Evaluation of 3D Registration Reliability and Speed – A Comparison of ICP and NDT

Martin Magnusson, Andreas Nüchter, Christopher Lörken, Achim J. Lilienthal, and 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 experience obtained during the original development process. This paper presents a thorough comparison of 3D scan registration algorithms based on a 3D mapping 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). We also present an improved version of NDT with a substantially larger valley of convergence than previously published versions.**

I. Introduction

Experimental methodologies for robotic mapping have recently received a lot of attention in the community: Firstly, scientists have started to define rules for experiments [1]. It is important to follow certain guidelines so that experiments are replicable and show all relevant characteristics of the investigated system. Secondly, research in mapping is fostered by open source projects such as *Radish: The Robotics Data Set Repository* [8] and OpenSLAM [14]. 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 allow 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. As an attempt to provide such a detailed subsystem analysis, this paper presents a thorough comparison of 3D scan registration algorithms based on 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 [10], [12]. When constructing a metric map from sequential laser scans, it is necessary to know how the robot has moved between scans. Odometry or inertial sensors typically drift wildly even over short distances, so in general it is necessary to use pairwise registration to determine the exact transformation of the robot.

Pairwise scan registration is the process of aligning two overlapping scans based on the shapes of their overlapping sections, given an estimate of the relative transformation (or pose) 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.

The algorithms investigated here are local optimisation algorithms, which means that they only converge to the correct pose if given an initial estimate close enough to the correct pose. Typically, robot odometry is used to supply the initial estimate, but in many cases, and especially in mines or when traversing other uneven terrain, the pose given by odometry may be too far from the right solution for pairwise registration to succeed. Therefore we have specifically investigated how wide the valley of convergence is for mine tunnel data using different registration algorithms; that is, how large a range of poses that leads to good registration. We have also compared the execution speed by running the authors' best implementations on the same hardware. This leads to an unbiased comparison.

II. Related work

We have investigated two algorithms for matching pairs of independently acquired 3D scans: ICP [2], [4], [11] and NDT [3], [10]. ICP is the de facto standard 3D registration algorithm used in the robotics community, and NDT has appeared to be a compelling alternative in previous comparisons [10]. Our groups have long experience working with the two algorithms under review, which is why we can confidently say that this is a fair comparison.

We are interested in full 3D metric mapping with six degrees of freedom. Other previous work on metric mine maps [15] has used semi-3D mapping with rigidly mounted 2D laser scanners, assuming that one or more rotation angles (such as the roll angle) are static.

Global registration algorithms, which do not require an initial pose estimate, also exist, and have been used in mine mapping scenarios [9]. However, as long as some initial estimate from odometry is available, local registration is much more computationally efficient than global registration.

A. ICP

The iterative corresponding point algorithm (ICP) iteratively minimises point-to-point distances between the two

This work was supported in part by Atlas Copco Rock Drills.

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

Andreas Nüchter is with the Jacobs University Bremen, Campus Ring 1, 28759 Bremen, Germany. Contact: andreas@nuechti.de

Andreas Nüchter, Christopher Lörken and Joachim Hertzberg are with the Knowledge Systems Research Group of the Institute of Computer Science, University of Osnabrück, Germany.

scans (see [2], [4]). In each iteration, the algorithm selects the closest points as correspondences and calculates the transformation $(\mathbf{R}, \mathbf{t})$ for minimising the equation

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

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

B. NDT

The normal distributions transform (NDT) uses another representation of the model scan (see [3], [10]). 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.

C. NDT with trilinear interpolation

NDT has an inherent limitation in the fact that space is subdivided into regular cells. The discretisation artifacts that come from the subdivision process, leading to discontinuities in the surface representations at cell edges, can sometimes be problematic. It is often the case that the probability distribution function (PDF) for a cell is substantially non-zero even at points outside the cell's borders. 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. The weight of the contribution from each cell is determined by trilinear interpolation. In other words, if $\mathbf{x}'$ is point $\mathbf{x}$ transformed by the current transformation parameters $\mathbf{p}$, the score function from [10],

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

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 [3] without the need to compute more PDFs (see Fig. 1). Because up to eight PDFs 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 execution time increased by around 450%.

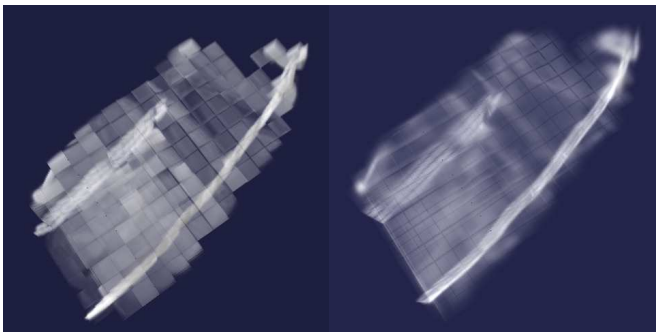


Fig. 1. Illustration of applying NDT to the model scan in data set A, with (right) and without (left) trilinear interpolation. Denser regions represent larger score values. (The dark grid pattern does not represent smaller score values, but only shows the borders of the underlying cells.)

III. Technical Approach

To compare the performance of ICP and NDT with respect to mine mapping, we performed three experiments: firstly we thoroughly investigated the valley of convergence for one pair of scans, secondly we examined the success rate using a larger data set using initial start poses from robot odometry, and thirdly we evaluated the cumulative registration accuracy for the larger data set.

A. Valley of convergence

In order to investigate the valley of convergence, we selected one scan pair and determined a reference pose. The registration algorithms were run from a number of start poses, each of them offset from the reference pose, both with respect to translation and rotation. 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: firstly, in a typical mine mapping scenario, the largest part of the error will lie in the horizontal plane; secondly, 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); thirdly, 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". For scans with "lock and key" features, such as walls in different directions, it is easier to determine a rather exact reference pose. In the mine tunnel application, however, many scans are from relatively featureless tunnels. For such scans it is more difficult to find the best pose parameters, especially the translation along the direction of the tunnel. We chose two translation thresholds: a

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

Fig. 3. The Kurt3D robot scanning underground.

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, but may still be considered good enough for some applications. For navigation, 1 m accuracy should be sufficient as long as relative positioning is more accurate.

B. Manual intervention

In addition to this scan-to-scan evaluation we executed both algorithms with incremental pairwise scan matching on all scans of the data set; i.e., each scan was registered against the previous scan. 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 a usable final pose estimate. As another measurement of robustness, we counted the number of occasions where the odometry had to be corrected for a sequence of scans. The correct poses in this scenario were also selected manually from visual inspection, marking as incorrect only those attempts that resulted in gross registration error.

C. Registration accuracy

During the experiment, we closed a loop, and therefore we can measure the transformation that is necessary to match the first scan against the last one when returning to the starting point. By doing so, we measured the accumulated error. If the accumulated error is small, the final scan should align well with the first scan.

IV. 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° × 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 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 travelled around the loop is about 150 m.

The data are available online for download [13].

B. Experiments

The results from the pairwise 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. Each group of points shows the results from nine start poses with the same translation but different rotations. (See Fig. 6 for clarification.)

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

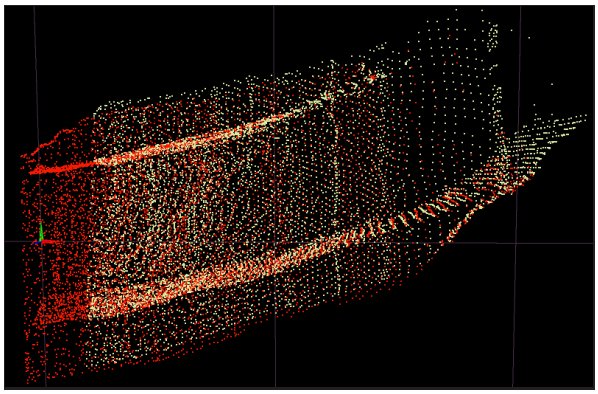


Fig. 4. The two scans of data set A at the reference pose, seen from above. The data scan is light (yellow) and the model scan is dark (red). In this figure, the x axis points to the right, the y axis points to the top of the page, and the z axis points towards the viewer.

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

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. In other words, this threshold aims to minimise the tendency to drag the two scans to a maximally overlapping pose.
- 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 the initial poses for which the algorithms converged to a good solution. ICP failed for most of the attempts where the

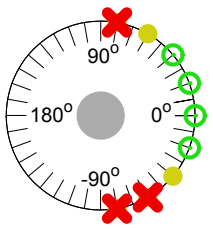
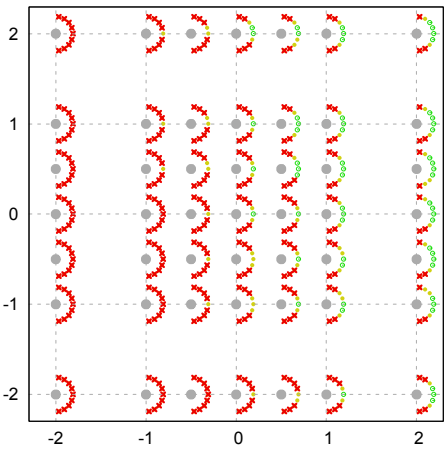


Fig. 6. Legend to the plots in Fig. 7–8. Each sub-plot represents a set of initial poses with the same translation offset and varying rotation offsets. Green circles represent successful registrations using the strict translation threshold, solid yellow dots represent successes using the loose threshold, and red crosses represent failures. For each translation offset, poses with initial rotation error ranging from -80° to +80° in 20° increments were tested. The grey dot marks the translation offset.

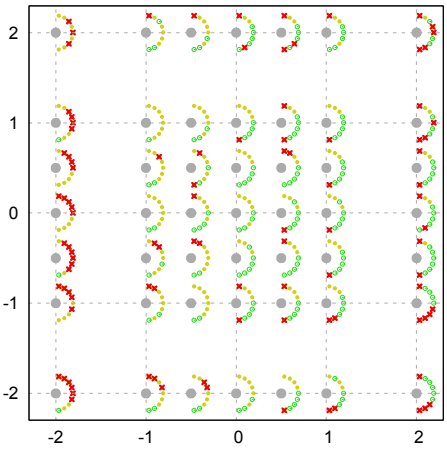
initial pose was translated backwards (in the $-x$ direction). Although the rotation of the pose estimate after registration was generally correct, the algorithm stopped prematurely in these cases at a pose with maximum overlap between the two scans. NDT overcame this local optimum in more cases. However, 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 acceptable than a result with large rotation error. If the rotation error only is used as the criterion for successful registration, the results look different, as can be seen in Fig. 8.

The execution times are shown in Fig. 9. 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.

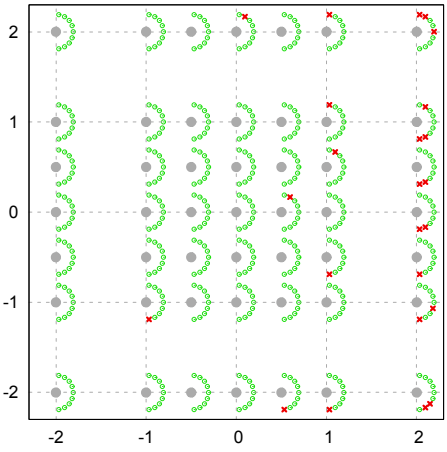
2) Manual intervention: When registering data set B, the initial pose of one scan had to be adjusted both for ICP



(a) ICP

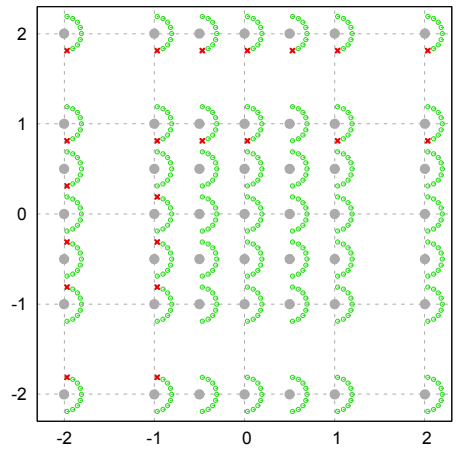


(b) NDT

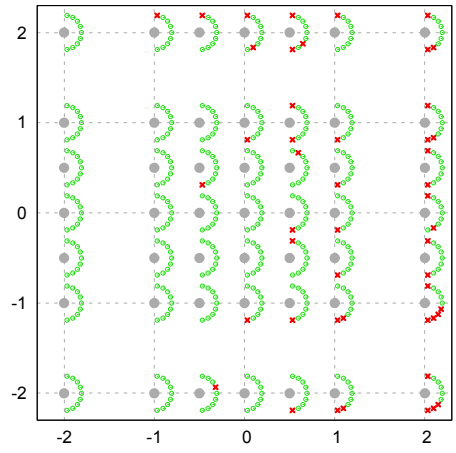


(c) NDT with trilinear interpolation

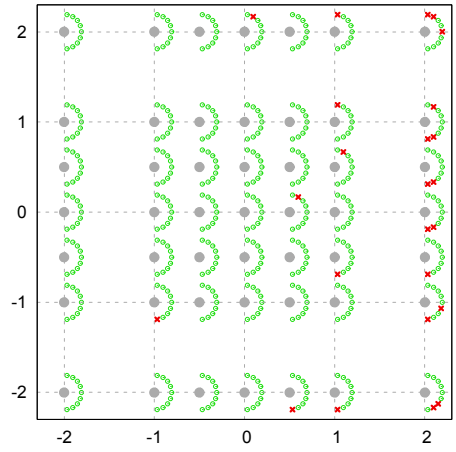
Fig. 7. Results from data set A. Success rate for loose/strict translation threshold: 30%/13% for ICP, 77%/37% for NDT, 95%/95% for trilinear NDT.



(a) ICP



(b) NDT



(c) NDT with trilinear interpolation

Fig. 8. Registration results from data set A, judging by rotation error only (disregarding the translation threshold). Success rate: 95% for ICP, 89% for NDT, 95% for trilinear NDT.

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. It should be noted that judging registration accuracy in this way can only give a rather uncertain indication of registration quality due to the fact that errors

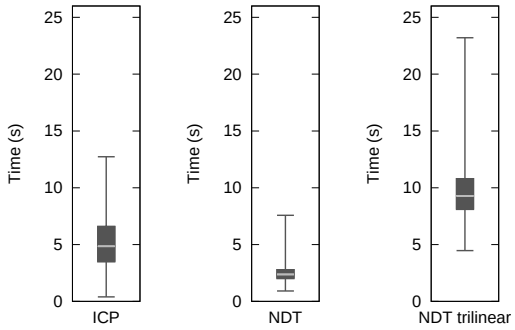


Fig. 9. Execution times for data set A. The light bar shows the median execution time from the 441 runs, the "whiskers" extend to the extreme values, and the edges of the box show the first and third quartile.

at different points in the loop can cancel out so that the final error is small even if there are inaccuracies for some scan pairs.

For NDT, the accumulated translation error was 2.26 m (using the altered initial pose for scan 23). The translation error vector was $[1.12, -0.02, -1.97]$, which means that the accumulated vertical error was almost 2 m.

For NDT with trilinear interpolation, the accumulated error was slightly larger in this case: 3.99 m (using the original pose estimates). The translation error vector was $[3.22, -0.56, -2.30]$. Most of the the horizontal translation error was because the more feature-less tunnel segments were somewhat "shortened".

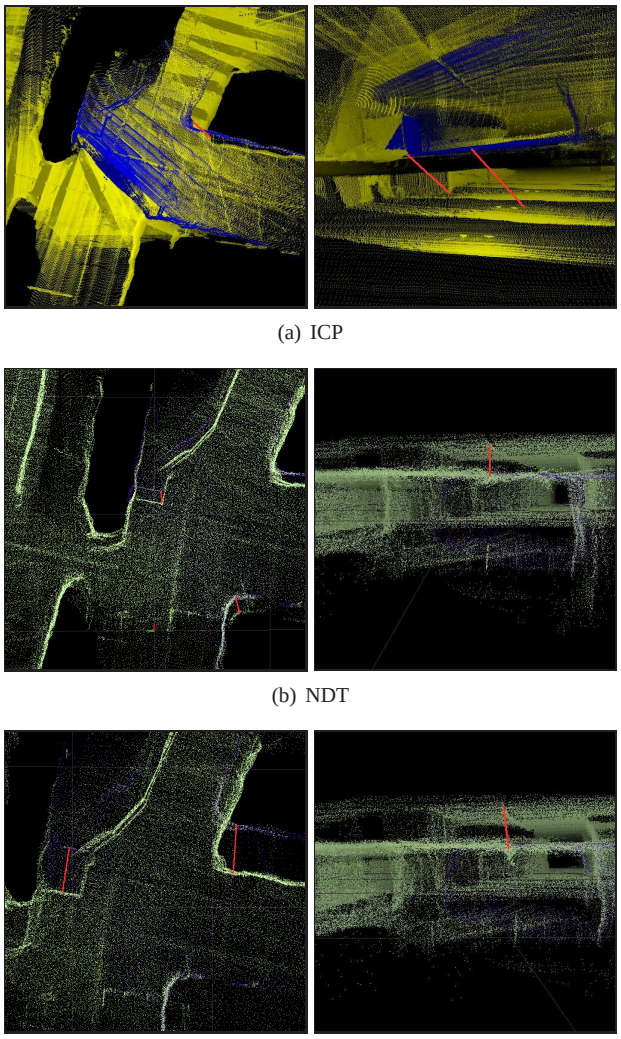
For ICP an accumulated translation error of 2.97 m can be reported (using the altered pose estimate for scan 33).

Close-ups of the registration results are shown in Fig. 10.

V. Conclusions

We have evaluated the performance of ICP and NDT for 3D mapping, and presented an improved version of NDT.

In our 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 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 substantially increased the success rate of NDT, at the expense of longer execution times.



(a) ICP

(b) NDT

(c) NDT with trilinear interpolation

Fig. 10. The accumulated error after registering all of the scans in data set B. 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.

References

[1] F. Amigoni, S. Gasparini, and M. Gini. Good experimental methodologies for robotics mapping: A proposal. In *Proc. ICRA*, 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 *Proc. IROS*, volume 3, pages 2743–2748, 2003.

[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] Daniel F. Huber and Nicolas Vandapel. Automatic 3D underground mine mapping. 2003.

[10] 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.

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

[12] 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.

[13] Osnabrück Robotic 3D Scan Repository. `http://kos.informatik.uni-osnabrueck.de/3Dscans/`.

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

[15] S. Thrun, D. Hähnel, D. Ferguson, M. Montemerlo, R. Triebel, W. Burgard, C. Baker, Z. Omohundro, S. Thayer, and W. Whittaker. A system for volumetric robotic mapping of underground mines. In *Proc. ICRA*, 2003.