# Survey on 2D Lidar Feature Extraction for Underground Mine Usage

Kristin Nielsen and Gustaf Hendeby


Nielsen, K., Hendeby, G., (2022), Survey on 2D Lidar Feature Extraction for Underground Mine Usage, *IEEE Transactions on Automation Science and Engineering*.
https://doi.org/10.1109/TASE.2022.3172522



# Survey on 2D Lidar Feature Extraction for Underground Mine Usage

Kristin Nielsen*† and Gustaf Hendeby†
*Rocktec Automation; Epiroc Rock Drills AB; Örebro, Sweden
kristin.nielsen@epiroc.com
†Dept. Electrical Engineering; Linköping University; Linköping, Sweden
gustaf.hendeby@liu.se

***Abstract*—Robust and highly accurate position estimation in underground mines is investigated by considering a vehicle equipped with 2D laser scanners. A survey of available methods to process data from such sensors is performed with focus on feature extraction methods. Pros and cons of the usage of different methods for the positioning application with 2D laser data are highlighted, and suitable methods are identified. Three state-of-the-art feature extraction methods are adapted to the scenario of positioning in a predefined map and the methods are evaluated through experiments conducted in a simulated underground mine environment. Results indicate that feature extraction methods perform in parity with the method of matching each ray individually to the map, and better than the point cloud scan matching method of a pure ICP, assuming a highly accurate map is available. Furthermore, experiments show that feature extraction methods more robustly handle imperfections or regions of errors in the map by automatically disregarding these regions.**

Note to Practitioners: Robust positioning in GNSS denied environments is a complex and real problem experienced by practitioners in many fields. The focus of this paper is underground mining; however, the findings and discussions have bearing on many other applications where reliable GNSS is unavailable and lidar is used for positioning in unstructured environments. An important take home message, of interest in any application with dynamically changing environments, is that by using feature extraction map errors are automatically handled as features are simply not matched in erroneous regions. The experiments in this paper are performed with a realistic laser simulation model trained on real data; hence the gap between simulation and reality is relatively small ensuring the results are relevant for practical purposes. In general, the feature extraction methods are sensitive to the parameter settings, and would have to be properly tuned for the specific application. Furthermore, the computational complexity, which is only mentioned briefly in this paper, varies a lot between the methods and has to be investigated further to ensure meeting real-time requirements. Since the feature extraction methods robustly handle errors in the map, future research will explore how this can be used to enable automatic updates of the map.

***Index Terms*—underground positioning, 2D lidar, feature extraction, position estimation, scan matching, data association**

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

## I. Introduction

Estimation of the position of a mobile platform is an important part of many autonomous systems, which is specifically targeted as a challenge for the underground mining industry [1]. In an underground mine there is no coverage from *global navigation satellite systems* (GNSS) and the dynamic and harsh environment demands as little infrastructure as possible. A self contained localization system is therefore preferred [2].

Existing commercial autonomous mine vehicles are equipped with 2D lidars, odometry and *inertial measurement units* (IMUs) [2–5], and fuses the information from the different data sources by filtering algorithms to determine the position of the vehicle in a map. Matching all laser rays from a lidar scan individually to a predefined map is computationally expensive, and methods for selecting sub-sets of available rays without losing estimation precision are developed in [6]. However, these methods are sensitive to outliers in the data and imperfections in the map. Alternatively, direct scan matching algorithms that seek a transformation which aligns two point clouds can be used. For this the *iterative closest point* (ICP) algorithm is the de-facto standard and many variations have been developed [7].

The computer vision community has successfully built an arsenal of feature extraction methods for exactly the purpose of getting robust data associations in noisy data with low computational effort [8]. Inspired by this, feature extraction methods specifically adapted to 2D laser data have been developed in the last decades [9–11], which are also proved to perform well in clean indoor environments.

The feature extraction procedure consists of first detecting interest points, also known as *keypoints*, in the sensor data. Secondly, computing a distinctive signature for each of them, called a *descriptor*. A keypoint is ideally viewpoint invariant, repeatable, and descriptive over the region of detection. The descriptor is usually a vector encoding the neighborhood of the keypoint enabling more robust matching among points acquired from different viewpoints, with different sensor noise, which might also be affected by occlusion. Incorporating the feature extraction procedure into the positioning problem gives the workflow of a measurement update, as depicted in Fig. 1, starting with the input sensor data, resulting in a position estimate of the moving object.

This paper investigates how available feature extraction

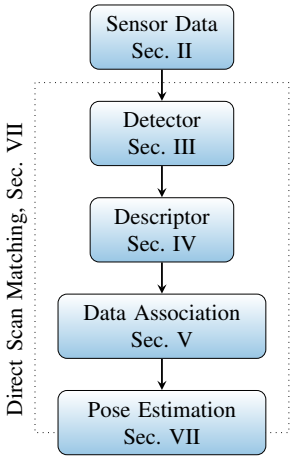



Fig. 1. The work-flow when a feature extraction method is used for positioning. A *detector* searches the input sensor data for keypoints, and a *descriptor* computes a distinctive signature for each of them. The descriptors are used in the data association step to enable more robust correspondence matching when data is acquired from different viewpoints.

methods can improve the robustness of an underground position estimate, in comparison to process each lidar ray individually, or to direct scan matching algorithms. The contributions are the following:

- A survey of available feature extraction methods suitable for 2D laser data. The survey is focused on the detection of keypoints, which is identified as the most crucial step for state estimation performance.
- Three selected feature extraction methods are adapted to fit the problem of positioning in a predefined map. Also evaluation metrics are adapted to this scenario to enable fair comparison of feature detectors.
- Estimation experiments conducted in a simulated mine environment that show that feature extraction methods perform on a par with the method of matching each ray individually and better than ICP assuming a good map. However, feature extraction methods are much more robust to errors in the map.

The structure of this paper follows the work-flow of feature based position estimation, see Fig. 1. In Sec. II, the type of sensor data is presented. Available feature detectors and descriptors are presented in Sec. III and IV, respectively. How correspondences between identified keypoints can be found is discussed in Sec. V, followed by a systematic evaluation of (for this application) suitable feature detectors and descriptors in Sec. VI. Sec. VII presents how position estimates can be obtained utilizing feature extraction methods in the context of a predefined map and Sec. VIII contains an experimental evaluation of the complete work-flow. Finally, concluding remarks are given in Sec. IX.

## II. Sensor Data

The sensor data considered in this paper is from 2D lidars, which can easily be converted to point clouds. Often when a point cloud is considered it is assumed to be compact, dense, and, in many cases, 3D point cloud [14], see Fig. 2 for an example. A point cloud produced by a 2D lidar contains much less information and is very sparse in comparison to a 3D point cloud. The point cloud is, at least approximatively, describing a 2D plane, and depending on the environment the data is more or less scattered.

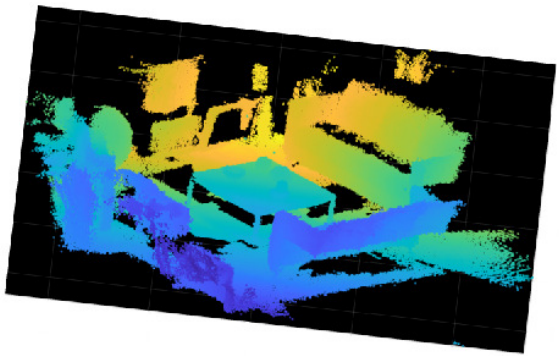


Fig. 2. 3D point cloud produced by a RGB-D (Kinect-style) camera, showing a scene containing furnitures. [12]

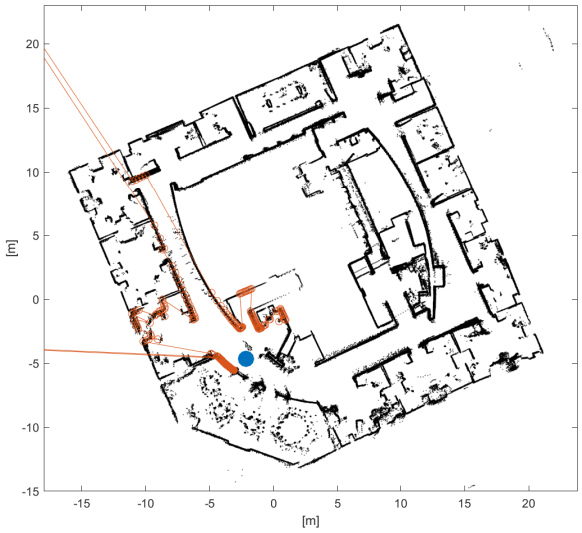
2D range datasets suitable for feature extraction purposes are presented in [9] and later used in [10, 11]. The datasets are either from indoor environments with well-defined (often $90°$) corners connecting structured walls with smooth surfaces, as in Fig. 3(a) which is a snap-shot from the *Intel*-dataset. Or, the datasets are from open outdoor environments without continuous walls resulting in scattered points with lots of no-return measurements in between, as in the *Victoria park*-dataset, shown in Fig. 3(b).

A point cloud produced by a 2D lidar in an underground mine is provided in Fig. 3(c), with a map of the operation area overlaid. The corners are not as well-defined as in the indoor dataset and the surfaces of the walls are not smooth, yet they are continuous. This distinguishes the underground mine environment from the typical indoor- as well as outdoor-scenario. As seen in the figures, the characteristics of measurement outliers and no-return data, follows the indoor datasets, while the scale of the operation area (a couple of $100\,\mathrm{m}$) have more in common with the outdoor environment.

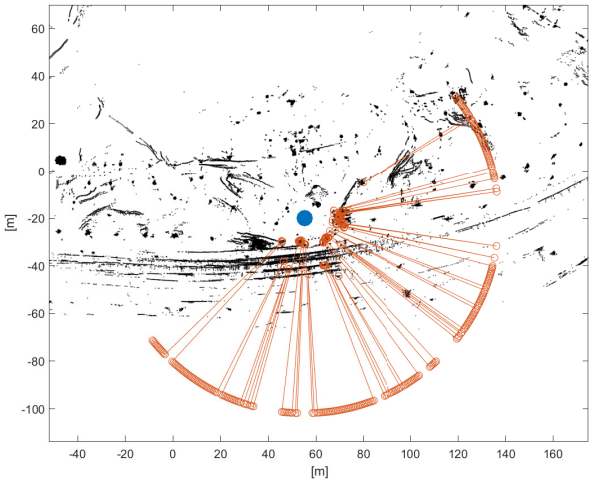
## III. Detectors

Detectors and descriptors are typically presented in pairs as complete feature extraction methods. However, it is in general possible to use any detector in combination with any arbitrarily chosen descriptor, since a detector merely finds points of interest in the data [8]. Therefore, the detectors and descriptors are here presented separately. It should though be noted that it might be more computationally efficient to use a specific detector-descriptor pair since partial results from computations might be reused.

A substantial amount of research effort is put into object detection and recognition in images. For 3D point clouds, several keypoint feature detectors have been proposed that mainly originates, or at least are strongly influenced by, methods from image processing, see for example [15–18]. Since this type of data is not the focus of this paper we will merely refer to some of the many surveys and performance evaluations available on this topic [19–21]. From now on, only 2D point clouds are considered.




(a) The *Intel* dataset is freely available in the Radish repositiry and the github repository accompanying [9] provides the ground truth data used to obtain this picture.




(b) The *Victoria park* dataset, available through [13] and the ground truth data used to generate this picture is provided in the github repository accompanying [9].


(c) Laser data from Epiroc's underground test area in Kvarntorp outside of Örebro in Sweden, with a map of the operation area overlaid.

Fig. 3. Examples of point clouds acquired by 2D lidars in different environments. Blue circles mark sensor positions.

### A. Detectors Inspired by Image Features

The computer vision community has done major progress in detecting general keypoints in an image. Through rasterization, 2D laser data can be converted into an image, and feature detectors designed for images can be applied. Detailed descriptions of different algorithms for image feature detection can be found in [8]. Some worth mentioned by name are:

- The *Harris corner* detector [22], which defines a corner as the crossing of two different edge directions, where an edge is a sudden change in image intensity. A structure tensor is computed consisting of the covariance of directional derivatives in the local neighborhood of a pixel. From the eigenvalues of this tensor it can be identified if the pixel is uninteresting, on an edge, or a corner point.
- The *Shi-Tomasi* detector [23], optimizes the Harris corner detection. By only considering the minimum eigenvalue the computational complexity is considerably reduced.
- In the *scale-invariant feature transform* (SIFT) [24], a raw image is convolved by Gaussian kernels at different scales, $t$, forming blurred or smoothed images. Differences of successive smoothed images at different scales are formed, and keypoints are taken as extreme points of the difference images.

In [25, 26], 2D laser data is converted to an image and Shi-Tomasi corner detection is applied. This approach enables the re-use of computer vision algorithms, however it is computationally expensive to perform rasterization of laser measurements and the rasterization introduces inaccuracies.

### B. 2D Range Data Detectors

Data acquired from a 2D range finder consist of relatively few points, the point density is highly non-uniform and not view-point invariant. The nature of the data affects how to extract stable and distinguishable keypoints, and there are few publications specialized on 2D laser data keypoint features.

One of the first attempts to detect and describe keypoints in 2D range data is [27], where keypoints are selected at locations of high curvature in submaps defined as a collection of multiple laser scans. In the subsequent work [28] different types of keypoint detectors are tested, but all of them search for keypoints in a submap rather than in a single scan. This makes their approach more of a submap characterization technique for place recognition rather than a feature extraction method.

Although, general feature detectors specialized on 2D range data are few, there exists three quite recently developed such methods:

*1) FLIRT:* In [9] the *fast laser interest region transform* (FLIRT) is presented as a detector for locally defined keypoints for 2D laser data. The FLIRT method adopts the scale space theory used in SIFT. First, the original data is represented by a family of smoothed signals parametrized by $t$, the size of a Gaussian smoothing kernel. Then, differences of the smoothed signals are formed and peaks are identified to locate keypoints. Three different detectors are presented in [9]. The first one operates on the raw range data and the second one on a local approximation of the normal direction in each point. The third, and also the best performing one, according to simulations

performed in the paper, adapts the work done for 3D point clouds in [18] to 2D range data. The range data defines a curve in Cartesian 2D space and the scale space theory is applied to this curve. An integral operator maps the input curve into a multi-scale parametrization

$$S(\alpha(s);t) = \int_{\Gamma} k(s,u;t)\alpha(u)\,du, \tag{1}$$

where $\Gamma$ is the curve, $\alpha(s)$ is the parametrization of the curve by the geodesic coordinate $s$, and $k(s,u;t)$ is a Gaussian kernel with mean $s-u$ and standard deviation $t$. The integral is approximated by a sum and to compute the Gaussian kernels in the sum the geodesic distance between each observed point pair is required. As suggested by [18], *disjoint minimum spanning trees* (DMST) are used for this purpose in the FLIRT algorithm. This introduces a design parameter that defines how many disjoint trees to use. When DMST is used for the curvature approximations, the range data is considered an unordered point cloud, and extreme outliers such as no-return data are automatically ignored, see [29] for details.

Keypoints are then detected as local maxima of the exponential damping expression

$$F(p_i,t) = \frac{2\|p_i - S(\alpha(s),t)\|}{t} e^{-\frac{2\|p_i - S(\alpha(s),t)\|}{t}}, \tag{2}$$

where $p_i$ represents each scan point. To find peaks of $F(x,t)$ two design parameters are defined, the minimum value of $F(x,t)$ to be considered a peak ($F_{\min}$) and the minimum separation between peaks ($F_{\text{dist}}$).

The scales to consider for the curve fitting are determined by

$$t = t_0(t_i)^k, \tag{3}$$

where $t_0$ and $t_i$ are scalar design parameters, $k \in 0,1,\dots n_k-1$ is the current scale, and $n_k$ gives the number of scales.

*2) FALKO:* In [10] a novel keypoint detector specialized on 2D range data is proposed, the *fast adaptive laser keypoint orientation-invariant* (FALKO) detector. This method focuses on finding stable view-point invariant keypoints like corners, rather than gaps and isolated points which could be a result of occlusion. The detector is designed to be orientation invariant and measurement sparsity independent by considering neighboring data points within a variable radius of the candidate point. A set of neighboring points of an observed point $p_i$, is defined as

$$C(p_i) = \left\{p_j \in \overline{S} : \|p_j - p_i\| < r_i\right\}, \tag{4}$$

where the radius $r_i$ is computed as

$$r_i = ae^{b\|p_i\|} \tag{5}$$

with $\|p_i\|$ defines as the distance from the observed point to the sensor origin. The design parameters $a$ and $b$ scale the neighborhood radius, $a$ has most effect on short range measurements whereas $b$ has a greater impact on longer ranges. The two endpoints in each neighbor set $p_{j_{\min}}$ and $p_{j_{\max}}$ do, together with the point $p_i$ itself, form a triangle. As a first rough approximation of a corner, this triangle is used to discard many points as keypoints candidates. If the base or the height of the triangle is less than $\frac{r_i}{\beta}$ the point is discarded. Greater values of $\beta$ allows both wider and sharper corners as candidate points.

For the remaining candidate points, the set $C(p_i)$ is divided into two subsets

$$C_L(p_i) = \left\{p_j \in C(p_i) : j < i\right\}, \tag{6a}$$

$$C_R(p_i) = \left\{p_j \in C(p_i) : j > i\right\}. \tag{6b}$$

For each point in any of the sets, a quantized orientation $\phi_j$, with respect to the candidate point is computed by forming a grid with $s_n$ circular sectors centered on the candidate point. The orientation is then defined as in which circular sector the point $p_j$ lays in, and a discrete distance function $d_{\theta_j,\theta_k}$ is defined as how many sectors are between $\theta_j$ and $\theta_k$. A *cornerness score* is then computed for each of the two subsets in (6) as the sum of all point pair distances in the subset

$$\text{score}_L(p_i) = \sum_{h=i-1}^{j_{\min}} \sum_{k=h-1}^{j_{\min}} |d_\theta(\phi_h,\phi_k)| \tag{7a}$$

$$\text{score}_R(p_i) = \sum_{h=i+1}^{j_{\max}} \sum_{k=h+1}^{j_{\max}} |d_\theta(\phi_h,\phi_k)|. \tag{7b}$$

Such a score is small if the points in the subset are aligned with each other. A total score for a candidate point is then defined as the sum of $\text{score}_L(p_i)$ and $\text{score}_R(p_i)$ and keypoints are chosen as local minima of the sum. To avoid ambiguous keypoints a *non-maxima suppression* (NMS) procedure is applied defining a minimum distance between keypoints, $\text{NMS}_{\min}$.

*3) BID:* The *B-spline based interest point detector* (BID) is introduced in [11]. It uses the same concept as FALKO by extracting high curvature candidate points using a set of neighbor points defined as in (4) with a radius growing with measured range as in (5). The points in $C(p_i)$ are used to approximate a B-spline curve that in turn is used to determine corner existence. First a B-spline curve is constructed with four control points, evenly distributed along the axis corresponding to the largest eigenvalue of the covariance matrix of $C(p_i)$. Then, the iterative *point distance minimization* (PDM) algorithm is applied to the initial B-spline to fit it to the data points in $C(p_i)$. This resulting B-spline is denoted $S(t)$. The normalized Euclidian distance between the two median control points $P_1$ and $P_2$, defined by the construction of the B-spline, of the approximated curve $S(t)$,

$$E = \frac{\|P_1 - P_2\|}{\sum_{i=0}^{2}\|P_i - P_{i+1}\|}, \tag{8}$$

is used as a measure of curvature. Keypoints are defined as local maxima of the inverse of the normalized Euclidean distance $E^{-1}$ with scores above a threshold $T_{\text{th}}$.

An advantage with the BID method is that the approximated curve can be interpolated along the B-spline to get a more precise location of the keypoint, that also depends less on the point density. A disadvantage is the computational effort it takes to approximate the B-spline curve. This computational burden is somewhat lowered by a threshold when PDM is applied. If the error, according to (8), of the initial curve, is lower than a threshold $E_{\text{th}}$, the PDM is not applied and

the candidate point is disregarded. Such situation occurs in low curvature regions. As for the FALKO detector a NMS procedure is suggested to avoid ambiguous keypoints.

*4) Computation Time Comparison:* The FALKO method includes only simple computations and is therefore well suited for real-time applications. This is experimentally verified in [11], where computation times for FLIRT and FALKO are compared to their own contribution BID. FALKO is roughly 50 times faster than FLIRT and 100 times faster than BID [11].

### C. Geometric Detectors

The extraction of various high-level geometric features have been used in association with 2D laser scanners, *e.g.*, line segments [30–33], poly-lines [34], curve segments [35], circles [36, 37], Bezier curves [38] or B-splines [39]. Strictly, these are not feature keypoints as they are not view-point invariant points of interest with an associated descriptor. Due to the use of dedicated geometric features, these approaches do not provide robust solutions in a general environment of any shape, since they are design and tailored to a specific environment or application, *e.g.*, indoor or outdoor environments. Because of this, these methods are not expected to perform well with the special characterization of data from an underground mine.

## IV. Descriptors

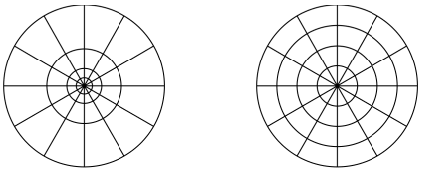
A good descriptor should capture the signature of a keypoint regardless of view-point, sensor noise, and occlusions. When used for localization or place recognition, the detector is usually defined as a fixed-length vector or a histogram, accompanied by a distance metric. This is to make the descriptor computationally efficient when the intention is to robustify the data association step. Although, for image processing there exists more complex descriptors. Two examples are the basis space descriptors, where the descriptor is defined in another basis, and the polygon shape descriptors, where the features are based on the perimeter of a polygon shape [8].

Since the same feature can be perceived from different view-angles, descriptors can also include additional information about the orientation of the keypoint. Thus, for matching, the descriptors can be rotated and aligned before the distance metric is evaluated.

As for detectors there are substantial work performed in the computer vision community [8, 40, 41], and in [26] the SIFT descriptor is used for 2D range data. However, the results in [10, 11] indicate that stable detection of feature points is much more important than a sophisticated description to obtain good feature matching in 2D laser data. Therefore the main focus of the reminder of this section is on the descriptors accompanying the FLIRT, FALKO, and BID detectors in [9], [10], and [11], respectively. All originating from the same concept of a shape context.

### A. Shape Context

In the early range data feature extraction attempt in [28], multiple detectors are discussed and evaluated. They consider a region with a fixed radius of 9 m and include data from multiple scans when computing description vectors. This is suitable for their place recognition purpose, but it cannot really be viewed as a local keypoint feature descriptor. Nevertheless, one of the best performing descriptors in the experiments conducted in [28] is based on the concept of shape contexts, which is also the case for the descriptors used in state-of-the-art range data feature extraction work [9–11]. The concept of shape contexts is developed for object recognition in images in [42]. Objects are treated as a set of points sampled from the internal or external contours of the object. Then a reference point is chosen and the distribution of the relative coordinates of the remaining points is represented in a histogram. The bins of the histogram are uniform in log-polar space with $r_n$ bins in the radial direction and $\alpha_n$ bins in the angular direction, see Fig. 4(a). This histogram is defined as the *shape context* descriptor. This type of descriptor suits the range data well since the data is already points and detected keypoints are natural choices of reference points.



(a) Log-polar histogram (b) Linear-polar histogram

Fig. 4. Histograms used by a shape context descriptor, both with $r_n = 4$ and $\alpha_n = 12$.

### B. $\beta$-grid

The $\beta$-grid descriptor is introduced in [9]. A linear-polar histogram (see Fig. 4(b)) is suggested to reduce the impact of measurement noise that typically occurs in radial direction in lidar data. This noise would be captured in the bins near the centre in a log-polar histogram. The radius of the histogram $\rho_{\max}$ as well as an inner threshold $\rho_{\min}$ for when data points are ignored and assumed ambiguous, are design parameters for the $\beta$-grid descriptor, together with $r_n$ and $\alpha_n$ defined in Sec. IV-A above. The symmetric $\chi^2$-distance function defined as

$$d = \frac{1}{2} \sum_i \frac{(x_i - y_i)^2}{(x_i + y_i)}, \tag{9}$$

where $x_i$ and $y_i$ are elements in different histograms, is used to compare descriptors. The $\beta$-grid descriptor also encodes the probabilities of free-space into the histogram by applying Bayesian parameter learning. This gives a distinction between a concave and a convex structure, but it only gives slightly improved performance in experiments, and it could be questioned if the increased computational complexity is worth it.

### C. Binary Shape Context (BSC)

The *binary shape context* (BSC) descriptor accompanying the FALKO detector is introduced in [10]. This descriptor uses the same structure with the linear-polar histogram as the $\beta$-grid descriptor, to reduce the risk of capturing measurement

noise in the bins close to the centre. However, BSC limits the histogram to be binary. With bins that are either filled or empty, the influence of the point density is reduced. A natural choice of $\rho_{\max}$ is the region for where the set of neighboring points are defined, resulting in different histogram sizes for different keypoints, $\rho_{\max} = r_i$. The distance metric associated with this descriptor is a Hamming-like function and the orientation of each keypoint can be computed using a variant of the *intensity centroid* presented in [43].

### D. Spline Distribution Histogram (SDH)

In [11] the similar *spline distribution histogram* (SDH) descriptor is proposed, again using the linear-polar histogram to minimize measurement noise impact. The SDH descriptor utilizes results from the accompanying B-spline based detector (BID). Since B-spline approximations of the curve in the vicinity of a keypoint is performed in the detection step, the curve can be interpolated to remove the point density dependency. The histogram can be kept non-binary not to lose information about the alignment of neighboring points. The interpolation is done with equal distance $d_t$ between points along the spline. The symmetric $\chi^2$-distance is used to measure the distance between two SDH descriptors. The orientation of keypoints are computed using the centroids as in [10], but now applied to uniformly sampled points along the approximated spline instead of neighboring points from the point cloud.

## V. Data Association

The purpose of data association is twofold; it should match keypoints detected in different point clouds corresponding to the same actual feature, and it should filter out and ignore outliers not having a true match in any of the point clouds.

### A. Feature Matching

Even though feature descriptors are designed to provide robust feature matching between scans, pure descriptor association usually gives poor results in experiments [9, 10, 44]. Therefore some other data association technique is often used in conjunction with the feature descriptors. For position estimation purposes, where a good initial estimate is known, the descriptors can be used as a gating rule for geometric associations. For place recognition and loop closure applications other data association algorithms are often used, *e.g.*, *geometrical landmark relations* (GLARE) [10, 11, 45], where 2D laser scans are transformed into pose invariant histogram representations, or, as in [11, 46], a graph theoretic approach. The relative geometry between points in a scan is used to build graphs and then the *maximum common subgraph* is used to match different scans. In [47] a comparison of different data association techniques based on extracted FALKO features is performed. The *Hungarian algorithm*, a combinatorial optimization method for solving assignment problems, is suggested as a data association method well suited for shape context descriptors [8, 10], but no experiments provide any insight of the performance when applied to 2D laser data.

### B. Outliers and Occlusions

Data association in noisy data containing outliers or occlusions can be a serious problem for robust state estimation. When individual matching of laser rays are used, an outlier or occlusion gives false information as input to the state estimation method. In scan matching algorithms acting on raw data, outliers or occlusions can cause inaccurate transformations or possibly incorrect matching. *Random sample consensus* (RANSAC) [48] is an iterative method designed to robustly handle outliers in data. This method is general and applies to various kinds of data. It is a common strategy to increase robustness towards outliers in scan matching algorithms, and in [9] it is used for loop closure detection on keypoint features.

Using a feature extraction approach, sporadic outliers can be filtered out by a sophisticated detector/descriptor (as *e.g.*, the DMST in FLIRT). In the case of more regular or continuous outliers within a scan, *e.g.*, as the result of occlusion, the effect is keypoints that cannot be matched and therefore ignored for further processing. If many outliers or occlusions are present this can of course result in severe information loss, but a state estimation method will in neither case be fed with false information. In addition to this already built in robustness against outliers, RANSAC can of course also be applied to detected keypoints to further improve the outlier rejection capabilities.

## VI. Detector and Descriptor Evaluation

In the literature the stability is emphasized as the most important property of a feature, *i.e.*, the detection of features in the same location after changing viewpoint, and regardless of laser sensor properties such as noise level and resolution. In [9] a set of metrics are assessed, which in turn follows the approach in image processing [44, 49]. This set of metrics is also used later in [11, 47]. Using these criteria as a starting point, this section presents quality metrics for feature extraction methods adapted to the situation with a predefined map.

### A. Evaluation Data

2D laser data for feature evaluation is obtained in a simulated environment. Fig. 5 depicts a map of Epiroc's underground test area outside of Örebro in Sweden, an area having many attributes of a typical underground mine. A path is defined so that all parts of the operation area are in the field-of-view of the lidar at some point along the trajectory. For 2504 poses along the trajectory, lidar scans are simulated in two different ways. First, a virtual range finder is placed in the environment resulting in noise-free lidar scans simulating a perfect scenario with no uncertainties neither in the map nor the measurements. Secondly, measurement errors are added utilizing a *support vector machine* (SVM) model, trained on real data collected by a vehicle located in the physical area the map is covering, as described in [50]. In this model the bias and variance of the error depends on range and inclination angle to the wall, resulting in each simulated laser measurement being one realization of a stochastic distribution. The model also includes a probability of producing no-return measurements, resulting

TABLE I
PARAMETER VALUES IN THE SVM MODELS USED TO SIMULATE RANGE MEASUREMENTS.

| | $k_s$ | $\varepsilon$ | $C$ |
|---|---|---|---|
| $\overline{y}$ | 1.92 | 0.026 | 45.6 |
| $\Sigma_y$ | 0.145 | 0.025 | 0.248 |
| $P_{\text{no}}$ | 2.71 | 0.0047 | 4.33 |

in a very realistic simulation of laser data, representative for this particular environment.

In detail, three separate SVM model are trained. One for the mean value of a range measurement $\overline{y}$, one for the variance $\Sigma_y$, and, one for the probability of no-return $P_{\text{no}}$. The models use Gaussian kernel functions where the kernel scale $k_s$, residual margin $\varepsilon$, and, a box constraint constant $C$ controlling the penalty imposed on observations that lie outside the residual margin, are all hyperparameters determined by a five-fold cross-validation optimization. Values for the three models are presented in Table I.

### B. Adaptations to Map Positioning

In [9–11] the evaluation metrics are applied to datasets consisting of 2D lidar scans. For the underground application where a predefined map is available, every scan can instead be matched to interest points extracted in the map. The procedure is therefore adapted, resulting in:

1) Find keypoints in the map. The map consists of polygons whose edges represent walls. To extract keypoints, the map is first converted to a single point cloud by equidistant interpolation along the polygon edges. Then, a feature detector is applied to the point cloud to identify interest points.
2) Define a path in the map that will make as many keypoints as possible visible to the vehicle at some point along the path. It is possible to use real data collected along the path, but here simulated laser scans are obtained as described in Sec. VI-A.
3) Extract keypoints in each of the laser scans using the feature detector.
4) Associate a keypoint in a scan to a keypoint in the map if the Euclidean distance is smaller than a threshold, $d_{\max}$.

Note that the association is performed without using the descriptors, hence this procedure only measures how well different detectors can detect features in this particular environment. If the descriptors are to be evaluated the last step has to be adjusted.

### C. Detector Quality Measures

This section presents quality metrics for the detectors only.

*1) Global Detector Stability Test:* In [26] a procedure associate extracted features from different scans is presented. Ideally a keypoint should always be detected when in the field of view of the sensor. In practice this is not the case since no detector can perfectly find keypoints from different view points, occlusions or in noisy data. However, some keypoints are more often found than others by a specific detector,

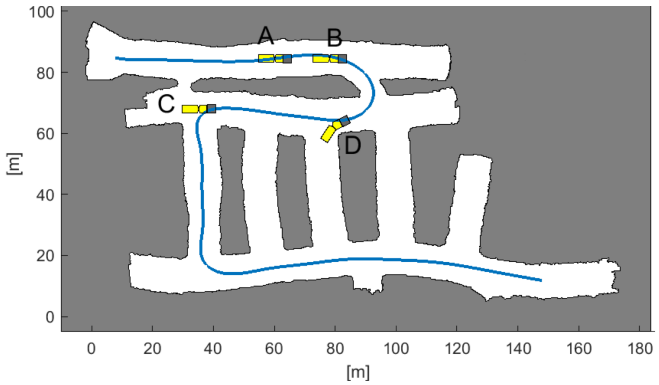



Fig. 5. The blue line is the path sampled with poses used for global stability test of feature detectors. The four vehicle poses marks the positions where the measurement update experiments are performed.

and these are referred to as stable keypoints. The covariance matrix of all keypoints in the different laser scans associated to the same feature in the map provides a measure of the stability of the feature point. The maximum eigenvalue $\lambda_{\max}$ of the covariance matrix defines a scalar metric. By taking the average over all scans in a path, this gives a measure of how sensitive a detector is to data sampling.

Another stability measure is the percentage of single points. That is the ratio of map keypoints never matched to a keypoint in any scan and gives an indication of how many unnecessary keypoints a detector finds.

How many scans along a complete path that have zero matching features is of special importance for the localization application, since in a dynamic filter approach this leads to no new information in the measurement update. Therefore, this is also included here as an additional stability measure.

*2) Detector Repeatability:* The repeatability of a point detected from two different scans is a frequently used metric for the stability of a feature point. Consider the sets of keypoints $\mathcal{K}_R$ and $\mathcal{K}_S$ which are detected from a reference scan and from a similar scan, respectively. Assuming correct associations, the repeatability is defined as the ratio of common keypoints over the smaller of the two keypoint sets [9], *i.e.*,

$$\frac{|\mathcal{K}_R \cap \mathcal{K}_S|}{\min\{|\mathcal{K}_R|, |\mathcal{K}_S|\}}, \tag{10}$$

where $|\mathcal{K}|$ denotes the cardinality of $\mathcal{K}$.

When matching laser data to a predefined map repeatability is defined as the ratio between the number of associated interest points in a scan and the total number of feature points in the map visible from a particular pose of the vehicle. Again the association is purely geometric without utilizing any descriptor, to keep this as a metric for the detectors only.

*3) Detector Evaluation:* The methods FALKO and BID are evaluated both using the parameter setup suggested in the corresponding papers [10, 11], and with parameters manually chosen. Good parameter values are determined according to the noise levels in the input data, to the quality and smoothness of the walls in the environment/map, and to the average range of the measurement in this environment. The goal is to have as few scans without matching features in the map as possible,

TABLE II
PARAMETER CONFIGURATIONS USED FOR THE DETECTORS.

| Detector | $t_0$ | $t_i$ | $n_k$ | $F_{\min}$ | $F_{\text{dist}}$ |
|---|---|---|---|---|---|
| FLIRT [9] | 0.2 | 1.4 | 5 | 0.34 | 0.001 |
| | $a$ | $b$ | $\beta$ | $s_n$ | $\text{NMS}_{\min}$ |
| FALKO [10] | 0.2 | 0.07 | 4.0 | 16 | 0.2 |
| FALKO Tuned | 0.2 | 0.1 | 2.25 | 16 | 0.2 |
| | $a$ | $b$ | $T_{\text{th}}$ | $E_{\text{th}}$ | $\text{NMS}_{\min}$ |
| BID [11] | 0.2 | 0.07 | 3.1 | 0.03 | 0.2 |
| BID Tuned | 0.4 | 0.1 | 0 | 0.03 | 0.2 |

but without including insignificant keypoints that are never matched. With the environment used for all simulations in this paper (see Fig. 5) this amounts to increasing $b$, which results in larger neighboring areas for longer measurements. The value in the original paper is designed to fit indoor range measurements of $\sim 10\,\text{m}$, but in the underground application valid range measurements can be up to $\sim 70\,\text{m}$, see Fig. 3. For BID also $a$ is increased to allow for larger neighborhoods also for shorter ranges, and $T_{\text{th}}$ is lowered, since BID otherwise finds relatively few keypoints. On the contrary, the $\beta$ parameter for FALKO is adjusted to only detect the most significant keypoints.

There is always a trade-off between detecting many keypoints and only detecting the most significant ones. From the aspect of repeatability, a theoretically perfect detector would consider every point in the point cloud a keypoint, as this maximizes the repeatability score in (10). However, this is not desirable in practice since the keypoints will not be distinctive and thus difficult to match. The detectors are therefore tuned to detect as few keypoints as possible without leaving large areas of the map without keypoints. To give a fair comparison of the three methods equal effort has been put into the non-trivial task of hand tuning the parameter values for each of the methods. However, the sensitivity in parameter values differs for specific parameters and the different methods, resulting in varying changes of parameter values. The suggested parameter setup in the corresponding papers is worst suited for the BID method for this particular application. Therefore, this is the method where the parameter values are changed the most. See Table II for a complete list of parameter values.

All detectors have been implemented by the author in `Matlab`. For the FLIRT method a C++ implementation is provided together with the paper [9]. Unfortunately this implementation cannot be used with the subsampled map as input, since it assumes laser scans containing range information. The range information is not used for the detector but only later to compute the $\beta$-grid descriptor. Therefore a `Matlab` implementation of a FLIRT detector taking a point cloud as input is used here, with all parameters set according to [9].

Global stability is evaluated on data produced according to Sec. VI-A and results are presented in Table III based on $d_{\max} = 0.1\,\text{m}$. It is in general hard to fine-tune the methods. The tuned version of FALKO performs worse than the standard version on all measures, but still gives a slightly better position estimate as will be presented later in Sec. VIII. The repeatability scores are in absolute values significantly lower than the results in [9–11]. That is because they are comparing scans, while the keypoints are extracted from a map in a slightly different way. The FLIRT method looks promising but is sensitive to noise. The repeatability score is halved when using the realistic laser simulation. On the contrary, BID is the worst performing method, but it is also the least sensitive to noise.

TABLE III
GLOBAL STABILITY OF IDENTIFIED KEYPOINTS WITH DIFFERENT FEATURE DETECTION METHODS.

| Detector | Keypoints in Map | $\bar{\lambda}_{\max}$ | Single Points | Scans without match | Repeat-ability |
|---|---|---|---|---|---|
| FLIRT | 892 | 0.0006 | 32.7% | 0.8% | 0.080 |
| FALKO Std | 1434 | 0.0019 | 48.1% | 0.0% | 0.084 |
| FALKO Tuned | 827 | 0.0016 | 55.5% | 1.0% | 0.070 |
| BID Std | 363 | 0.0020 | 60.9% | 19.1% | 0.066 |
| BID Tuned | 557 | 0.0020 | 60.0% | 11.2% | 0.067 |

(a) Noise-free laser simulation

| Detector | Keypoints in Map | $\bar{\lambda}_{\max}$ | Single Points | Scans without match | Repeat-ability |
|---|---|---|---|---|---|
| FLIRT | 892 | 0.0024 | 35.9% | 8.7% | 0.042 |
| FALKO Std | 1434 | 0.0024 | 47.6% | 0.2% | 0.057 |
| FALKO Tuned | 827 | 0.0022 | 53.5% | 5.0% | 0.048 |
| BID Std | 363 | 0.0022 | 59.8% | 28.6% | 0.049 |
| BID Tuned | 557 | 0.0024 | 59.8% | 17.4% | 0.053 |

(b) SVM laser simulation

### *D. Descriptor Quality Measures*

This section presents quality measures for descriptors.

*1) 1-Precision-Recall:* To evaluate the performance of feature descriptors a 1-precision-recall curve is often used [9–11, 44]. Recall is defined as the number of correctly matched keypoints with respect to the total number of true corresponding keypoints between two scans, and 1-precision is defined as the number of false matches with respect to the total number of matches found in two scans.

*2) Descriptor Evaluation:* The $\beta$-grid, BSC, and SDH descriptors designed for 2D laser data, are evaluated according to the 1-precision-recall criteria in [10, 11] with weak and comparable results. Conclusions are unison in that a pure descriptor data association is not recommended for positioning purposes, but should rather be used as gating rule in a geometric association method.

All techniques studied in this paper are designed to compare two scans, when instead a scan is matched to features extracted in a map, the concept of a "true" correspondence is nontrivial. Furthermore, it is problematic that the $\beta$-grid descriptor cannot be computed according to its original definition in this setting. The construction of the $\beta$-grid descriptor takes the nature of the laser ray into consideration, where the approaching angle of the ray is used to estimate probabilities of sectors in the linear-polar histogram being occupied or not. With features extracted from a map, this data is not available. This inflexibility is a huge drawback for this technique and in this study the descriptor gated association for FLIRT-$\beta$-grid is performed by matching detected keypoints to a noise free

scan, simulated at the true position. This can only overestimate the performance causing correspondences to be better, since this decreases the demands on view-point invariance.

Further, since the descriptors are designed to perform well on features found by a specific detector, it is hard to define an objective quality measure that does not also depend on the quality of the detector. Therefore, the descriptors are in this paper only evaluated according to how they perform together with their accompanying detector, as a gating rule in a geometric association method. This is done in Sec. VIII-A1, once details for the position estimation procedure have been presented.

## VII. Underground Positioning Application

For the study of the underground positioning problem a mid-articulated vehicle equipped with odometry, *inertial measurement unit* (IMU) and two 2D lidars is considered. The pose is represented by the state vector $\mathbf{x} = \begin{bmatrix} x, y, \theta \end{bmatrix}^T$ comprising the Cartesian $(x, y)$-coordinates and the heading, $\theta$. An *unscented Kalman filter* (UKF) [51] performs the estimation with the lidar data being utilized in the correction step. This section provides details on how the measurement update is implemented.

### A. Pose Estimation from Feature Points

When a feature extraction method is applied to the raw lidar data, the measurement in each update makes up a point cloud in the vehicle coordinate frame (see Fig. 6 for definitions of coordinate systems),

$$\mathbf{y} = \begin{bmatrix} \mathbf{p}_1^T & \mathbf{p}_2^T & \dots & \mathbf{p}_N^T \end{bmatrix}^T. \tag{11}$$

Each $\mathbf{p}_i,\ i = 1, \dots, N$, is a point in the vehicle coordinate system represented by its Cartesian coordinates,

$$\mathbf{p}_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}. \tag{12}$$

$N$ is the number of feature points where a matching feature in the map is found. The number of found features in each laser scan varies and not all detected features will have a match in the map.

To perform the UKF correction step with this setup the following preprocessing steps are performed.

1) The map is preprocessed by applying the feature detection method of choice on an interpolated version of the map, see Fig. 6. If the method requires, *descriptors* for each detected feature are computed. This is done once and for all with a specific map.
2) For each new available laser scan, features are extracted according to the chosen feature extraction method. The identified features are then converted to points in the vehicle coordinate frame.
3) For each found feature a descriptor is computed according to the chosen feature extraction method.
4) The features from the laser scan are then matched to features in the map and saved as a list of currently visible map features, $\mathbf{m}$, represented as points in the global frame.

The measurement equation can then be formulated as a transform of visible feature points in the map, given in the global coordinate system, to points in vehicle coordinate frame for a given state $\mathbf{x}$,

$$\mathbf{y} = \mathcal{R}(\mathbf{x})(\mathbf{m} - \mathbf{x}_{1,2}) + \mathbf{e}. \tag{13}$$

The Cartesian components $(x, y)$ of the state vector is denoted by $\mathbf{x}_{1,2}$, $\mathcal{R}$ is a rotation matrix, and $\mathbf{e} \sim \mathcal{N}(0, \mathbf{R})$ is additive measurement noise, assumed white with covariance matrix $\mathbf{R}$.

### B. Direct Scan Matching

There are many examples of methods for scan matching of 2D range data that use the sensor data directly for data association, without first extracting features, see Fig. 1. The output from these direct methods is often a rigid body transform $(R, t)$ of the position. A well established method worth introducing is the *iterative closest point* (ICP) algorithm [52, 53], which is included in this paper for validation purposes.

Given two point clouds

$$\mathbf{p}_k = \left\{ p_i^k \right\}_{i=1}^{N_k}, \tag{14a}$$

$$\mathbf{p}_l = \left\{ p_i^l \right\}_{i=1}^{N_l}, \tag{14b}$$

with $p_i \in \mathbb{R}^D$, the following optimization problem is solved

$$\min_{R,t} \mathbf{C}(\mathbf{p}_k, \mathbf{p}_l) = \min_{R,t} \sum_{i=1}^{N_k} \sum_{j=i}^{N_l} w_{i,j} \left\| p_i^k - \left( R p_j^l + t \right) \right\|^2, \tag{15}$$

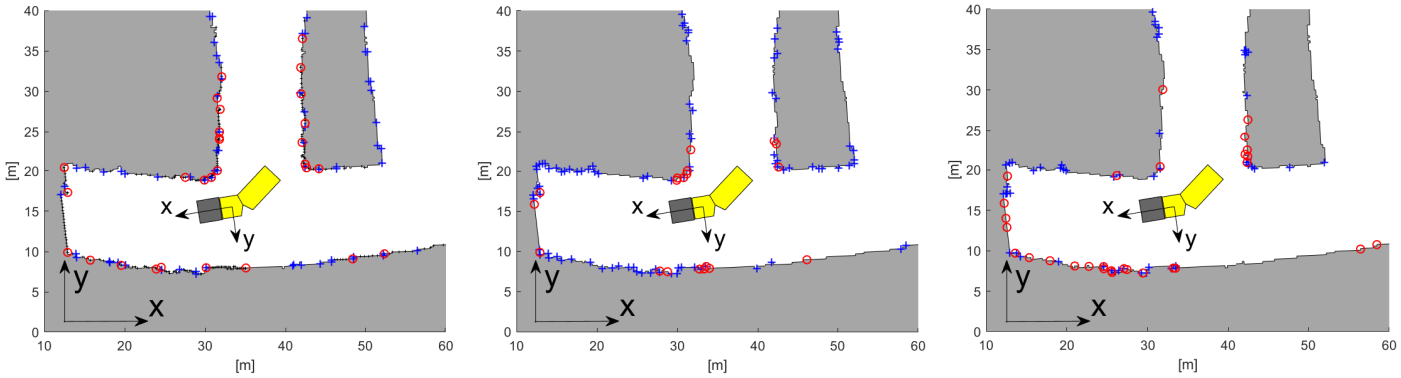
where $w_{i,j}$ is 1 if point $p_i^k$ and point $p_j^l$ are assumed to describe the same point in space, and 0 otherwise. The ICP algorithm iteratively finds the nearest neighbor point pair, then computes $(R, t)$ by solving the optimization problem and repeats until convergence.

The ICP algorithm is fast and in general produces good results when the initial offset between the scans is small [54]. The original version of the algorithm provides no estimate of the uncertainty of the resulting transform, but various versions have been developed over the years to improve performance and specializations for specific applications [7, 27, 55–58]. In [59] ICP is used for underground navigation.

Other scan matching algorithms are *e.g.*, the *normal distribution transform* (NDT) [60] where the 2D plane is divided into cells and each cell is assigned a normal distribution modelling the probability of obtaining an observation, or the *conditional random field* (CRF) [61] used in [54]. Locally defined features such as, relative distances between points in the scan, angle between the segments connecting a point to its neighbors, or sum of distances to neighboring points, together comprise a descriptor for a complete scan.

If one of these direct data association methods are used, a linear measurement equation can be used in the filter, that now acts as a low pass filter for the estimate. The outcome from a method operating directly on the raw lidar data is often formulated as a transform and can be converted to an indirect measurement of the state variable. The measurement equation is then reduced to the identity matrix,

$$\mathbf{y} = \mathbf{x} + \mathbf{e}. \tag{16}$$




(a) Features found with FLIRT detector. (b) Features found with tuned FALKO detector. (c) Features found with tuned BID detector.

Fig. 6. Detected features in a laser scan (red circles), are represented as a point cloud in the coordinate system with origin on the front wheel axis of the vehicle. Blue crosses marks the features found in the map and are defined in the global coordinate frame. Note that for the FLIRT detector only keypoints associated to the same scale are considered for matching even though all detected keypoints have the same notation in this figure.

with the same assumption on the measurement noise $\mathbf{e}$, as in (13). In this paper a UKF is always used even when the measurement equation is linear. This is partly to keep all experiments homogeneous and comparable, but also because the time update is nonlinear for this application. Note that a UKF update basically reduces to a standard Kalman filter update in case of a linear function.

## VIII. Positioning Experiments

This section presents results from experiments conducted in the simulated mine environment described in Section VI-A. All laser data used in this section is simulated using the SVM model.

The UKF parameterization with $\alpha$, $\beta$, and $\kappa$ as suggested in [51] is used in this evaluation. The suggested parameter values in [51] proved to perform poorly in initial tests, whereas $\alpha = 0.8$, $\beta = 2.0$, $\kappa = 0.0$ obtained from hand tuning of the filter perform significantly better. Hence, these values are used in the experiments together with the process noise covariance matrix set to $Q = \text{diag}\left(0.002, 0.002, 10^{-6}\right)$, and the measurement noise $R = 0.5^2\mathbf{I}$.

### A. Measurement Update

Monte Carlo simulations of a single measurement update in the UKF algorithm is performed to analyze what accuracy the different feature extraction methods give in a position estimation application. The vehicle is positioned in four different poses, see Fig. 5, with different capabilities of obtaining informative laser measurements. The initial pose is perturbed from the true state with $\mathbf{x}_0 = \mathbf{x}_{\text{true}} + \delta_{\mathbf{x}}$, with $\delta_{\mathbf{x}} \sim \mathcal{N}(0, \mathbf{P}_0)$, to simulate the position uncertainty originating from prior odometry errors.

*1) Descriptor Data Association:* As mentioned in Sec. VI-D2, an evaluation of the descriptors is conducted by comparing pure geometric data association to geometric association using the descriptors for gating. The descriptors $\beta$-grid, BSC, and SDH are used in conjunction with the detectors FLIRT, FALKO and BID, respectively, and 10 000 realizations of a complete UKF measurement update are

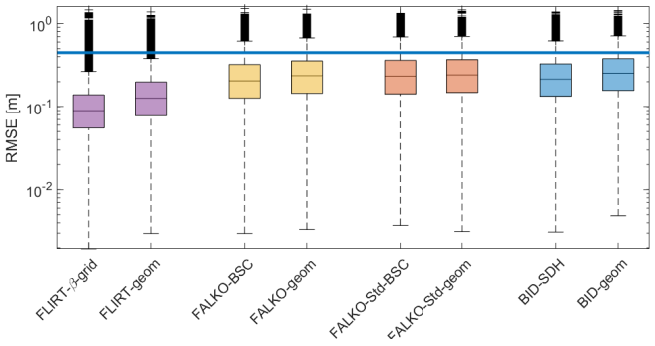



Fig. 7. RMSE after one measurement update based on 10 000 realizations in each of the four poses in Fig. 5. Different feature detectors paired with a descriptor-gated or pure geometric data association are used. The blue horizontal line represents the *a priori* error, $\sqrt{\text{tr}(\mathbf{P}_0)}$. The edges of the boxes indicates 25th and 75th percentiles, and the median is marked with a horizontal line. The whiskers extend to the most extreme data points not considered outlier, and outliers are marked individually with a cross.

performed with an initial perturbation of the position obtained by sampling from $\delta_{\mathbf{x}} \sim \mathcal{N}(0, \mathbf{P}_0)$ with,

$$\mathbf{P}_0 = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.00076 \end{bmatrix}. \tag{17}$$

Each realization also sample its own laser scan measurement noise from the SVM model. All descriptor histograms are constructed with $r_n = 4$ and $\alpha_n = 12$ as suggested in the original papers, and features in the map within a radial distance ($d_{\max}$) of three standard deviations of the *a priori* uncertainty are considered matching candidates. For keypoints with a difference in orientation smaller than a threshold $d_\alpha$, the keypoint with the minimum distance according to the distance function associated with the used descriptor, is considered a match. Matched features are then used as measurements in the UKF update. For FALKO-BSC, $d_\alpha = \pi/6$ and for FLIRT-$\beta$-grid and BID-SDH, $d_\alpha = \pi/10$ is used. The values are chosen manually by performing simulations and with knowledge of the methods and the characteristics of the specific problem. A larger threshold is used for FALKO-BSC since the orientation

TABLE IV
DESCRIPTOR DATA ASSOCIATION. NUMBER OF LASER SCANS UNABLE TO FIND A SINGLE KEYPOINTS MATCH IN THE SET OF KEYPOINTS EXTRACTED IN THE MAP, AND THE MEAN NUMBER OF MATCHED FEATURE IN EACH SCAN IS PRESENTED FOR EACH OF THE METHODS.

| | Scans without match | | |
|---|---|---|---|
| | # | % | #matches/scan |
| FLIRT-$\beta$-grid | 38 | 0.1 | 14.7 |
| FLIRT-geom | 5 | 0.01 | 20.4 |
| FALKO-BSC | 283 | 0.7 | 8.1 |
| FALKO-geom | 5 | 0.01 | 14.9 |
| FALKO-Std-BSC | 32 | 0.08 | 10.6 |
| FALKO-Std-geom | 1 | 0.003 | 20.0 |
| BID-SDH | 267 | 0.7 | 5.4 |
| BID-geom | 4 | 0.01 | 16.8 |

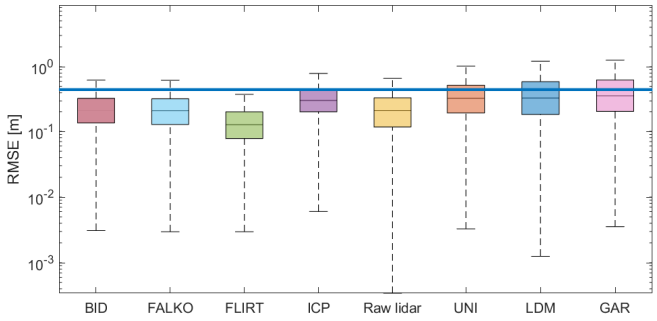



Fig. 8. RMSE for 10 000 realizations of a measurement update step using different methods. The horizontal blue line is the square-root of the trace of the initial covariance matrix $\mathbf{P}_0$. The edges of the boxes indicates 25th and 75th percentiles, and the median is marked with a horizontal line. The whiskers extend to the most extreme data points.

of these features is more uncertain, due to how they are computed.

In Fig. 7 the *root mean square error* (RMSE) for the descriptor gated data association is compared to pure geometric nearest neighbor data association with $d_{\max} = 0.1$ m as threshold, and Table IV presents numerical data. Fig. 7 confirms earlier results that the descriptors do not add much value when a fairly good initial value can be fed to a pure geometric association. FLIRT is the only method where the usage of the descriptor actually makes a significant difference. However, the descriptor experiment is in this case expected to overestimate the performance due to the customization that had to be made to the evaluation of the $\beta$-grid descriptor described earlier in Section VI-D2. The tuned parameter setup of FALKO is in this experiment, unlike the results in Table III, resulting in marginally lower errors than the default parameter setup. This is despite the fact that there are on average a lower number of matches with the tuned parameter setup, which implies that the standard parameter setup for FALKO results in many erroneous matches since the detected keypoints are not significant enough. For BID, the tuned parameter setup in Table II is used. With standard parameter values, BID gives so few keypoints in the map that for many of the poses there exists no keypoints in the field-of-view.

*2) Compare to Raw Range Data Association:* In this section the feature extraction methods are compared to the case when raw range data is sent to the UKF, *i.e.*, each ray is matched individually to the map. This is similar to the

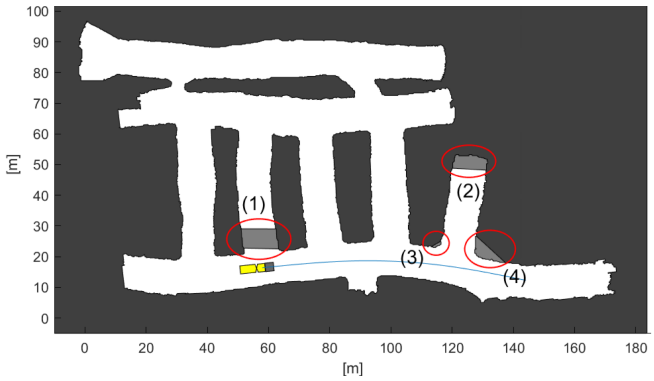



Fig. 9. The vehicle travels along the blue line starting from left to right. The trajectory both starts and stops with zero velocity and light grey areas show modification done to simulate errors in the map. (1) A tunnel is filled with material since the map was created. (2) The tunnel is shortened as if rocks have fallen down. (3) A small chunk of material is cut-off in a corner. (4) A larger chunk of material is cut-off in a corner.

navigation system described in [4], but instead of matching the rays to locally created occupancy grid maps, the data is matched to a given global metric map of the operation area. The approach of matching raw lidar data is also used in the sensor selection methods described in [6], where the aim is to cut computational complexity. The methods are also compared to a standard ICP variant using the point-to-plane error metric without outlier rejection. The implementation used is provided as an open-source project [62] where details are provided in the related paper [63].

The RMSE of 10 000 realizations is depicted in Fig. 8, again using the SVM to simulate laser data. (Note that this is not 10 000 replicated laser scans, but 10 000 realizations of the distribution given by the SVM model.) Data association for FALKO and BID are performed with the descriptor gated geometric data association, and for FLIRT a pure geometric association is used. For FALKO, the tuned parameter setup is used in this case.

The results are presented in Fig. 8. Note that the methods UNI, LDM, and, GAR also uses raw range data association, but only utilizing a subset of available data in each scan. The feature extraction methods perform in this evaluation in parity with the 'Raw lidar'-method. It is worth noting that FALKO requires significantly less computational resources than all of the other methods, still performing better than the methods of selecting subsets.

### B. Trajectory

A trajectory is prepared with accompanying laser data, simulated using the SVM model, to investigate how the feature extraction methods perform in realistic dynamic position estimation. The trajectory is given in Fig. 9 and the vehicle starts from standing still at the leftmost point. This trajectory is representative for movements of the vehicle in this environment, exposing the system to, *e.g.*, long tunnels, cross-sections, and irregularities of the walls captured in the laser data. Since the UKF outputs an estimate of the state covariance in each time increment, the radial distance $d_{\max}$ will change dynamically.

TABLE V
MEAN RMSE AND THE MEAN NUMBER OF MATCHED FEATURE IN ONE SCAN IS AVERAGED OVER ALL MODIFICATIONS PRESENTED IN FIG. 10. THE MEAN NEES IS GIVEN FOR THE ORIGINAL MAP (FIG. 10A) AND WHEN ALL MODIFICATION ARE APPLIED (FIG. 10F).

| Method | RMSE | # matches | NEES (orig) | NEES (mod) |
|---|---|---|---|---|
| Raw lidar | 1.17 | – | 320 | 1 000 000 |
| FLIRT-geom | 0.023 | 3.4 | 0.17 | 0.21 |
| FALKO-BSC | 0.042 | 6.2 | 1.3 | 0.94 |
| BID-SDH | 0.053 | 5.1 | 6.3 | 8.1 |
| ICP | 0.27 | – | 7.0 | 26 |

For FLIRT where only pure geometric association is used, the threshold is fixed to 0.1 m.

In Fig. 10(a) the RMSE of the estimated state is presented for each time increment along the trajectory. The poor result for the basic ICP variant [63] could probably be improved by a sophisticated outlier rejection principle. For example, as mentioned in Sec. VII-B many variations of the algorithm addressing specific properties do exist, or as discussed in Sec. V-B, ICP can be used in conjunction with RANSAC to decrease the influence of outliers. However, this adds complexity to the system and it is something that is somewhat automatic in the feature extraction approach. Since RANSAC can also be used to boost the performance of the feature extraction approach, a better understanding of the performance of the actual methods is given by comparing the pure versions of them, without adding extra steps.

In Fig. 10(b)–(f), laser data is still simulated in the original map but the UKF is handed a modified version to perform the position estimate in. This mimics a situation when a tunnel is extended/closed, a corner is worn down, or some other change in the environment that is not updated accordingly in the map. The ICP approach has a relatively high RMSE in this setting. It is somewhat robust to small changes, while the performance is considerably worse for larger modifications. The estimates produced by the feature extraction methods are all relatively unaffected by the modifications in the map, while the approach where each ray individually enters the measurement equation is highly sensitive to imperfections in the map. For some of the modifications the filter diverges with this approach.

Out of the feature extraction methods evaluated in this experiment FLIRT-geom performs the best, despite the fact that only a pure geometric data association is used. In Table V the mean RMSE for all maps and time increments are presented for each of the methods, and FLIRT gives an estimate with only half the error of the others. Note here that the average number of matched features in each scan is significantly lower for the FLIRT method, compared to Table IV, this is since the prior uncertainty, that now changes in each time step, highly affect the possibility to extract stable keypoints. However, the FLIRT method still manages to find the most significant ones, since the position estimates still has a high quality.

Table V also presents the *normalized estimated error square* (NEES) averaged over the trajectory states and realizations for the case with the original map (Fig. 10a) and when all modifications are added (Fig. 10f). The NEES is defined as

$$\text{NEES} = (\mathbf{x}_{\text{true}} - \hat{x})^\top \hat{P}^{-1} (\mathbf{x}_{\text{true}} - \hat{x}), \tag{18}$$

where $\hat{x}$ and $\hat{P}$ are the state and covariance estimates obtained from the UKF. Under the Gaussian assumption this becomes $\chi^2$-distributed with mean $\dim(\mathbf{x}) = 3$ [64]. NEES-values above 3 indicates inconsistent estimates and values below 3 indicates an over-conservative uncertainty estimate. All feature extraction methods have NEES values unaffected by the modifications whereas 'raw lidar' and ICP gives inconsistent results with errors in the map. The NEES measure is highly dependent on the tuning of the UKF itself (this also applies to the RMSE measure), especially the process and measurement noise, $Q$ and $R$, respectively. There is a potential to improve the performance of the 'raw lidar' method by tuning these matrices or even making it less sensitive to errors in the map by using an adaptive tuning technique such as the ones described in [65, 66].

## IX. CONCLUSION

In this paper we have addressed the problem of robust and highly accurate position estimation in an underground mine. By considering a vehicle equipped with 2D laser scanners we have surveyed available methods to process data from such sensors with the goal of positioning the vehicle in a predefined map.

In particular, methods of extracting features from 2D laser data have been investigated and three state-of-the-art methods have been identified, FLIRT-$\beta$-grid, FALKO-BSC and BID-SDH. The methods have been adapted to fit the scenario of localization in a predefined map. Experiments conducted in a simulated underground mine environment have shown that all of these methods perform better than the point cloud scan matching method of a pure ICP, and in parity with the method of matching each ray individually to the map in situations without map modifications. Results also demonstrates that all the feature extraction methods are much more robust to errors and imperfections in the map, simply because no matches are found in those regions. This is an important advantage since the underground mine environment changes dynamically and with a method insensitive to errors, the predefined map is not forced to change accordingly.

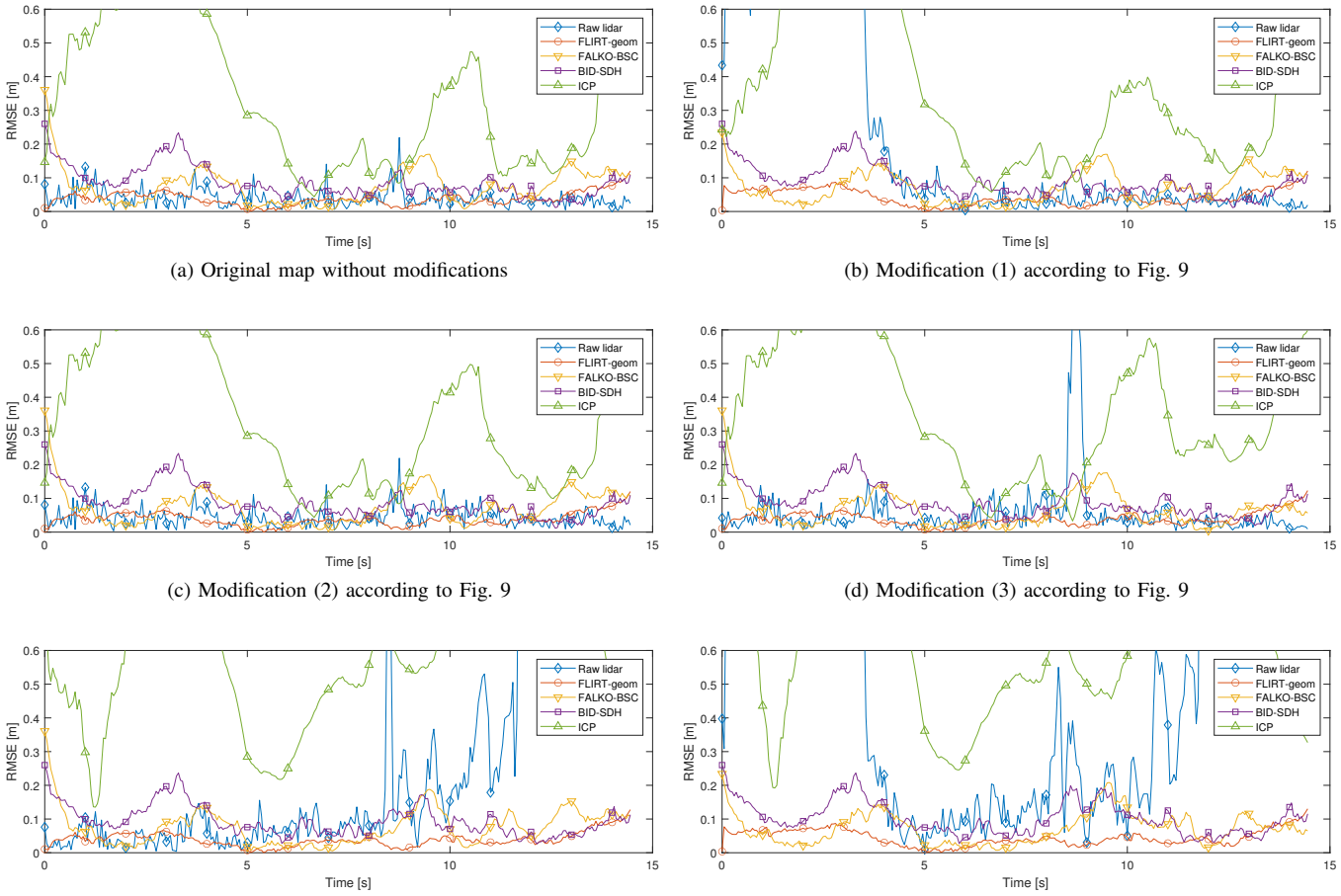
In general, the feature extraction methods are sensitive to parameter settings. Hence, all parameters have been manually tuned to fit this particular application. The best performing method is the FLIRT feature detector in combination with a pure geometric data association with default parameters suggested in [9]. The used nearest neighbor geometric data association could probably be improved by applying a more sophisticated method for outlier rejection *e.g.*, RANSAC.

Since the feature extraction methods are robust to errors in the map, they have a great potential of performing well in a dynamic environment. Future work will explore how such methods can be used to enable automatic updates of the map.

## REFERENCES

[1] L. Thrybom, J. Neander, E. Hansen, and K. Landernas, "Future challenges of positioning in underground mines," in *Proc. 2nd IFAC Conf. on Embedded Systems, Computer Intelligence and Telematics CESCIT*, vol. 48, no. 10, Maribor, Slovania, 2015, pp. 222–226.

[2] J. Marshall, A. Bonchis, E. Nebot, and S. Scheding, *Robotics in mining*. Cham: Springer International Publishing, 2016, ch. 59, pp. 1549–1576.




(a) Original map without modifications

(b) Modification (1) according to Fig. 9

(c) Modification (2) according to Fig. 9

(d) Modification (3) according to Fig. 9

(e) Modification (4) according to Fig. 9

(f) All modifications applied at once.

Fig. 10. The RMSE of the pose in each time increment along the trajectory in Fig. 9, estimated with different methods. This is the average result of 10 realizations.


[3] J. Paraszczak, A. Gustafson, and H. Schunnesson, "Technical and operational aspects of autonomous LHD application in metal mines," *International Journal of Mining, Reclamation and Environment*, vol. 29, no. 5, pp. 391–403, 2015.

[4] J. Marshall, T. Barfoot, and J. Larsson, "Autonomous underground tramming for center-articulated vehicles," *Journal of Field Robotics*, vol. 25, no. 6-7, pp. 400–421, 2008.

[5] H. Mäkelä, "Overview of LHD navigation without artificial beacons," *Robotics and Autonomous Systems*, vol. 36, no. 1, pp. 21–35, 2001.

[6] K. Nielsen and G. Hendeby, "Sensor management in 2D lidar-based underground positioning," in *Proc. 23th IEEE Int. Conf. Inform. Fusion*, Virtual conference, Jul. 6–9 2020.

[7] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," in *Proc. 3rd Int. Conf. on 3-D Digital Imaging and Modeling*, Quebec City, Canada, 2001.

[8] S. Krig, *Computer vision metrics: survey, taxonomy, and analysis*. Berkeley, CA: Apress, 2014, ch. Interest Point Detector and Feature Descriptor Survey, pp. 217–282.

[9] G. D. Tipaldi and K. O. Arras, "FLIRT - interest regions for 2D range data," in *Proc. IEEE Int. Conf. on Robotics and Automation*, Anchorage, AK, USA, 2010, pp. 3616–3622.

[10] F. Kallasi, D. L. Rizzini, and S. Caselli, "Fast keypoint features from laser scanner for robot localization and mapping," *IEEE Robot. Autom. Lett.*, vol. 1, no. 1, pp. 176–183, 2016.

[11] M. Usman, A. M. Khan, A. Ali, S. Yaqub, K. M. Zuhaib, J. Y. Lee, and C.-S. Han, "An extensive approach to features detection and description for 2-D range data using active B-splines," *IEEE Robot. Autom. Lett.*, vol. 4, no. 3, pp. 2934–2941, 2019.

[12] K. Lai, L. Bo, X. Ren, and D. Fox, "A large-scale hierarchical multi-view RGB-D object dataset," in *Proc. Int. Conf. on Robotics and Automation*, Shanghai China, 2011, pp. 1817–1824.

[13] J. Guivant. (2006) Victoria Park data set. The University of Sydney. [Online]. Available: http://www-personal.acfr.usyd.edu.au/nebot/victoria_park.htm

[14] R. B. Rusu and S. Cousins, "3D is here: point cloud library (PCL)," in *Proc. IEEE Int. Conf. on Robotics and Automation*, Shanghai, China, 2011, pp. 1–4.

[15] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.

[16] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: speeded up robust features," in *Proc. European Conference on Computer Vision*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 404–417.

[17] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF," in *Proc. Int. Conf. on Computer Vision*, Barcelona, Spain, 2011, pp. 2564–2571.

[18] R. Unnikrishnan and M. Hebert, "Multi-scale interest regions from unorganized point clouds," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Anchorage, AK, USA, 2008, pp. 1–8.

[19] S. Salti, F. Tombari, and L. D. Stefano, "A performance evaluation of 3D keypoint detectors," in *Proc. Int. Conf. on 3D Imaging, Modeling, Processing, Visualization and Transmission*, Hanghou, China, 2011, pp. 236–243.

[20] V. K. Ghorpade, P. Checchin, L. Malaterre, and L. Trassoudaine, "Performance evaluation of 3D keypoint detectors for time-of-flight depth data," in *Proc. 14th Int. Conf. on Control, Automation, Robotics and Vision*. Phuket, Thailand: IEEE, 2016, pp. 1–6.

[21] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, and J. Wan, "3D object recognition in cluttered scenes with local surface features: a survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2270–2287,

2014.
[22] C. G. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vision Conference*, vol. 1, Manchester, UK, 1988, pp. 147–151.
[23] J. Shi and Tomasi, "Good features to track," in *Proc. Conf. on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 1994, pp. 593–600.
[24] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. Int. Conf. on Computer Vision*, vol. 2, Kerkyra, Greece, 1999, pp. 1150–1157.
[25] Y. Li and E. B. Olson, "Extracting general-purpose features from LIDAR data," in *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, Anchorage, AK, USA, May 2010, pp. 1388–1393.
[26] Y. Li and E. B. Olson, "Structure tensors for general purpose LIDAR feature extraction," in *Proc. 2011 IEEE Int. Conf. on Robotics and Automation*, Shanghai, China, 2011.
[27] M. Bosse and R. Zlot, "Map matching and data association for large-scale two-dimensional laser scan-based SLAM," *The International Journal of Robotics Research*, vol. 27, no. 6, pp. 667–691, 2008.
[28] ——, "Keypoint design and evaluation for place recognition in 2D lidar maps," *Robotics and Autonomous Systems*, vol. 57, no. 12, pp. 1211–1224, 2009.
[29] M. A. Carreira-Perpiñán and R. S. Zemel, "Proximity graphs for clustering and manifold learning," in *Proc. of the 17$^{th}$ Int. Conf. on Neural Information Processing Systems*. MIT Press, 2004, pp. 225–232.
[30] D. Rodriguez-Losada, F. Matia, and R. Galan, "Building geometric feature based maps for indoor service robots," *Robotics and Autonomous Systems*, vol. 54, no. 7, pp. 546–558, 2006.
[31] V. Nguyen, A. Martinelli, N. Tomatis, and R. Siegwart, "A comparison of line extraction algorithms using 2D laser rangefinder for indoor mobile robotics," in *2005 Proc. Int. Conf. on Intelligent Robots and Systems*, 2005, pp. 1929–1934.
[32] S. T. Pfister, S. I. Roumeliotis, and J. W. Burdick, "Weighted line fitting algorithms for mobile robot map building and efficient data representation," in *Proc. Int. Conf. on Robotics and Automation*, vol. 1, Taipei, Taiwan, 2003, pp. 1304–1311.
[33] P. Smith, I. Reid, and A. J. Davison, "Real-time monocular SLAM with straight lines," in *Proc. the British Machine Vision Conference*. Edinburgh, UK: British Machine Vision Association, 2006, pp. 17–26.
[34] M. Veeck and W. Burgard, "Learning polyline maps from range scan data acquired with mobile robots," in *Proc. Int. Conf. on Intelligent Robots and Systems (IROS)*, vol. 2, Sendai, Japan, 2004, pp. 1065–1070.
[35] P. Nunez, R. Vazquez-Martin, J. C. del Toro, A. Bandera, and F. Sandoval, "Feature extraction from laser scan data based on curvature estimation for mobile robotics," in *Proc. Int. Conf. on Robotics and Automation (ICRA)*, Orlando, FL, USA, 2006, pp. 1167–1172.
[36] J. Vandorpe, H. V. Brussel, and H. Xu, "Exact dynamic map building for a mobile robot using geometrical primitives produced by a 2D range finder," in *Proc. Int. Conf. on Robotics and Automation (ICRA)*, vol. 1, Minneapolis, MN, USA, 1996, pp. 901–908.
[37] S. Zhang, M. Adams, F. Tang, and L. Xie, "Geometrical feature extraction using 2D range scanner," in *Proc. 4$^{th}$ Int. Conf. on Control and Automation*, Montreal, Canada, 2003, pp. 901–905.
[38] K. Meier, S.-J. Chung, and S. Hutchinson, "Visual-inertial curve simultaneous localization and mapping: creating a sparse structured world without feature points," *Journal of Field Robotics*, vol. 35, no. 4, pp. 516–544, 2017.
[39] L. Pedraza, D. Rodriguez-Losada, F. Matia, G. Dissanayake, and J. V. Miro, "Extending the limits of feature-based SLAM with b-splines," *IEEE Trans. Robot.*, vol. 25, no. 2, pp. 353–366, 2009.
[40] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: binary robust independent elementary features," in *Proc. European Conference on Computer Vision*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 778–792.
[41] M. E. Leventon, W. E. L. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours," in *Proc. Conf. on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, Hilton Head Island, SC, USA, 2000, pp. 316–323.
[42] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, apr 2002.
[43] P. L. Rosin, "Measuring corner properties," *Computer Vision and Image Understanding*, vol. 73, no. 2, pp. 291–307, 1999.
[44] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, 2005.
[45] M. Himstedt, J. Frost, S. Hellbach, H.-J. Bohme, and E. Maehle, "Large scale place recognition in 2D LIDAR scans using Geometrical Landmark Relations," in *Proc. Int. Conf. on Intelligent Robots and Systems*, Chicago, IL, USA, 2014, pp. 5030–5035.
[46] T. Bailey, E. M. Nebot, J. K. Rosenblatt, and H. F. Durrant-Whyte, "Data association for mobile robot navigation: a graph theoretic approach," in *Proc. Int. Conf. on Robotics and Automation (ICRA)*, vol. 3, San Francisco, CA, USA, 2000, pp. 2512–2517.
[47] F. Kallasi and D. L. Rizzini, "Efficient loop closure based on FALKO lidar features for online robot localization and mapping," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, Daejeon, South Korea, 2016.
[48] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381––395, 1981.
[49] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, pp. 43–72, 2005.
[50] A. Tallavajhula, "Lidar simulation for robotic application development: modeling and evaluation," Ph.D. dissertation, The Robotics Institute Carnegie Mellon University Pittsburgh, May 2018.
[51] E. A. Wan and R. V. D. Merwe, "The unscented Kalman filter for nonlinear estimation," in *Proc. IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium*, Lake Louise, Alberta, Canada, 2000.
[52] Y. Chen and G. Medioni, "Object modeling by registration of multiple range images," in *Proc. IEEE Int. Conf. on Robotics and Automation*, vol. 3, Sacramento, CA, USA, 1991, pp. 2724–2729.
[53] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, 1992.
[54] F. Ramos, D. Fox, and H. Durrant-Whyte, "CRF-Matching: conditional random fields for feature-based scan matching," in *Proc. Robotics: science and Systems*, Atlanta, GA, USA, 2007.
[55] F. Lu and E. Milios, "Robot pose estimation in unknown environments by matching 2D range scans," *Journal of Intelligent and Robotic Systems*, vol. 18, no. 3, pp. 249–275, 1997.
[56] O. Bengtsson and A. J. Baerveldt, "Localization in changing environments - estimation of a covariance matrix for the IDC algorithm," in *Proc. Int. Conf. on Intelligent Robots and Systems (IROS)*, vol. 4, Maui, HI, USA, 2001, pp. 1931–1937.
[57] S. T. Pfister, K. L. Kriechbaum, S. I. Roumeliotis, and J. W. Burdick, "Weighted range sensor matching algorithms for mobile robot displacement estimation," in *Proc. Int. Conf. on Robotics and Automation (ICRA)*, vol. 2, Washington, DC, USA, 2002, pp. 1667–1674.
[58] N. Gelfand, L. Ikemoto, S. Rusinkiewicz, and M. Levoy, "Geometrically stable sampling for the ICP algorithm," in *Proc. 4$^{th}$ Int. Conf. on 3-D Digital Imaging and Modeling (3DIM)*, Banff, Alta., Canada, 2003.
[59] R. Madhavan, M. W. M. G. Dissanayake, and H. F. Durrant-Whyte, "Autonomous underground navigation of an LHD using a combined ICP-EKF approach," in *Proc. IEEE Int. Conf. on Robotics and Automation*, vol. 4, Leuven, Belgium, 1998, pp. 3703–3708.
[60] P. Biber and W. Strasser, "The normal distributions transform: a new approach to laser scan matching," in *Proc. 2003 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, USA, 2003.
[61] J. Lafferty, A. Mccallum, and F. Pereira, "Conditional random fields: probabilistic models for segmenting and labeling sequence data," in *Proc. Int. Conf. on Machine Learning (ICML)*, Williamstown, MA, USA, 2001, pp. 282–289.
[62] A. Geiger. [Online]. Available: https://github.com/symao/libicp
[63] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, USA, 2012, pp. 3354–3361.
[64] F. Gustafsson, *Statistical sensor fusion*. Studentlitteratur AB, 2018.
[65] Q. Song, Z. Jiang, and J. Han, "Noise covariance identification based adaptive UKF with application to mobile robot systems," in *Proc. IEEE Int. Conf. on Robotics and Automation*, Roma, Italy, 2007, pp. 4164–4169.
[66] Z. Chen, C. Heckman, S. Julier, and N. Ahmed, "Weak in the NEES?: auto-tuning Kalman filters with Bayesian optimization," in *Proc. 21$^{st}$ Int. Conf. on Information Fusion (FUSION)*, Cambridge, UK, 2018, pp. 1072–1079.

**Kristin Nielsen** is a Phd student at the division of Automatic Control, Department of Electrical Engineering, Linköping University. She received her MSc in Engineering Physics in 2011 from Uppsala University and has since been employed as Automation Software Developer at Epiroc Rock Drills AB. Epiroc is a company manufacturing vehicles and equipment, as well as offering automation solutions, for the mining industry. Kristin's Phd studies are conducted in collaboration with Epiroc and is partially sponsored by the Wallenberg AI, Autonomous Systems and Software Program (WASP).

**Gustaf Hendeby** (S'04-M'09-SM'17) is Associate Professor and Docent in the division of Automatic Control, Department of Electrical Engineering, Linköping University. He received his MSc in Applied Physics and Electrical Engineering in 2002 and his PhD in Automatic Control in 2008, both from Linköping University. He worked as Senior Researcher at the German Research Center for Artificial Intelligence (DFKI) 2009–2011, and Senior Scientist at Swedish Defense Research Agency (FOI) and held an adjunct Associate Professor position at Linköping University 2011–2015. Dr. Hendeby's main research interests are stochastic signal processing and sensor fusion with applications to nonlinear problems, target tracking, and simultaneous localization and mapping (SLAM). He has experience of both theoretical analysis as well as implementation aspects.