# Appearance-Based Loop Detection from 3D Laser Data Using the Normal Distributions Transform

Martin Magnusson, Henrik Andreasson, Andreas Nüchter, and Achim J. Lilienthal

***Abstract*— We propose a new approach to appearance based loop detection from metric 3D maps, exploiting the NDT surface representation. Locations are described with feature histograms based on surface orientation and smoothness, and loop closure can be detected by matching feature histograms. We also present a quantitative performance evaluation using two real-world data sets, showing that the proposed method works well in different environments.**

## I. Introduction

Being able to detect loop closure is essential for autonomous mobile robot navigation, especially with respect to the problem of simultaneous localisation and mapping (SLAM). There are algorithms that can distribute the accumulated pose error of pairwise registered scans in order to render a consistent map once the robot has detected that it has closed a loop. Some examples include the tree based relaxation methods of Frese et al. [1], [2] and the 3D relaxation method of Grisetti et al. [3]. However, detecting loop closure when faced with large pose errors remains an open problem. We propose a loop detection approach using only on 3D point cloud data, based on surface shape and orientation histograms. The histograms can be compared using a difference metric, and 3D scans with similar histograms are assumed to be from nearby locations.

## II. Technical approach

Our method is inspired by NDT: the normal distributions transform. NDT is a method for representing a scan surface as a piecewise continuous function. It has previously been used for efficient pairwise 2D and 3D scan registration [4], [5]. However, the NDT surface representation can also be used as a compact description of the appearance of a 3D scan, as will be explained in this section.

### A. The normal distributions transform

The input to NDT is a 3D point cloud. The points are represented by the 3D coordinates of their position in space (and we use the location of the laser scanner as the origin of each scan's local coordinate system). The point cloud is transformed into a collection of smooth functions in the following fashion. The space occupied by the scan is subdivided into a regular grid of cells (squares in the 2D case, cubes in the 3D case). Each cell stores the mean vector and covariance matrix of the scan points within the cell; in other words, the parameters of a normally distributed probability density function (PDF) describing the local surface shape. The covariance matrix can encode either a round, linear (stretched ellipsoid) or planar (squashed ellipsoid) shape. Our appearance descriptor is created from histograms of these local surface shape descriptions.

In order to minimise the issues with spatial discretisation, the cells are overlapping, so that if the side length of each cell is $q$, the distance between each cell's centre point is $q/2$.

### B. Appearance descriptor

We classify the NDT cells based on the shapes of their PDFs. For each cell, the eigenvalues $\lambda_1 > \lambda_2 > \lambda_3$ and corresponding eigenvectors $\vec{e}_1, \vec{e}_2, \vec{e}_3$ of the covariance matrix are computed. There are three main cell classes. Distributions are assigned to a class based on the relations between their eigenvalues with respect to a threshold $t_e \in (0, 1)$ that quantises a "much smaller" relation.

- Distributions are linear if $\lambda_2/\lambda_1 < t_e$.
- Distributions are planar if they are non-linear and $\lambda_3/\lambda_2 < t_e$.
- Distributions are spherical if they are non-linear and non-planar (in other words, if no eigenvalue is $1/t_e$ times larger than any other one).

It would be straightforward to use more classes such as different levels of "almost planar" distributions by using more eigenvalue ratio thresholds, but for the data presented here using more than one $t_e$ did not improve the result.

Each of the main classes can be divided into sub-classes, based on orientation for the planar and linear classes, and surface roughness for the spherical class. Using $s$ spherical sub-classes, $p$ planar sub-classes, and $l$ linear sub-classes, the basic element of the proposed appearance descriptor is the feature vector

$$\vec{f} = \big( \underbrace{f_1, \ldots, f_s}_{\text{spherical classes}}, \underbrace{f_{s+1}, \ldots, f_{s+p}}_{\text{planar classes}}, \underbrace{f_{s+p+1}, \ldots, f_{s+p+l}}_{\text{linear classes}} \big), \quad (1)$$

where $f_i$ is the number of cells that belong to class $i$.

For planar distributions, the eigenvector $\vec{e}_3$ (which corresponds to the smallest eigenvalue) coincides with the normal vector of the plane that is approximated by the PDF. Assume that there is a set of $p$ approximately evenly distributed lines $\mathcal{P} = \{P_1, \ldots, P_p\}$. For example, using an equal area partitioning [6] to distribute $p$ points on a half-sphere, $\mathcal{P}$ is the set of lines intersecting the origin and one of the points. The index for planar sub-classes is

$$i = s + \arg\min_j \delta(\vec{e}_3, P_j), \quad (2)$$

M. Magnusson, H. Andreasson, and A. J. Lilienthal are with AASS, Örebro University, Sweden {martin.magnusson, henrik.andreasson}@oru.se, achim@lilienthals.de

A. Nüchter is with Jacobs University Bremen, Campus Ring 1, Campus Ring 1, 28759 Bremen, Germany andreas@nuechti.de

where $\delta(\vec{e}, P)$ is the distance between a point $\vec{e}$ and a line $P$. In other words, we choose the index of the line $P_j$ that is closest to $\vec{e}_3$.

The same method can be used for linear distributions, but using $\vec{e}_1$ (which corresponds to the linear axis) instead of $\vec{e}_3$.

Spherical sub-classes can be defined by the ratio $\lambda_2/\lambda_1$, although for the data used here, one spherical class sufficed.

The distance from the scanner location to a particular surface is also important information. For this reason, each location is described by a matrix

$$\mathbf{F} = \left(\vec{f}_1{}^{\mathrm{T}} \cdots \vec{f}_r{}^{\mathrm{T}}\right)^{\mathrm{T}} \tag{3}$$

and a corresponding set of range intervals $\mathcal{R} = \{r_1, \ldots, r_r\}$, where each $\vec{f}_i$ is the histogram of all NDT cells within the range defined by interval $r_i$, measured from the origin.

### C. Rotation invariance

Because the appearance descriptor (3) explicitly uses the orientation of surfaces, it is not rotation invariant. In order for the appearance descriptor to be invariant to rotation, the orientation of the scan must first be normalised.

Starting from an initial histogram $\vec{f}'$ with $\mathcal{R} = \{[0, \infty)\}$, we want to find two peaks in plane orientations and orient the scan so that the most common plane normal is aligned along the $z$ axis, and the second most common plane normal is aligned in the $yz$ plane. The reason for orientations of planes instead of lines is that planar cells are much more common than linear ones. For environments with mostly linear structures, line orientations could be used instead.

There is not always an unambigous maximum, so we generate two *sets* of directions: $\mathcal{Z}$ and $\mathcal{Y}$. Given the planar part $\vec{p} = (p_1, \ldots, p_p)$ of $\vec{f}'$ and an ambiguity ratio treshold $t_a \in [0, 1]$ that determines which histogram peaks are "similar enough", $\mathcal{Z}$ and $\mathcal{Y}$ are generated as follows:

$$i' = \operatorname*{argmax}_i p_i, \tag{4}$$

$$\mathcal{Z} = \{i \in \{1, \ldots, p\} | p_i \geq t_a p_{i'}\}, \tag{5}$$

$$i'' = \operatorname*{argmax}_i p_i | i \notin \mathcal{Z}, \tag{6}$$

$$\mathcal{Y} = \{i \in \{1, \ldots, p\} | i \notin Z, p_i \geq t_a p_{i''}\}. \tag{7}$$

For each $i \in \mathcal{Z}$, we create a rotation $\mathbf{R}_z$ that encodes a rotation of $-\arccos(\vec{P}_i \cdot (0, 0, 1))$ radians around the axis $\vec{P}_i \times (0, 0, 1)$, where $\vec{P}_i$ is a unit vector along the line $P_i$. For each $i \in \mathcal{Y}$, the corresponding rotation $\mathbf{R}_y$ is $-\arccos((\mathbf{R}_z \vec{P}_i) \cdot (0, 1, 0))$ radians around the axis $(0, 0, 1)$. The descriptor $\mathbf{F}$ is created for the rotated scan $\mathbf{R}_y \mathbf{R}_z \mathcal{S}$.

This aligment is always possible to do, unless all planes have the same orientation. If it is not possible to find two main directions it is sufficient to use only $\mathbf{R}_z$, because in this case no subsequent rotation around the $z$ axis change which histogram bins are updated for any planar PDF. If linear sub-classes are used, it is possible to derive $\mathbf{R}_y$ from linear directions if not enough planar directions can be found.

It is possible to choose one of two rotations (in opposite directions) when aligning the scan. However, since the appearance histograms are based on sets of lines $\mathcal{P}$ and $\mathcal{L}$ with ambiguous orientations, as opposed to rays, it does not matter which of the two rotations is used.

In the case of ambiguous peaks (that is, when $\mathcal{Z}$ or $\mathcal{Y}$ has more than one member), we generate multiple histograms. For each combination $\{i, j | i \in \mathcal{Z}, j \in \mathcal{Z} \cup \mathcal{Y}, i \neq j\}$ we apply the rotation $\mathbf{R}_y \mathbf{R}_z$ to the original scan and generate a histogram. The outcome is a set of histograms $\mathcal{F} = \{\mathbf{F}_1, \ldots, \mathbf{F}_{|\mathcal{Z}||\mathcal{Z} \cup \mathcal{Y}| - |\mathcal{Z}|}\}$. For highly symmetrical scans, this could lead to a large number of histograms. For a scan generated at the centre of a sphere, where the histogram bins for all directions have the same value, $p^2 - p$ histograms would be created. In practice, this has not been a problem.

### D. Difference metric

To quantify the difference between two appearance descriptors $\mathbf{F}$ and $\mathbf{G}$ we normalise $\mathbf{F}$ and $\mathbf{G}$ with their entrywise 1-norms, compute the sum of Euclidean distances between each of their rows (that is, each range interval), and weight the sum by the ratio of the number of occupied NDT cells in the scans:

$$\sigma(\mathbf{F}, \mathbf{G}) = \sum_{i=1}^{r} \left( \left|\left| \frac{\vec{f}_i}{||\mathbf{F}||_1} - \frac{\vec{g}_i}{||\mathbf{G}||_1} \right|\right|_2 \right) \frac{\max(||\mathbf{F}||_1, ||\mathbf{G}||_1)}{\min(||\mathbf{F}||_1, ||\mathbf{G}||_1)}. \tag{8}$$

The factor $\max(||\mathbf{F}||_1, ||\mathbf{G}||_1) / \min(||\mathbf{F}||_1, ||\mathbf{G}||_1)$ is used to differentiate between large scans (with many NDT cells) and scans of more confined spaces (with few cells).

Given a scan pair $(\mathcal{S}_1, \mathcal{S}_2)$ with appearance descriptor sets $(\mathcal{F}, \mathcal{G})$, all members are compared to each other using (8), and the minimum $\sigma$ is used as the difference measure.

$$\sigma'(\mathcal{F}, \mathcal{G}) = \min_{i,j} \sigma(\mathbf{F}_i, \mathbf{G}_j) \qquad \mathbf{F}_i \in \mathcal{F}, \mathbf{G}_j \in \mathcal{G} \tag{9}$$

### E. Parameters

The parameters of the proposed appearance descriptor are

- class counts: $s$, $p$, and $l$,
- eigenvalue ratio threshold $t_e$,
- range limits $\mathcal{R}$,
- ambiguity ratio threshold $t_a$,
- NDT cell size $q$.

We have chosen the values of these parameters empirically. Some parameters depend on the scale of the environment, but we found that a single parameter set worked well for all our data. If using a scanner with different resolution or different max range, $\mathcal{R}$ and $q$ should probably be adjusted.

We found that using one spherical class, nine planar classes, and one linear class worked well for our different data sets. The reason for using only one spherical and linear class is that these classes tend to be less stable than planar ones. Linear distributions with unpredictable directions tend to occur at the far ends of a scan, where the point density is too small. Spherical distributions often occur at corners and edges, depending on where the boundaries of the NDT cells end up, and may shift from scan to scan. However, using only the planar features ($s = l = 0$) decreased the obtainable recall rate without false positives or mismatches with around one third for our data.

The eigenvalue ratio threshold $t_e$ and ambiguity ratio threshold $t_a$ were also chosen empirically. In our experiments, using $t_e = 0.10$ and $t_a = 0.60$ produced good results.

The best cell size $q$ depends mostly on the scanner configuration. If the cell size is too small, planes at the further parts of scans (where the scan points are sparse) may show up in the histogram as lines with unpredictable orientation. Previous work [5] has shown that cell sizes between 0.5 m and 2 m work well for registering scans of the scale encountered by mobile robots. We have used $q = 0.5$ m and $\mathcal{R} = \{[0, 3), [3, 6), [6, 9), [9, 15), [15, \infty)\}$.

Two more parameters determine the outcome when examining the similarity matrix for detection of loop closure:

- minimum loop size $S$,
- difference threshold $t_d$.

If $S$ is too small, a number of correct but uninteresting "loops", consisting only of consective scans, may be detected. We are only interested in detecting proper loops that contain more than some minimum number of scans. The minimum loop size $S$ should therefore be set to the minimum number of scans that can be expected to be recorded between two visits to any location. Each scan $\mathcal{S}_i$ is compared to all other scans, except for the closest ones $\{\mathcal{S}_{i-S}, \ldots, \mathcal{S}_{i+S}\}$. We set $S$ to 30 when testing the algorithm.

It is important to find a good value for the difference threshold $t_d$, which determines which pairs are considered overlapping (positives). An "overlapping scan" in this context is a scan that is taken in a region that overlaps with another visited region. Setting $t_d$ too small decreases the number of true positives. Setting it too large increases the number of false positives. Fig. 1 shows how the numbers of true positives, false positives, and mismatches change with various difference thresholds. Mismatches are overlapping scans that are matched to the wrong scan.

A method for determining $t_d$ that has been useful for our experiments is to perform expectation maximisation (EM) to fit a mixture of three Gaussian curves to the smallest difference values of all scans of a data set and choosing the point where the first and second curves intersect (see Fig. 2). The reasoning for using three kernels is that we assume that difference value comes from one of three distributions: one with overlapping scans, one with non-overlapping ones, and one with random values where the proposed method fails to give a meaningful difference measure.

Finally, one more parameter was used when evaluating the classification result with respect to the ground truth data:

- distance threshold $t_r$.

This parameter is only added for convenience in order to save the labour of manually judging which scans are from overlapping regions. Instead, all scans that have at least one other scan outside the minimum loop size $S$ such that the ground truth distance between the two scans is less than the distance threshold $t_r$ are considered overlapping.

## III. Experiments

In order to evaluate the performance of the proposed algorithm, we used two data sets: one outdoor set from a campus area, and one from an underground mine.

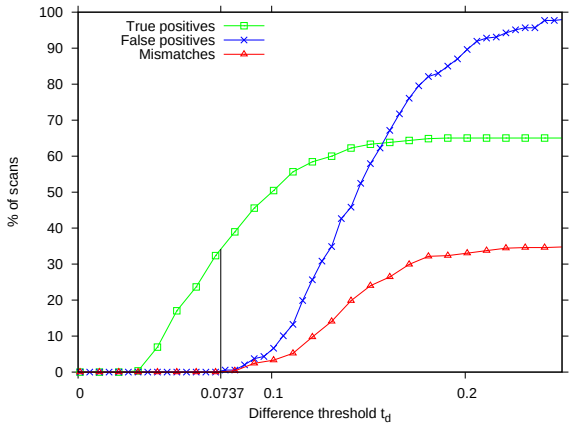



Fig. 1. Relationship between difference threshold and success rate for the Hannover2 data set. The threshold giving the maximum number of true positives with no errors is marked with a bar.

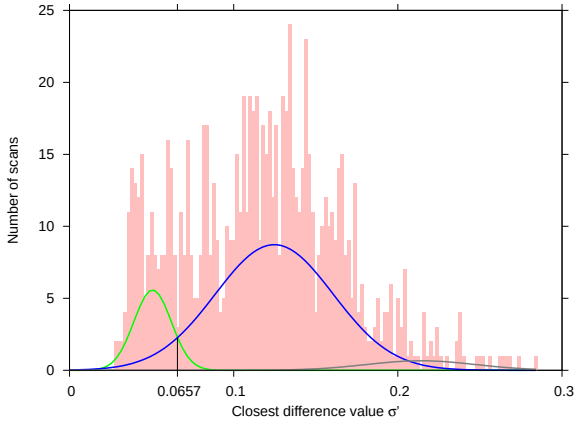



Fig. 2. Determining $t_d$ for the Hannover2 data set using EM with three Gaussian kernels. A histogram showing the difference values of all scans' most similar neighbour is printed in the background. Three kernels fitted to the histogram are overlayed. In this case, $t_d = 0.0657$ would be used.

To quantify the performance of the loop detection algorithm, we counted the number of correctly detected overlapping scans (true positives), scans incorrectly regarded as overlapping (false positives) and mismatched scans (those that were correctly regarded as overlapping, but whose corresponding most similar scan was incorrect). Please refer to Table I to see how scans were labelled.

### A. Data sets

The Hannover2 (Fig. 3(a)) data set was recorded at the university campus of Leibniz Universität Hannover, Germany. It contains 922 3D omni-scans (with 360° field of view), covering a trajectory of about 1.24 km. Each scan contains approximately 15 000 points.

The Kvarntorp data set (shown in Fig. 4(a)) was recorded in the Kvarntorp mine outside Örebro, Sweden. The data set is divided into four "missions". For the experiments presented in this paper, we used "mission 4" followed by "mission 1". This combined mission sequence has 131 3D scans, each

TABLE I

TAXONOMY FOR EVALUATING RESULTS. GIVEN A SCAN $\mathcal{S}$, $\hat{\mathcal{S}}$ IS THE SCAN NEAREST TO $\mathcal{S}$ (EXCEPT THOSE WITHIN THE MINIMUM LOOP SIZE $S$), AND $\bar{\mathcal{S}}$ IS THE MOST SIMILAR SCAN TO $\mathcal{S}$.

| $\mathcal{S}$ is | if $\sigma'(\mathcal{S}, \bar{\mathcal{S}})$ | and distance to $\hat{\mathcal{S}}$ | and distance to $\bar{\mathcal{S}}$ |
|---|---|---|---|
| true positive | $< t_d$ | $< t_r$ | $< t_r$ |
| mismatch | $< t_d$ | $< t_r$ | $\geq t_r$ |
| false positive | $< t_d$ | $\geq t_r$ | any |
| true negative | $\geq t_d$ | $\geq t_r$ | any |
| false negative | $\geq t_d$ | $< t_r$ | any |

covering a 180° field of view and containing around 70 000 data points. The total trajectory is about 370 m.

The Kvarntorp data set is rather challenging for a number of reasons. Firstly, the mine environment is highly self-similar. Without knowledge of the robot's trajectory, it is very difficult to tell different tunnels apart. The fact that the scans of this data set are not omnidirectional also makes it more difficult, because the same location looks quite different depending on which direction the scanner is pointing towards. The median distance travelled between consecutive scans was also longer for this data set: around 2.5 m, compared to 1.5 m for Hannover2.

All of the scan data are available for download [7]. The ground truth poses are available from the authors on request.

### B. Results

The results are summarised in Table II.

*1) Outdoor data:* For the Hannover2 data set, ground truth pose measurements were acquired by registering every 3D scan against a point cloud made from a given 2D map and an aerial lidar scan made while flying over the campus area. Fig. 3(c) shows the similarity matrix for our algorithm and Fig. 3(b) shows the ground truth distance matrix. For this data set, we used $t_r = 10$ m.

For the parameter values stated in Section II-E, the difference threshold $t_d = 0.0737$ gives the maximum number of true positives without any false positives: a recall rate of 35.3%. These results are comparable to visual place recognition methods using SIFT features from camera images [8]. At this point it should be noted that a recall rate of 30% is often sufficient to close all loops as long as the number of false positives and mismatches is low, because several scans are usually taken from each location.

Using $t_d = 0.0657$ instead, as determined by expectation maximisation (Fig. 2), the result is 29.2% true positives (and no errors). The parameters of the Gaussian mixture model were initalized by running a maximum of 50 EM iterations from randomly initialized start parameters and selecting the parameters providing the best likelihood among those trials.

Using minimum loop size $S = 0$ (which entails that 100% of the scans are overlapping) and $t_d = 0.0737$, the result is 45.9% true positives and 0.5% mismatches.

The most difficult part of the Hannover2 data set is when the stretch H–I is revisited (scans 480–513 and 815–847). Only a few of those scans were detected. However, there is some distance between the first and second run through that area (as may be seen from the lighter shade in the circled area of Fig. 3(b)), so those scans only barely overlap.

TABLE II

SUMMARY OF CLASSIFICATION RESULTS FOR MANUALLY SELECTED $t_d$. IN ALL CASES, THERE WERE NO MISMATCHES.

| Data set | pos. | neg. | $t_d$ | **true pos.** | false pos. |
|---|---|---|---|---|---|
| Hannover2 | 575 | 347 | 0.0737 | **35.3%** | 0% |
| Kvarntorp | 35 | 95 | 0.0894 | **31.4%** | 1.1% |

The stretch E–F (scans 251–350) is revisited while travelling in the opposite direction (scans 612–715). These are the longest sequences of scans that are taken in different directions, and should give a good indication of the algorithm's robustness under viewpoint changes. The recall rate when examining only E–F and F–E is 45.1% using $t_d = 0.0737$, and the maximum recall rate with no errors is attained using $t_d = 0.0821$ which gives a recall rate of 53.4%.

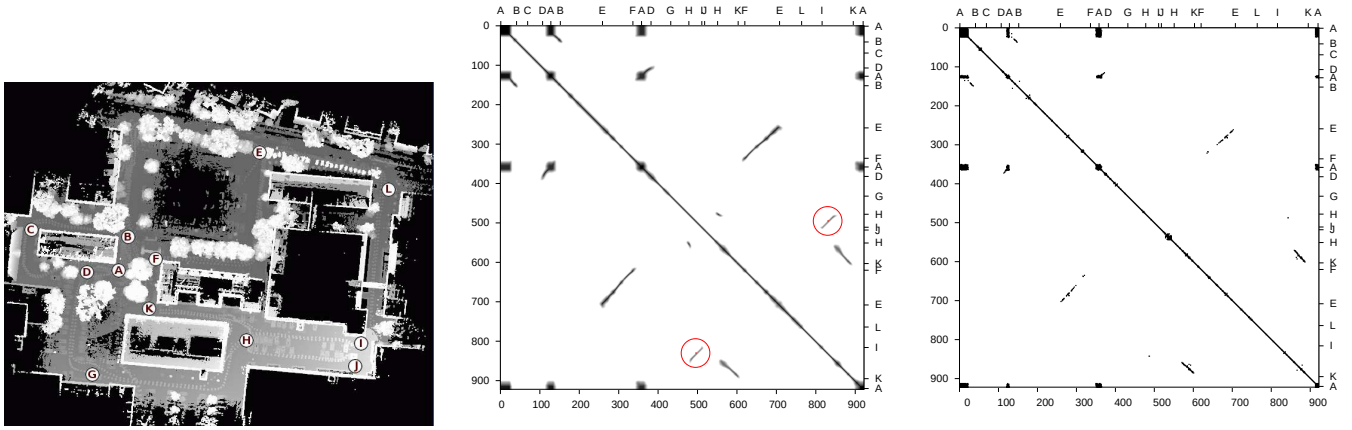
*2) Underground mine data:* Ground truth poses for the Kvarntorp data set were provided using the algorithm presented in [9]. It is a network based global relaxation method for 3D laser scans. To generate a genuine truth, the network was manually given to the algorithm and the result was visually inspected for correctness.

The loop detection algorithm described in this paper cannot be rotation invariant if the input scans are not omnidirectional. When looking in opposite directions from the same place, the view is generally very different. Because an omnidirectional scanner was not used to record Kvarntorp, only scans taken in similar directions were counted as overlapping when evaluating the algorithm for this data set. The distance matrix shown in Fig. 4(b) only shows scan pairs that were taken with a maxiumum orientation difference of 20°. We also chose $t_r = 5$ m instead of 10 m. The reason for selecting a smaller distance threshold is firstly because of the scanner's limited field of view and secondly because of the more confined spaces of the mine environment. These two factors make the appearance of scenes change more drastically than in the open-air scans of Hannover2.

We used the same parameters for this data set as for Hannover2, except for $t_d = 0.0894$. The recall rate with this $t_d$ was 31.4% and there was one false positive. The ground truth distance matrix is shown in Fig. 4(b), and the similarity matrix of our algorithm is shown in Fig. 4(c).

### C. Execution time

The experiments were run using a C++ implementation on a laptop computer with a 1600 MHz Intel Celeron CPU and 2 GiB of RAM. For the Hannover2 data set, average times for computing the surface shape histograms were 0.18 s per call to the histogram computation function, and in total 0.94 s per scan to generate histograms (this includes transforming the scan, generating $\vec{f}'$ and the histograms that make up $\mathcal{F}$). The average size of $\mathcal{F}$ is 3.2 histograms. In total, 61.2 s were spent to compute similarity measures for scan pairs. There

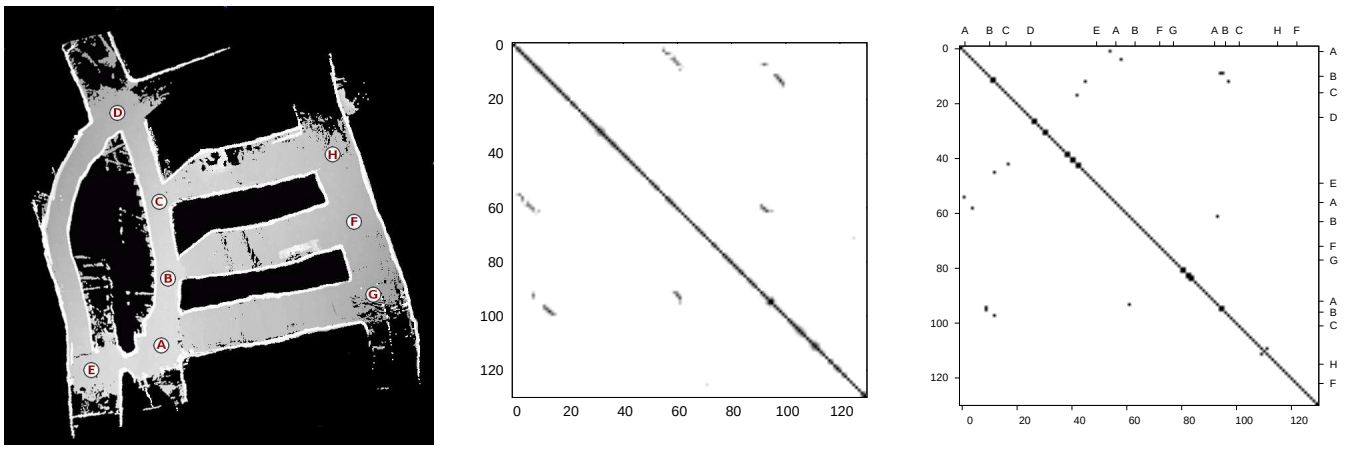


(a) Overview of the Hannover2 data set, seen from above with parallel projection. The robot traveled along the sequence A–B–C–D–A–B–E–F–A–D–G–H–I–J–H–K–F–E–L–I–K–A.

(b) Ground truth distance matrix, showing all scan pairs taken less than 10 m apart. The circled area marks where H–I is revisited.

(c) Similarity matrix, showing all scan pairs whose difference value $\sigma' < 0.0737$. Because of the large matrix and the small print size, this image has been morphologically dilated by a $3 \times 3$ element in order to better show the values.

Fig. 3. The Hannover2 data set.



(a) Overview of the Kvarntorp data set. The robot travelled along the sequence A–B–C–D–E–A–F–G–A–B–C–H–F–H.

(b) Ground truth distance matrix, showing all scan pairs taken less than 5 m apart and with an orientation difference of less than 20°.

(c) Similarity matrix, showing all scan pairs with a difference value $\sigma' < 0.0894$.

Fig. 4. The Kvarntorp data set.

are 922 scans in the data set, 2 947 histograms were created, $2\,947^2$ similarity measures were computed, so the average time per similarity comparison was around 7 $\mu$s. So if each scan requires the generation of 3.2 histograms on average, a new scan can be compared to roughly 13 800 other scans in one second to test for loop closure, disregarding the time needed to compute the histograms.

The requirements for both data sets are summarised in Table III, showing the number of scans in each data set and the average point count per scan, as well as the average time to create a single histogram and the average number of histograms generated per scan. The time for creating the histograms and the number of histograms required for rotation invariance depend on the sizes of the point clouds, but the time required for similarity comparisons depend only on the number of histogram bins.

TABLE III
SUMMARY OF RESOURCE REQUIREMENTS.

| Data set | Scans | Avg. $\lvert\mathcal{S}\rvert$ | Avg. creation time | Avg. $\lvert\mathcal{F}\rvert$ |
|---|---|---|---|---|
| Hannover2 | 922 | 15k | 0.18 s | 3.2 |
| Kvarntorp | 130 | 70k | 0.27 s | 2.8 |

## IV. Related work

Previous work on loop detection has focused mostly on data from camera images and 2D range data.

Cummins and Newman [10] have presented a bag-of-words based method using “visual words” from camera images. Scenes are represented as a collection of such words (local visual features) drawn from a “dictionary” of available features. The appearance descriptor is a binary

vector indicating the presence or absence of all words in the dictionary. The appearance descriptor is used within a probabilistic framework together with a generative model that describes how informative each visual word is and common co-occurrences of words. Cummins and Newman have reported recall rates of 35% to 46% on urban outdoor data sets. However, it should be noted that it is difficult to compare recall rates from different data sets. It is not possible to say how their method would perform in the environments used in our experiments, and vice versa.

A method that is more similar to the approach presented here is the 2D histogram matching of Bosse et al. [11], [12]. While our method may also be referred to as histogram matching, there are several differences. For example, Bosse et al. create 2D histograms with one dimension for the spatial distance to the scan points and one for scan orientations. The angular histogram bins cover all possible rotations of a scan in order to achieve rotation invariance. With the parameters used in their papers, 240 000 histogram bins are required for the 2D case. For unconstrained 3D motion with angular bins for the $x$, $y$, and $z$ axes, a similar discretisation would lead to many millions of bins. In contrast, the 3D histograms presented here require only a few dozens of bins. The histogram matching of Bosse et al. is reported to work well for the kidnapped robot problem, but they have not provided a quantitative performance evaluation yet.

Daniel Huber has described a method based on spin-images [13] for matching multiple 3D scans without initial pose estimates [14]. Such global registration is closely related to the loop detection problem. An important difference between spin-images and the surface shape histograms proposed in this paper is that spin-images are local feature descriptors, describing the surface shape around one point. In contrast, our surface shape histograms are global appearance descriptors, describing the appearance of a whole 3D point cloud. Comparing spin-images to the local PDF features used in this work, spin-images are more descriptive and invariant to rotation. Normal distributions are unimodal functions, while spin-images can capture arbitrary surface shapes if the resolution is high enough. The initial step of Huber's multi-view surface matching method is to compute a model graph by using pairwise global registration with spin-images for all scan pairs. The model graph contains potential matches between pairs of scans, some of which may be incorrect. Surface consistency constraints on sequences of matches are used to distinguish correct matches from incorrect ones because it is not possible to distinguish the correct and incorrect matches at the pairwise level. The algorithm proposed in this paper can be seen as another way of generating the initial model graph and evaluating a local quality measure. Using data sets scanned from small objects with the background removed, the recall rate of Huber's method is up to 80%. (Again, it should be noted that it is difficult to compare recall rates from very different data sets.) However, the execution time is much longer. Using data sets containing 32 scans with 1000 mesh faces each, as done by Huber, the time to compute the initial model graph using spin-image matching can be estimated to $1.5 \cdot 32^2 = 1\,536$ s (the complete time is not explicitly stated in [14], but pairwise spin-image matching is reported to require 1.5 s on average). With a data set of that size, a rough estimate of the execution time of the algorithm proposed in this paper is $32 \cdot 0.8 + (32 \cdot 3)^2 \cdot 7 \cdot 10^{-6} = 26$ s on similar hardware, based on the execution times in Table III.

## V. Conclusions and future work

We have described a novel approach to appearance-based place recognition using 3D range data. We have demonstrated its performance on two real-world data sets. We can conclude that the results are encouraging, and the performance is comparable to that of loop detection methods using visual data, with a recall rate of over 30% even for quite challenging underground mine data.

The purpose of this paper is to demonstrate the performance of the NDT-based appearance descriptor. To further improve performance, future work should include learning a generative model in order to learn how to disregard common, nondiscriminative, features, based on the general appearance of the current surroundings (see [10]).

It would also be interesting to do a more elaborate analysis of the similarity matrix than simple thresholding in order to better discriminate between overlapping and non-overlapping scans. A more detailed study of useful models for finding the difference treshold $t_d$ would also be interesting.

Further future work should include investigating how this approach performs when faced with dynamic changes, such as moving furniture or people.

## References


[1] U. Frese, P. Larsson, and T. Duckett, "A multilevel relaxation algorithm for simultaneous localisation and mapping," *IEEE Trans. Robotics*, vol. 21, no. 2, pp. 196–207, Apr. 2005.

[2] U. Frese and L. Schröder, "Closing a million-landmarsk loop," in *Proc. IROS*, 2006, pp. 5032–5039.

[3] G. Grisetti, S. Grzonka, C. Stachniss, P. Pfaff, and W. Burgard, "Efficient estimation of accurate maximum likelihood maps in 3D," in *Proc. IROS*, 2007.

[4] P. Biber and W. Straßer, "The normal distributions transform: A new approach to laser scan matching," in *Proc. IROS*, vol. 3, 2003, pp. 2743–2748.

[5] M. Magnusson, A. J. Lilienthal, and T. Duckett, "Scan registration for autonomous mining vehicles using 3D-NDT," *J. Field Robotics*, vol. 24, no. 10, pp. 803–827, 2007.

[6] E. B. Saff and A. B. J. Kuijlaars, "Distributing many points on a sphere," *The Mathematical Intelligencer*, vol. 19, no. 1, pp. 5–11, 1997.

[7] Oct. 9 2008. [Online]. Available: http://kos.informatik.uni-osnabrueck.de/3Dscans/

[8] M. Cummins and P. Newman, "Accelerated appearance-only SLAM," in *Proc. ICRA*, 2008.

[9] D. Borrmann, J. Elseberg, K. Lingemann, A. Nüchter, and J. Hertzberg, "Globally consistent 3D mapping with scan matching," *JRAS*, vol. 56, no. 2, pp. 130–142, Feb. 2008.

[10] M. Cummins and P. Newman, "Probabilistic appearance based navigation and loop closing," in *Proc. ICRA*, Rome, Apr. 2007.

[11] M. Bosse and J. Roberts, "Histogram matching and global initialization for laser-only SLAM in large unstructured environments," in *Proc. ICRA*, Apr. 2007.

[12] M. Bosse and R. Zlot, "Keypoint design and evaluation for global localization in 2D lidar maps," in *RSS*, Jun. 2008.

[13] A. E. Johnson, "Spin images: A representation for 3-D surface matching," Ph.D. dissertation, Carnegie Mellon University, 1997.

[14] D. F. Huber, "Automatic three-dimensional modeling from reality," Ph.D. dissertation, Carnegie Mellon University, 2002.