DiVA✿

http://www.diva-portal.org

Preprint

This is the submitted version of a paper presented at *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Workshop "3D Mapping", Nice, France, September 2008.*

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:oru:diva-3262

3D Mapping the Kvarntorp Mine – A Field Experiment for Evaluation of 3D Scan Matching Algorithms

Martin Magnusson   Andreas Nüchter   Christopher Lörken   Achim J. Lilienthal   Joachim Hertzberg

Abstract— To advance robotic science it is important to perform experiments that can be replicated by other researchers to compare different methods. However, these comparisons tend to be biased, since re-implementations of reference methods often lack thoroughness and do not include the hands-on experiences obtained during the original development process. This paper presents the results of a field experiment, carried out by two research groups that are leading in the field of 3D robotic mapping. The iterative closest points algorithm (ICP) is compared to the normal distributions transform (NDT).

I. INTRODUCTION

Experimental methodologies for robotic mapping received recently a lot of attention in the community: Firstly, scientists start to define rules for experiments [1]. Secondly, research in mapping is fostered by open source projects such as *Radish: The Robotics Data Set Repository* [8] and OpenSLAM [12]. These sites offer some interesting algorithms but currently cover only 2D mapping methods. Thirdly, comparing robotic systems in competitions like RoboCup [6], ELROB [7] or the Grand Challenge [5] is increasing. These kinds of competitions allows the level of system integration and the engineering skills of a certain team to be ranked, but it is not possible to measure the performance of a subsystem or a single algorithm.

This paper presents the results of a 3D mapping field experiment in the Kvarntorp mine outside of Örebro in Sweden. Automated mapping and localisation for underground mining vehicles is a current goal of the mining industry [9], [11]. In addition to the application of autonomous robots, accurate 3D mine models can also be used for other applications, such as verifying that tunnels have the desired shape and size, measuring the volume of removed material, and surveying old tunnels to investigate whether they are still safe.

3D mapping of the underground mine has been used to compare two scan matching methods, namely the iterative closest point algorithm (ICP) and the normal distributions transform (NDT). The experimental results of the algorithm are compared in terms of robustness and speed. For robustness we measure how reliably 3D scans are registered with respect to different starting pose estimates. Speed is evaluated running

Martin Magnusson and Achim J. Lilienthal are with the Centre for Applied Autonomous Sensor Systems in the Department of Technology at University of Örebro, Fakultetsg. 1, S-70182 Örebro, Sweden. Contact: martin.magnusson@tech.oru.se

Andreas Nüchter, Christopher Lörken and Joachim Hertzberg are with the Knowledge Systems Research Group of the Institute of Computer Science, University of Osnabrück, Germany. Contact: nuechter@informatik.uni-osnabrueck.de

the authors' best implementations on the same hardware. This leads to an unbiased comparison.

II. PROBLEM STATEMENT

Pairwise scan registration is the process of aligning two overlapping scans, given an estimate of the relative transformation needed to match one with the other. When the scans are properly aligned, they are said to be in registration. Following the nomenclature of Besl and McKay [2], the scan that serves as the reference is called the *model* and the scan that is moved into alignment with the model is called the *data* scan.

III. RELATED WORK

We have investigated two algorithms for matching pairs of independently acquired 3D scans [2]–[4], [9], [10]:

A. ICP

The iterative corresponding point algorithm (ICP) iteratively calculates the point correspondences (see [2], [4]). In each iteration, the algorithm selects the closest points as correspondences and calculates the transformation $(\mathbf{R}, \mathbf{t})$ for minimizing the equation

$$E(\mathbf{R}, \mathbf{t}) = \sum_{i=1}^{N_m} \sum_{j=1}^{N_d} w_{i,j} \, ||\mathbf{m}_i - (\mathbf{R}\mathbf{d}_j + \mathbf{t})||^2 \, ,$$

where N_m and N_d, are the number of points in the model set M and data set D, respectively, and w_{ji} are the weights for a point match. The weights are assigned as follows: $w_{ji} = 1$, if $\mathbf{m}_i$ is the closest point to $\mathbf{d}_j$ within a close limit, $w_{ji} = 0$ otherwise.

B. NDT

The normal distributions transform (NDT) uses another representation of the model (see [3], [9]). Instead of using the individual points of the model point cloud, it is represented by a combination of normal distributions, describing the probability of finding part of the surface at any point in space. The normal distributions give a piecewise smooth representation of the model point cloud, with continuous first and second order derivatives. Using this representation, it is possible to apply standard numerical optimisation methods for registration.
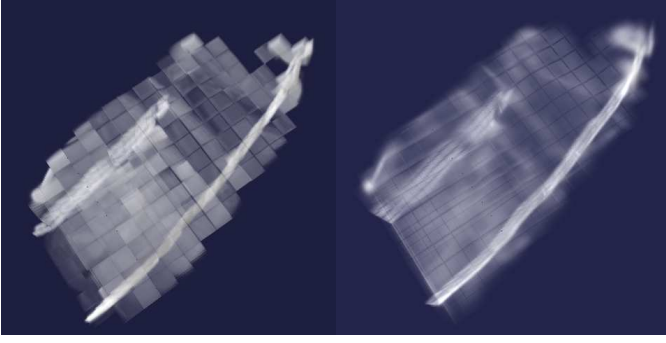
Fig. 1. Illustration of applying NDT to the model scan in data set A, with (right) and without (left) trilinear interpolation. Denser regions represent larger score values. (The dark grid pattern does not represent smaller score values, but only shows the borders of the underlying cells.)

C. NDT with trilinear interpolation

The discretisation artifacts that come from subdividing the space into cells, leading to discontinuities in the surface representations at cell edges, can sometimes be problematic. In the original 2D NDT implementation [3], the discretisation effects were minimised by using four overlapping 2D cell grids. A similar approach was implemented here, using the normal distributions from the eight neighbouring cells at each evaluation of the score function, with the weight of the contribution from each cell is determined by trilinear interpolation. In other words, if $\mathbf{x}' = T(\mathbf{p}, \mathbf{x})$ is point $\mathbf{x}$ transformed by the current transformation parameters $\mathbf{p}$, the score function from [9],

$$s(\mathbf{p}) = -\frac{1}{c} \exp\left(-\frac{(\mathbf{x}' - \mathbf{q})^T \mathbf{C}^{-1} (\mathbf{x}' - \mathbf{q})}{2}\right),$$

is replaced with

$$s(\mathbf{p}) = -\frac{1}{c} \sum_{b=1}^{8} w(\mathbf{x}', \mathbf{q}_b) \exp\left(-\frac{(\mathbf{x}' - \mathbf{q}_b)^T \mathbf{C}_b^{-1} (\mathbf{x}' - \mathbf{q}_b)}{2}\right),$$

where $\{\mathbf{q}_b\}$ and $\{\mathbf{C}_b\}$ are the means and covariances of the PDFs of the eight cells which are closest to $\mathbf{x}'$, and $w(\mathbf{x}', \mathbf{q}_b)$ is a trilinear interpolation weight function. This has a smoothing effect similar to the approach of Biber and Straßer without the need to compute more probability distributions (see Fig. 1). Because up to eight distribution functions have to be evaluated for each point (less then eight if the model surface does not occupy all of the surrounding cells), the algorithm takes up to eight times as long as NDT without trilinear interpolation (in our experiments, the median execution time increased by around 450% percent).

IV. TECHNICAL APPROACH

To compare the performance of ICP and NDT with respect to mine mapping, we proceeded as follows: For each of the selected scan pairs, a reference pose was determined and the registration algorithms were run at a number of start poses with varying translation and rotation offsets from the reference pose. We then counted which start poses resulted in an end pose sufficiently close to the reference pose. We limited the offsets of the initial pose estimates to rotations and translations in the horizontal plane. This constraint can be motivated for three reasons: first, in a typical mine mapping scenario, the largest part of the error will lie in the horizontal plane; second, it reduces the number of trials that must be run (we tried 441 start poses, using the same offsets on all transformation parameters would make 250 047 poses); third, it makes the results easier to visualise. No constraints were added to the registration algorithms; they still operate with six degrees of freedom. Unfortunately ground truth data are not available in this type of field experiment. The reference poses were therefore determined manually, by performing a number of registrations and choosing the mean of the poses that led to visually correct results. Because of the low accuracy of this referencing, all registrations resulting in a pose within a specified translation and rotation distance from the reference pose were regarded as "successful". We chose two translation thresholds: a stricter one (0.20 m), and a weaker one (1.0 m). The rotation threshold used was 5°. Poses within the stricter translation threshold are difficult to tell apart for a human observer. Poses with larger translation errors are clearly less exact matches, but may still be considered good enough for some applications.

In addition to this scan-to-scan evaluation we executed both algorithms with incremental pairwise scan matching, i.e., each scan was registered against the previous scan. During the experiments, we closed several loops, and therefore, we can measure the transformation that is necessary to match the first scan against the last on of a closed loop. By doing so, we measured the accumulated error of both methods.

Because of the sometimes large odometry errors that come from driving a small robot over loose rocks, the initial pose estimate had to be manually altered for some scan pairs in order to reach convergence. As another measurement of robustness, we counted the number of occasions where the odometry had to be corrected.

V. EXPERIMENTS AND RESULTS

A. Data

The 3D range data were acquired by a tiltable 3D laser scanner based on a SICK LMS 200. A small servo motor has been attached to the SICK to perform a controlled pitch motion. The resolution of a 3D scan is 361×226 data points covering the area of about $180° \times 116.3°$ in front of the robot. 3D scanning did proceed in a drive-scan-and-go fashion.

The data were collected by Kurt3D (cf. Fig. 3) in the Kvarntorp mine, south of Örebro in Sweden. This mine is no longer in production, but was once used to mine limestone. Fig. 2 shows a typical scene from one of the tunnels. The mine consists of around 40 km of tunnels, all in approximately one plane.

The following data sets were used for the comparisons:

Data set A: Two partly overlapping scans from a slightly curved tunnel section. Subsets of the original scans were used, with 8000 samples drawn from each scan so that the resulting

Fig. 2.   One of the tunnels in the Kvarntorp mine.

Fig. 3.   The Kurt3D robot scanning underground.

point clouds had relatively even densities (around 10% of the points were used). The scans are shown in Fig. 4.

Data set B: A sequence of 55 scans, going around a loop, with the last two scans partly overlapping the first scan. See Fig. 5. Again, each scan was subsampled to 8000 points. Data set A is scans number 32 and 33 from this set. The total distance traveled around the loop is about 150 m.

B. Experiments

The results from the scan-to-scan registration experiments are presented in plots where the translation offsets are layed out along the x and y axes of the plot and the rotation offsets are shown as points around a circle. In other words, each group of points shows the results from nine start poses with the same translation but different rotations. See Fig. 6.

To quantify the registration accuracy, a reference pose for the last scan of data set B was determined by registering it to the first scan. The difference between the reference pose and the resulting pose after pairwise registration of all scans of the data set was used as a measure of the algorithms'

accuracy. The initial pose estimate for each scan was taken from odometry.

C. Parameters

The following parameters were used:
NDT:
- Iterative NDT with cell sizes 2 m, 1 m, and 0.5 m. This means that for each registration attempt, NDT was run three times with successively smaller cell sizes, with the end pose from each run being used as the start pose for the next one. The first iterations roughly align scan pairs with large initial pose error, and the last iterations refine the result because the surface model is more precise.
- Linked cells (unoccupied cells store a pointer to the closest occupied cell) and infinite outer bounds (points that fall outside the cell grid during registration are matched to the closest occupied cell).
- Rotations parametrised as Euler angles with small-angle approximations. In other words, rotations are represented as triples $R(x, y, z)$ meaning three consecutive rotations around the main coordinate axis. This gives a six-dimensional optimisation problem (three from translation and three from rotation). Using the small-angle approximations $sin(x) \approx x$ and $cos(x) \approx 1$ is accurate enough when the rotation in each Newton iteration is small, and slightly decreases execution time.
- Optimisation using Newton's method with line search. Max step size $||\Delta \mathbf{p}|| = 0.2$, where $\mathbf{p}$ is the translation and rotation parameters of the current pose, measured in metres and radians. Max 100 iterations (but the iteration limit was never reached). Convergence threshold: step size $||\Delta \mathbf{p}|| < 10^{-6}$ or score decrement $\Delta s < 0$.

ICP:
- For closest point computation we used standard k-d tree search, employing a bucket size of 10 points per bucket.
- Distance threshold for point pairs 0.5 m. Data points whose current nearest neighbour in the model scan is beyond the distance threshold are treated as outliers and discarded. Furthermore, this threshold takes care of partially overlapping scans, i.e., using this threshold aims to
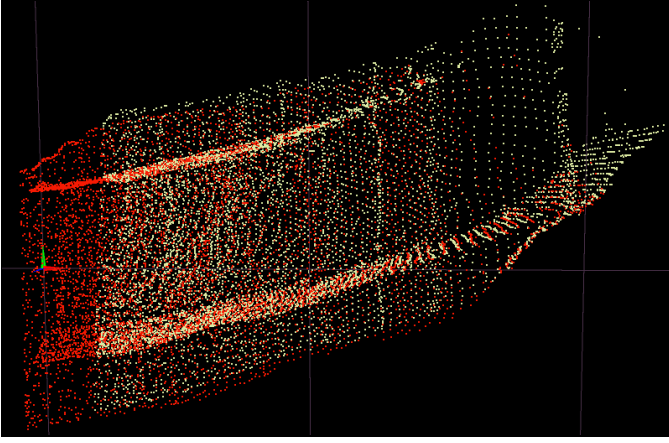
Fig. 4. The two scans of data set A at the reference pose, seen from above. The data scan is light (yellow) and the model scan is dark (red). The x axis points to the right, the y axis points up, and the z axis points towards the viewer in this figure.

Fig. 5. Data set B, seen from above after loop closure.

• Convergence threshold: step size $||\Delta E(\mathbf{R}, \mathbf{t})|| < 10^{-6}$.

D. Results

1) Valley of convergence: The sensitivity to error in the initial pose estimate was tested using data set A. Fig. 7 shows that ICP failed for most of the attempts where the initial pose was translated backwards (in the $-x$ direction). Although the rotation of the pose estimate after registration was generally correct, the algorithm stopped prematurely in these cases at a pose with maximum overlap between the two scans. NDT overcame this local optimum in more cases. However, for the cases where NDT did fail, it was sometimes the case that both the translation and rotation of the final pose were wrong. In other words, NDT succeeded more often, but for the cases where it failed, the result was sometimes worse than for ICP. A registration result where the rotation is well-aligned but the translation is off along the tunnel's direction is often more
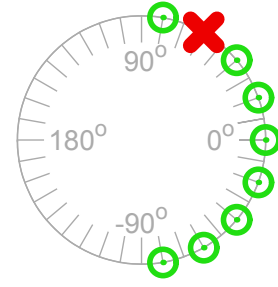

Fig. 6. Legend to the registration plots. Each sub-plot represents a set of initial poses with the same translation offset and varying rotation offsets. Green circles represent successful registrations, and red crosses represent failures. Poses with initial rotation error ranging from -80° to +80° in 20° increments were tested.

acceptable than a result with large rotation error. If the rotation error only is used as the criterion for successful registration, the results look quite different, as can be seen in Fig. 8.

The execution times are shown in Fig. 11. The reported times include all necessary preprocessing (including creation of the normal distributions for NDT, and a kd-tree for ICP) and all three iterations for NDT, but exclude the time needed for loading the scan data.

2) Outlier count: When registering data set B, the initial pose of one scan had to be adjusted both for ICP and standard NDT. For ICP, one scan (number 33) could not be aligned without adjusting the odometry. For NDT, scan number 23 had to be altered. NDT with trilinear interpolation successfully registered all scans from their original pose estimates.

3) Registration accuracy: The registration accuracy was measured by looking at the accumulated pose error at the end of data set B.

For NDT, the accumulated translation error was 2.26 m and the rotation error was 1.9°(using the altered initial pose for scan 23). The translation error vector was $[1.118, -0.02027, -1.965]$, which means that the accumulated vertical error was almost 2 m. Most of the the horizontal translation error was because the more feature-less tunnel segments were somewhat "shortened". A close-up of the registration result is shown in Fig. 12.

For NDT with trilinear interpolation, the accumulated error was slightly larger in this case: 3.99 m and 1.3°(using the original pose estimates). The translation error vector was $[3.217, -0.5633, -2.296]$. See Fig. 12.

For ICP an accumulated translation error of 2.97 m can be reported.

VI. CONCLUSIONS

In these experiments, NDT was shown to converge from a larger range of initial pose estimates than ICP, and to perform faster. However, the poses from which NDT converged were not as predictable as for ICP. In several cases, a scan would be
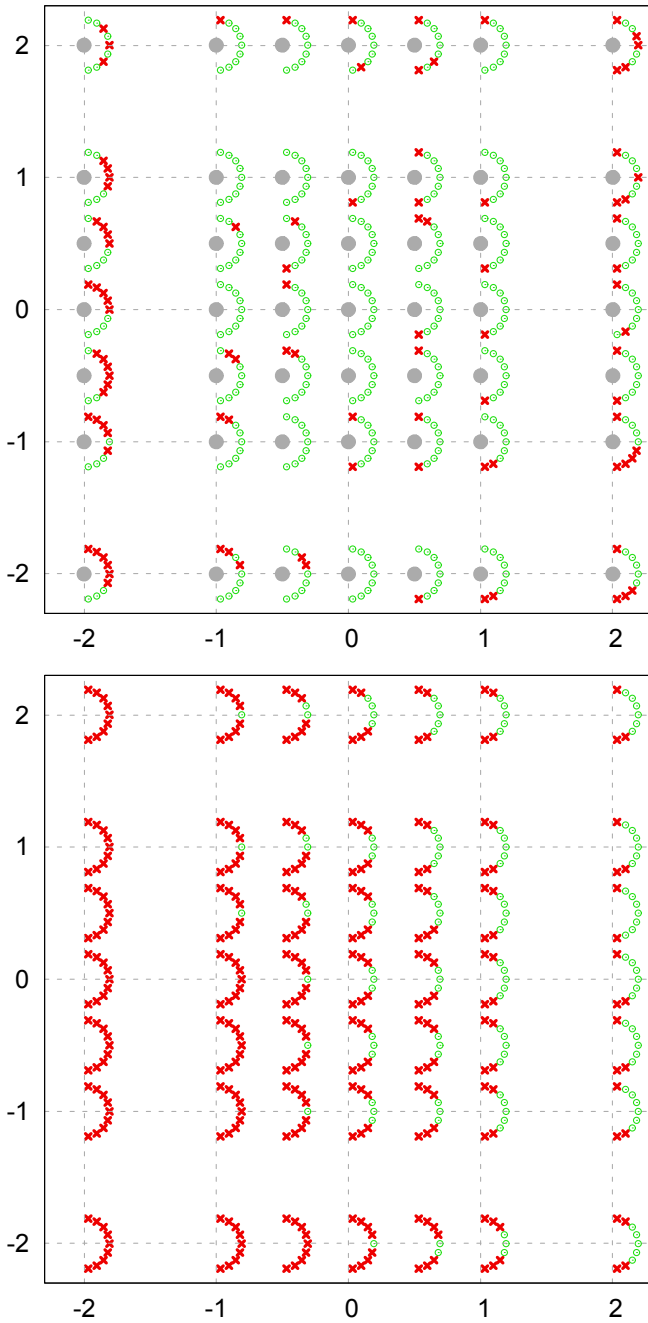
Fig. 7. Registration results from data set A, using the loose success threshold. Initial translation offsets along the two horizontal axes are on the x and y axis in the plot, and the initial rotation offsets are shown around the circle of each sub-plot, as described in Fig. 6. NDT above, ICP below. Success rate: 77% for NDT, 30% for ICP.
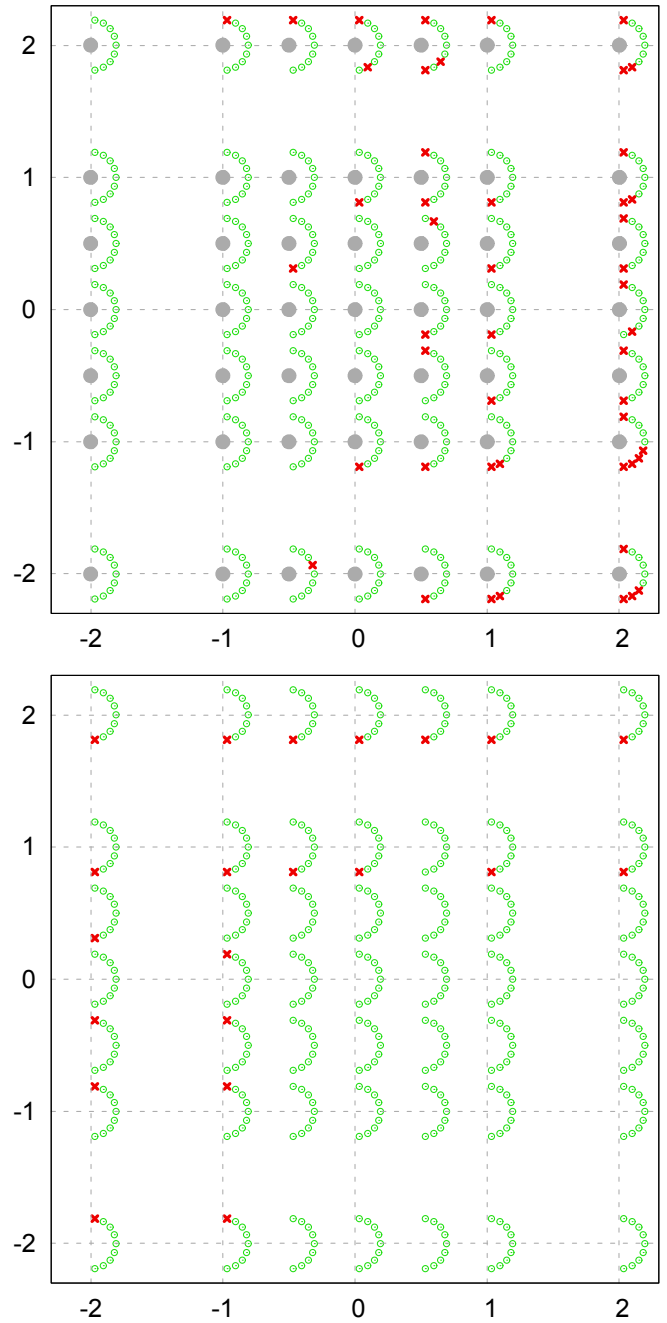

Fig. 8. Registration results from data set A, judging by rotation error only. NDT above, ICP below. Success rate: 89% for NDT, 95% for ICP.

successfully registered from a pose estimate with large initial error but fail from a pose estimate with less error. Also, in some cases where NDT failed, the resulting pose was worse than the result of ICP, because the rotation error was larger. Using NDT with trilinear interpolation further increased the success rate of NDT, at the expense of longer execution times.

REFERENCES

[1] F. Amigoni, S. Gasparini, and M. Gini. Good experimental methodologies for robotics mapping: A proposal. In *Proceedings of the 2007 IEEE International Conference on Robotics and Automation*, pages 4176–4181, Rome, Italy, April 2007.
[2] Paul J. Besl and Neil D. McKay. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239 – 256, February 1992.
[3] Peter Biber and Wolfgang Straßer. The normal distributions transform: A new approach to laser scan matching. In *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*, volume 3, pages 2743–2748, 2003.
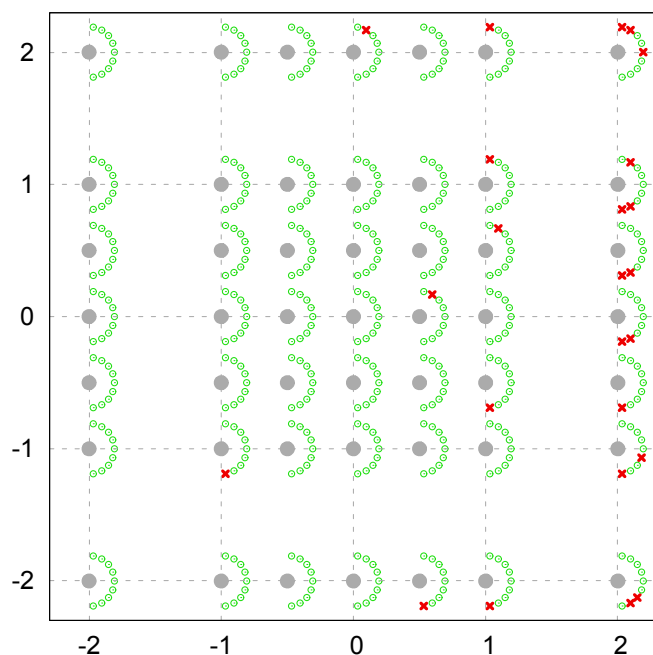
Fig. 9. Registration of data set A with NDT using trilinear interpolation. Please note that the strict translation threshold was used for this figure. All other parameters were the same in these runs and the ones shown in Fig. 7. Success rate: 95%.
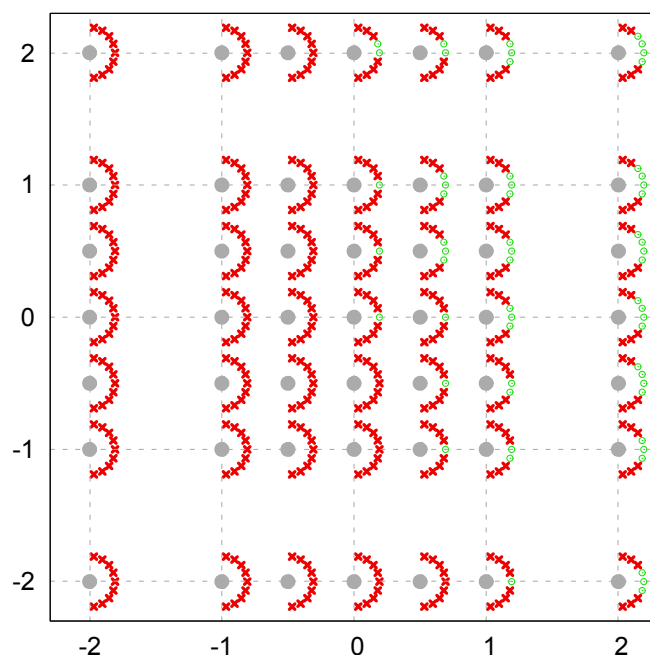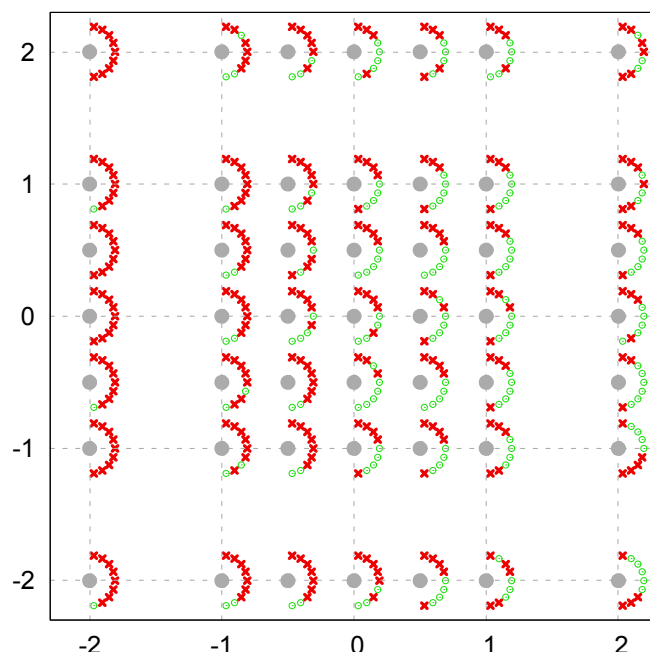
[4] Yang Chen and Gérard Medioni. Object modelling by registration of multiple range images. *Image and Vision Computing*, 10(3):145–155, April 1992.

[5] DARPA. www.darpa.mil/grandchallenge/.

[6] The RoboCup Federation. http://www.robocup.org/.

[7] FGAN. http://www.elrob2006.org/.

[8] A. Howard and N. Roy. http://radish.sourceforge.net/.

[9] Martin Magnusson, Achim Lilienthal, and Tom Duckett. Scan registration for autonomous mining vehicles using 3D-NDT. *Journal of Field Robotics*, 24(10):803–827, 2007.

[10] Andreas Nüchter, Kai Lingemann, Joachim Hertzberg, and Hartmut Surmann. 6D SLAM – 3D Mapping Outdoor Environments. *Journal of Field Robotics*, 2007.

[11] Andreas Nüchter, Hartmut Surmann, Kai Lingemann, and Joachim Hertzberg. 6D SLAM with an application to autonomous mine mapping. In *Proceedings of the 2004 IEEE International Conference on Robotics & Automation*, April 2004.

[12] C. Stachniss, U. Frese, and G. Grisetti. http://www.openslam.org/.

Fig. 10. Registration of data set A using the strict translation threshold. NDT above, ICP below. Success rate: 37% for NDT, 13% for ICP.
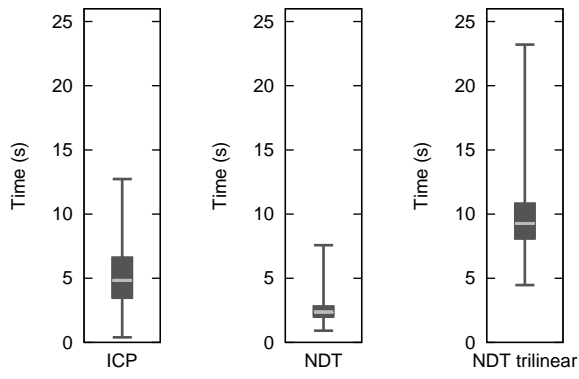
Fig. 11. Execution times for data set A. The light bar shows the median execution time from the 441 runs, the "whiskers" extend to the extreme values, and the edges of the box show the first and third quartile.

ICP

NDT

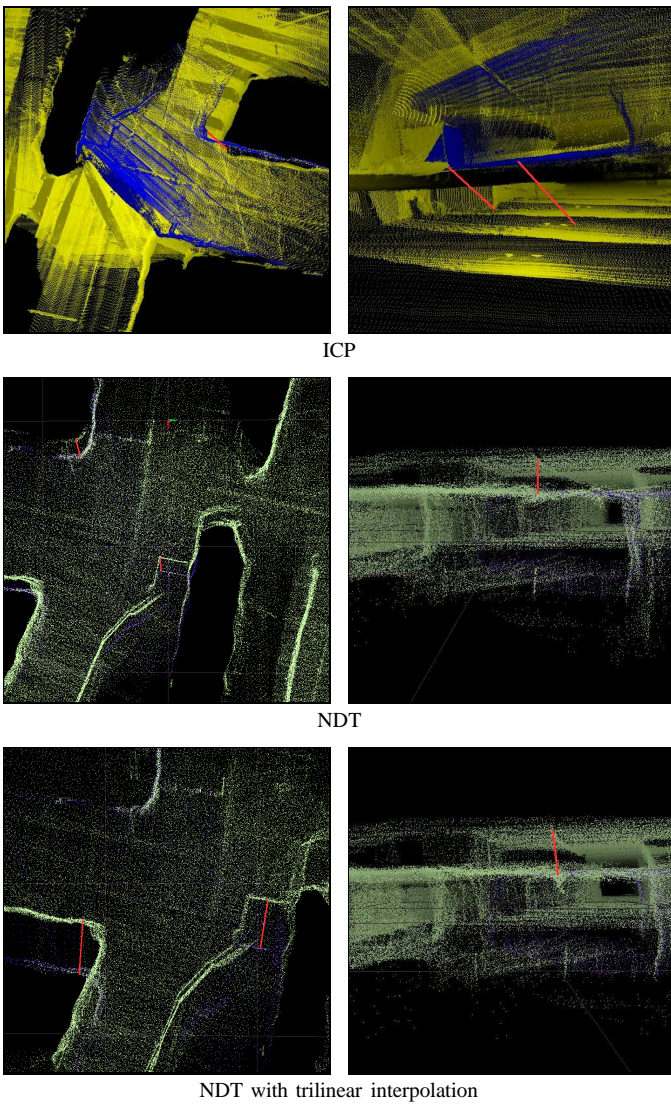NDT with trilinear interpolation

Fig. 12. The accumulated error after registering all of the scans in data set B, using NDT. The last scan is shown in blue. The red lines connect features that should line up if there were no error. The left column shows a top view, and the right column shows a horizontal view.