

Semi-structured (SS): The environment has both geometric and complex elements (e.g. a partially collapsed building or a park essentially composed of flat ground and some trees).

Considering a static sensor pose, we also define three types of dynamic element:

Intra-scan motions (AM): An element is moving while the data are captured. The more time it takes to capture the data, the more deformed the element will be (e.g. pedestrians or cars). This is comparable to motion blur for a fixed camera.

Inter-scan motions (EM): A dynamic event occurs punctually with respect to data acquisition (e.g. furniture is moved or a door is opened).

Global motions (GM): An event affects the environment on a global scale, and dynamic elements are detected by multiple views recorded at different time periods (e.g. seasonal changes or a building collapsing).

Finally, environment locations are divided into two categories: **Outdoors** (OUT) and **Indoors** (IN).

The sequences were recorded over half a year (between August 2011 and January 2012). Figures 6 and 7 present a visual overview of all sequences showing the variety of environments covered. Table 1 gives the number of 3D scans, the average number of points per 3D scan and the number of times the theodolite was relocated for each data set. The two last columns give an indication of the volumes covered with a bounding box in which the scanner was moved (Pose volume) and with a bounding box of the global map (Scene volume).

4.1. *Unstructured environments*

The sequence named *Wood* is a good example of a challenging environment for registration algorithms that

Table 2. Relative transformations between frames.

Transformation	Sensor	Estimated pose		Estimated precision
$T_{T\leftarrow G}$	*Global* to *Theodolite*	t	variable	$0.0006 < \sigma_t < 0.0018$ m
		q	variable	$0.0030 < \sigma_\theta < 0.0060$ rad
$T_{B\leftarrow T}$	*Theodolite* to scanner *Base*	t	$[0.016\ -0.024\ 0.606]$ m	residual = 0.004 m
		q	$[0.000\ 0.010\ -0.006\ -0.999]$	residual = 0.004 rad
$T_{B\leftarrow A}$	Tilting *Axis* to scanner *Base*	t	$[0.000, 0.000, 0.220]$ m	by construction
		q	variable	$\sigma_h < 0.001$ rad
$T_{B\leftarrow I}$	*IMU* to scanner *Base*	t	$[0.000, 0.000, -0.085]$ m	by construction
		q	$[0.000, 0.000, 0.000, 1.000]$	by construction
$T_{A\leftarrow L}$	*Laser* to tilting *Axis*	t	$[0.000, 0.000, 0.040]$ m	by construction
		q	$[0.001, 0.000, -0.003, 0.999]$	by construction
$T_{L\leftarrow P}$	*Point* to *Laser*	$r \in [0.1, 10)$ m		$\sigma_r < 0.01$ m
		$r \in [10, 30]$ m		$\sigma_r < 0.03$ m

Table 3. Overview of the data sets with their characteristics.

Sequence name	IN	OUT	S	SS	US	AM	EM	GM	Particulars
ETH Hauptgebaude	✓		✓			✓			Repetitive elements like pillars.
Apartment	✓		✓				✓		Single-floor apartment with five rooms.
Stairs	✓	✓	✓						Rapid variations of scanning volumes.
Gazebo (×2)		✓		✓		✓	✓	✓	Recorded in summer and in winter.
Mountain Plain		✓			✓				Pasture with few vertical structures.
Wood (×2)		✓			✓	✓		✓	Recorded in summer and in autumn.

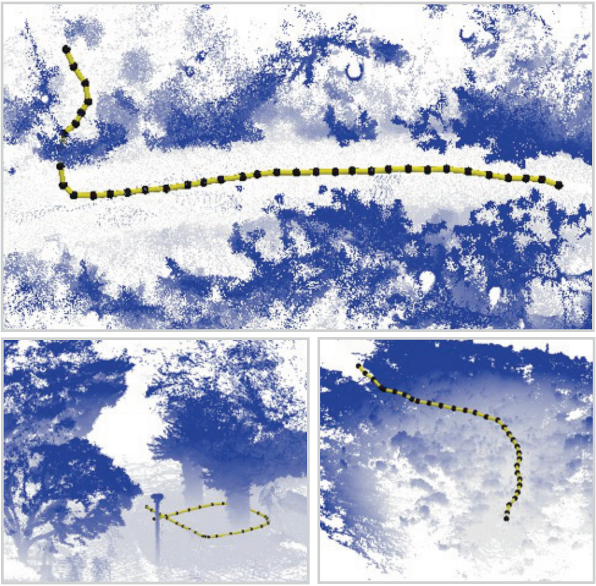


Fig. 6. Unstructured and semi-structured data sets. Top: aerial view of the sequence *Wood* with the upper part of the vegetation removed. Bottom left: part of the sequence *Gazebo* with some wine trees on the right and some large trees on the left. Bottom right: aerial view of the sequence *Mountain Plain*. For all figures, the yellow lines and black circles correspond to the scanner poses, and point clouds were colored to emphasize the depth of the structure from the virtual camera perspective.

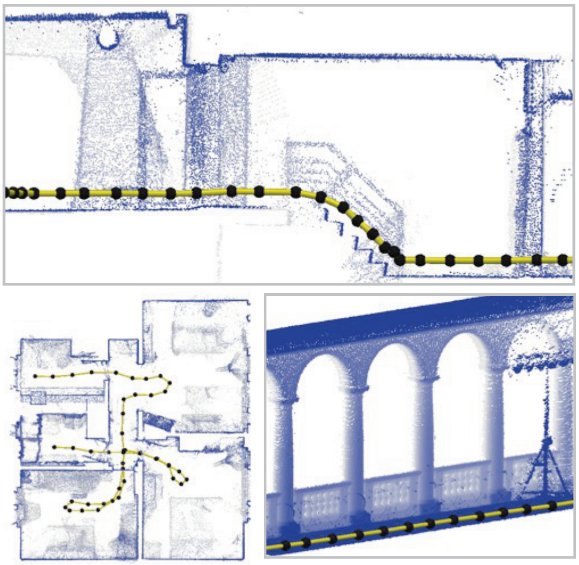


Fig. 7. Structured data sets. Top: side view of the sequence *Stairs*. Bottom left: top view of the sequence *Apartment* with the ceiling and floor removed. Bottom right: cut view of a hallway from the sequence *ETH Hauptgebaude* showing arches and pillars. For all figures, the yellow lines and black circles correspond to the scanner poses, and point clouds were colored to emphasize the depth of the structure from the virtual camera perspective.

contains both complex structures and intra-scan dynamic elements. Figure 1 shows the starting position of the recorded path. This environment mainly consists of vegetation (trees, bushes, etc.) with a small paved road crossing the wood as the only structural element. While recording

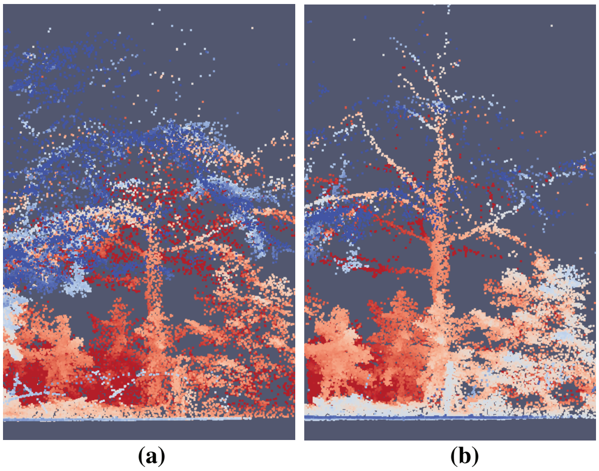


Fig. 8. Extracts of global representations highlighting seasonal changes: (a) summer, and (b) late autumn. Point cloud colors were selected to enhance the depth of the screenshot.

the data, some people were walking on the road. The scanner path starts in the wood and continues for approximately 12 scans before joining the small road for the next 14 scans. The sequence was recorded during two different seasons (i.e. summer and late autumn), which gave us the opportunity to test the registration algorithm's robustness against Global Motion (i.e. seasonal changes). Figure 8 shows a visual example of the impact of those changes on trees, which were manually extracted from the global map for each season.

Another sequence called *Mountain Plain* was recorded on a small area of an alpine plain located at 1920 m altitude. There is no major vertical structure in the environment and the main element on the ground is dry vegetation (around 50 cm height). The motivation behind this data set is to evaluate the robustness of registration algorithms against low-constrained, unstructured environments. The opposite of a low-constrained environment would be an apartment where the ceiling and walls are large enough to fix the position and orientation of the sensing platform easily. This data set is also very interesting because the hypothesis of the planar motion of the scanner does not hold since the scanner goes down a hill before ending up in a flat area.

4.2. Other environments

To ease comparisons between a more complete spectrum of environmental structures, we also provide five more sequences recorded with the same methodology. The two sequences named *Gazebo* were recorded in summer and winter in a park, in which there was grass, paved small roads and sparse trees. The main construction is a gazebo with rock walls and a ceiling covered with wines trees. This place is a good example of a semi-structured environment, with a mixture of man-made structures and vegetation. Some people were walking while the scanner was recording, whereas others stayed seated for several scans under the gazebo. The sequence called *Stairs* aims at evaluating the robustness of registration algorithms against rapid variations in scanned volumes. The path starts indoors, crosses some doorways and finishes outdoors. The scanner passes over five steps, which offers a more complex motion than a flat floor. The sequence *ETH Hauptgebaude* is able to tackle the problem of repetitive elements, including multiple pillars and arches in a hallway. Those elements may create multiple local minima, which can trigger interesting observations for registration algorithms. Finally, the sequence *Apartment* is a well-structured environment including: a kitchen, a living room, a bathroom, an office and a bedroom. Special care was taken to include outer-scan motion by moving a person, some furniture and boxes inbetween scans. The registration complexity of this environment is considered low, so it could be used as a reference for other types of environments. In both sequences *ETH Hauptgebaude* and *Apartment*, the scanner moved indoors on flat ground.

5. Data formats

All data are available as comma separated value (CSV) files with the first line consisting of a header. This format is natively supported by many languages and software (including Matlab and Python) and can be easily parsed. Point clouds are available in local coordinates [i.e. the frame named *Base* in Figure 3(b)] and in global coordinates. The provided ground truth poses directly give the transformation from the origin to the frame *Base* for a given scan. The axis origin of the global coordinate was selected to be the first scanner pose of each data set. Supporting data are given in the *IMU* frame. Moreover, we provide screenshots of all sensor information and photographs of the environments to facilitate understanding of the scene context. No ground truth information is available for dynamic elements included in the scenes. For a rapid overview of the data sets, we also provide VTK files in global coordinates. More explanation of all the file headers and content is available on the website.

6. Conclusion

In this paper, we introduced new data sets covering a diverse range of challenging environments for registration algorithms. Although some of these environments can be found in available data sets, our *Laser Registration Data Sets* englobe them all in a coherent group recorded with the same methodology and materials. We achieved precise localization of the scanner using a theodolite, which gave us the ability to record data sets in GPS-denied environments, indoors or outdoors with the same setup. The precision achieved is also higher than when using data sets that are already available to the community, which should ease the evaluation of registration algorithms on a fair base.

Note

1. The data sets and complementary information are publicly available under the section *Laser Registration Datasets* at: http://projects.asl.ethz.ch/datasets

Funding

This work was funded by the EU FP7 IP projects Natural Human-Robot Cooperation in Dynamic Environments (grant number ICT-247870). François Pomerleau was supported by a fellowship from the Fonds québécois de recherche sur la nature et les technologies.

Acknowledgements

The authors would like to thank Leica Geosystems and Benoit Lescot for their technical support, and also Dario Fenner and Markus Bühler for their input into the design and construction of the scanner. We thank ME Garneau for comments on the manuscript.

References

Chen Y and Medioni G (1991) Object modeling by registration of multiple range images. In: *IEEE international conference on robotics and automation (ICRA 1991)*, Sacramento, USA, 9–11th April 1991, pp. 2724–2729. Piscataway: IEEE Press.

Huang AS, Antone M, Olson E, Fletcher L, Moore D, Teller S et al. (2010) A high-rate, heterogeneous data set from the DARPA urban challenge. *The International Journal of Robotics Research* 29(13): 1595–1601.

Maimone M, Cheng Y and Matthies L (2007) Two years of visual odometry on the Mars exploration rovers. *Journal of Field Robotics* 24(3): 169–186.

Pandey G, McBride JR and Eustice RM (2011) Ford campus vision and lidar data set. *The International Journal of Robotics Research* 30(13): 1543–1552.

Pathak K, Birk A, Vaškevičius N and Poppinga J (2010) Fast registration based on noisy planes with unknown correspondences for 3-D mapping. *IEEE Transactions on Robotics* 26(3): 424–441.

Peynot T, Scheding S and Terho S (2010) The Marulan data sets: multi-sensor perception in a natural environment with challenging conditions. *The International Journal of Robotics Research* 29(13): 1602–1607.

Pomerleau F, Magnenat S, Colas F, Liu M and Siegwart R (2011) Tracking a depth camera: Parameter exploration for fast ICP. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS 2011)*, San Francisco, USA, 25–30 September 2011, pp. 3824–3829. Piscataway: IEEE Press.

Smith M, Baldwin I, Churchill W, Paul R and Newman P (2009) The New College vision and laser data set. *The International Journal of Robotics Research* 28(5): 595–599.

Tong CH and Barfoot TD (2011) A self-calibrating 3D ground-truth localization system using retroreflective landmarks. In: *IEEE international conference on robotics and automation (ICRA 2011)*, pp. 3601–3606. Piscataway: IEEE Press.

Wong U, Morris A, Lea C, Lee J, Whittaker C, Garney B et al. (2011) Comparative evaluation of range sensing technologies for underground void modeling. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS 2011)*, San Francisco, USA, 25–30 September 2011, pp. 3816–3823. Piscataway: IEEE Press.