Semi-Supervised 3D Place Categorisation by Descriptor Clustering

Martin Magnusson,[1] Tomasz Piotr Kucner,[1] Saeed Gholami Shahbandi,[2]
Henrik Andreasson,[1] and Achim J. Lilienthal[1]

***Abstract*— Place categorisation; i. e., learning to group perception data into categories based on appearance; typically uses supervised learning and either visual or 2D range data.**

This paper shows place categorisation from 3D data without any training phase. We show that, by leveraging the NDT histogram descriptor to compactly encode 3D point cloud appearance, in combination with standard clustering techniques, it is possible to classify public indoor data sets with accuracy comparable to, and sometimes better than, previous supervised training methods. We also demonstrate the effectiveness of this approach to outdoor data, with an added benefit of being able to hierarchically categorise places into sub-categories based on a user-selected threshold.

This technique relieves users of providing relevant training data, and only requires them to adjust the sensitivity to the number of place categories, and provide a semantic label to each category after the process is completed.

I. INTRODUCTION

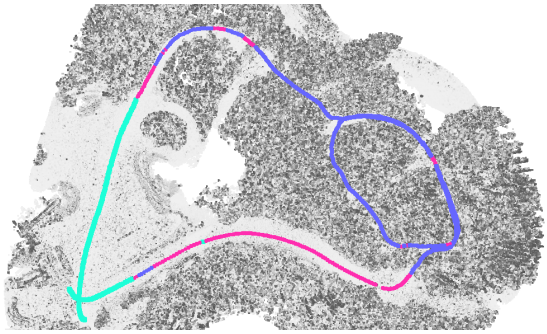
Place categorisation is the problem of labelling environment observations. E. g., it might be relevant for a service robot to determine whether it is in an office or a kitchen.

Place categorisation may improve service robots' communication capabilities with humans [5, 10], and holds a central place in semantic mapping. This ability may be particularly useful for teleoperation, so that the robot can tell operators what environment it is in, when that is hard to see from a video stream alone. Furthermore, by being able to detect that a new place has the same type as an already visited one, a field robot may generalise learned environment-specific parameters; such as odometry accuracy, traversability, etc.

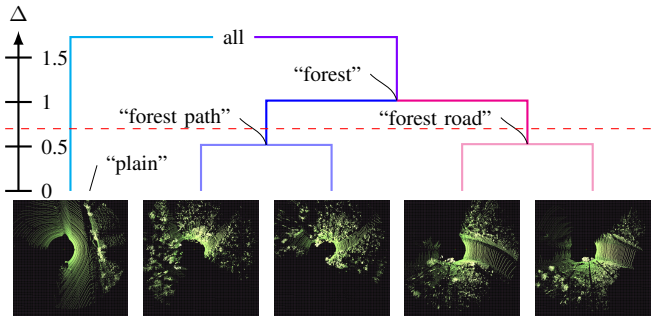
There are two main holes in the existing literature on place categorisation (see Sec. II) that this paper aims to fill.

1) Previous work typically uses fully supervised learning, where the classifier first needs to be trained on labelled data from known locations, with user-selected categories. In contrast, the method presented in this paper automatically computes a pertinent grouping of regions and lets the user assign meaningful labels afterwards.
2) Previous work mostly used 2D range sensors or cameras, and has been targeted for indoor environments. We focus on 3D data, comparing our results to the few existing results on 3D place categorisation. We also present outdoor results, and provide a new data set with a mix of forest and open environments (Fig. 1).

[1]MRO lab, AASS research centre at Örebro University, Sweden.
[2]IS lab, Halmstad University, Sweden.
This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 732737 (ILIAD).



(a) Map with trajectory coloured based on hierarchical k-means++ clustering of NDT histogram distances. Colours correspond to the links intersected by the dashed threshold line in Fig. 1b.



(b) Dendrogram showing the place hierarchy found by the proposed semi-supervised place categorisation method. The "plain" and "forest" parts of the environment are well separated, and the "forest" category is split into two semantically meaningful subcategories: "road" and "path". With a selectable threshold (dashed line), the user can select a suitably detailed categorisation. In this example, setting the treshold to 0.5 or lower further splits "road" and "path" categories.

Fig. 1: Qualitative results from the *EskilstunaField* data set.

To the best of our knowledge, this is the first work that demonstrates place categorisation for unstructured (outdoor) environments without supervised learning.

In brief, we show that place categorisation can be successfully addressed using established clustering methods and a global appearance descriptor, in contrast supervised learning.

After computing appearance descriptors, and clustering them based on appearance distance, the clusters should correspond to meaningful place categories, which can subsequently be labelled by the user. A thorough evaluation of the performance of this strategy, compared to two baseline methods, is given in Sec. V.

In addition to the two main contributions listed above, we demonstrate that by using hierarchical k-means++ clustering [1] and the NDT-histogram appearance descriptor [8],

it is possible to generate a semantically meaningful tree of sub-categories, with a single user-specified threshold for how fine the categories should be (see Fig. 1b). We also evaluate DBSCAN clustering and alternative k-means seeding.

II. RELATED WORK

Previous work on place categorisation mostly deals with methods using supervised learning to train classifiers based on labelled sets of images or 2D laser range scans.

In an early paper, Mozos et al. [12] trained an AdaBoost classifier to label indoor 2D scans as room, corridor, or doorway. Pronobis et al. [13] extend this work by combining laser and camera, and training a support vector machine (SVM) to combine image and range features. In later work, Mozos et al. [10, 11] use 3D scans. This work is particularly relevant for the present paper, and we use the same 3D data set [10] in our main quantitative results (Sec. V-B, Tab. Ia).

Recently, Goeddel and Olson [5] trained a convolutional neural network (CNN) on the same indoor classes as Mozos et al. [12]. By deploying specific training techniques to tackle the tendency of CNNs to overfitting and bias, they achieve good accuracy for the room and corridor classes. Sünderhauf et al. [16] instead use a CNN for place categorisation from camera images, and complement it with a set of classifiers in order to recognise new semantic classes online.

In contrast to all the methods above, our proposed method requires no supervised training in order to find semantically meaningful categories.

PLISS [14] uses bags of words from indoor image sequences, and exploits change-point detection to detect transitions between places. PLISS was shown to achieve good accuracy on a challenging video data set, but in contrast to our approach, it could not be used on non-sequential data, and it was designed for images, which are inherently less robust to changing light conditions than laser point clouds.

Gholami Shahbandi et al. [4] present an approach conceptually similar to ours, using k-means clustering to categorise places from a warehouse without supervised learning. Compared to their work, we employ a 3D appearance descriptor and study its performance for place categorisation in combination with a selection of clustering methods, and compare it quantitatively to state-of-the-art 3D place categorisation. We also demonstrate our method in unstructured outdoor environments.

III. APPEARANCE DESCRIPTOR

To provide a compact global appearance descriptor for 3D point clouds, we use surface-shape histograms based on the NDT (normal distributions transform) representation, which have previously been used for detecting loop closures [8].

NDT [2, 7] describes geometry as a set of Gaussian PDFs arranged in a voxel grid. NDT has previously been shown to be useful for applications of point cloud registration [9] and many other tasks. One can construct a *histogram* of NDT voxels to encode point cloud appearance, by classifying the PDFs based on orientation and shape, and distance from the sensor. We will briefly describe the appearance descriptor and its associated distance function. Please refer to Magnusson et al. [8] for a more comprehensive description.

A. NDT histograms

For each PDF in an NDT voxel grid, the eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3$ and eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are computed. From the relative magnitudes of the eigenvalues three surface classes can be discerned: spherical, planar, and linear. Distributions are assigned to a class with respect to a threshold $t_e \in [0, 1]$ that quantifies a "much smaller" relation, such that if $\lambda_2 < t_e\lambda_3$, the PDF is linear, and if $\lambda_1 < t_e\lambda_2$ it is planar.

Each of these shape classes can be subdivided, using orientation for the planar and linear classes, and roughness for the spherical class. Using n_s spherical subclasses, n_p planar ones, and n_l linear ones, the basic element of the appearance descriptor is the feature vector

$$\mathbf{f} = [\underbrace{f_1, \ldots, f_{n_s}}_{\text{spherical classes}}, \underbrace{\ldots, f_{n_s+n_p}}_{\text{planar classes}}, \underbrace{\ldots, f_{n_s+n_p+n_l}}_{\text{linear classes}}]^{\mathrm{T}}, \quad (1)$$

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

In addition to shape and orientation, the distance from the origin is also informative. Therefore, each point cloud is described by a matrix $\boldsymbol{F} = [\mathbf{f}_1 \cdots \mathbf{f}_{n_r}]$, where each column $\mathbf{f}_k$ is the histogram (1) of all voxels within a range interval.

To ensure rotation invariance, the orientation of the point cloud is first normalised by rotating so that the most common plane normal is aligned with the z axis and the second most common lies in the yz plane. There may not be a single unambiguous maximum, but it is possible to use two *sets* of directions, $\mathcal{D}_1$ and $\mathcal{D}_2$. Given an ambiguity threshold $t_a \in [0, 1]$ that determines which orientations are "similar enough," a set of dominant directions can be selected [8]. If there are several plane orientations that are equally common, multiple histograms are generated, one for each potential alignment. The outcome is a set of histograms $\mathcal{F} = \{\boldsymbol{F}_1, \ldots, \boldsymbol{F}_N\}$. This set is the appearance descriptor of the point cloud. The parameters are summarised in Tab. II.

B. Distance measure

To quantify the difference between two appearance matrices $\boldsymbol{F}$ and $\boldsymbol{G}$, we use the following function [8]:

$$\delta(\boldsymbol{F}, \boldsymbol{G}) = \sum_{i=1}^{n_r} \left(\left\| \frac{\mathbf{f}_i}{\|\boldsymbol{F}\|_1} - \frac{\mathbf{g}_i}{\|\boldsymbol{G}\|_1} \right\|_2 \right) \frac{\max(\|\boldsymbol{F}\|_1, \|\boldsymbol{G}\|_1)}{\min(\|\boldsymbol{F}\|_1, \|\boldsymbol{G}\|_1)}. \quad (2)$$

In other words, δ is the sum of Euclidean distances for each column (each column corresponds to one range interval). The right-most normalisation factor in (2) makes it possible to use a single threshold for data sets that both contain point clouds that cover a large area (with many occupied voxels) and scans of more confined spaces.

As said, multiple histograms may be generated for one point cloud when $t_a < 1$, to achieve rotation invariance. Given a point cloud pair $\mathcal{X}_1$ and $\mathcal{X}_2$ with histogram sets $\mathcal{F}$ and $\mathcal{G}$, all members of the sets are compared using δ (2), and the minimum δ is used as the difference measure for the pair.

$$\Delta(\mathcal{X}_1, \mathcal{X}_2) = \min_{i,j} \delta(\boldsymbol{F}_i, \boldsymbol{G}_j) \qquad \boldsymbol{F}_i \in \mathcal{F},\ \boldsymbol{G}_j \in \mathcal{G} \quad (3)$$

IV. CLUSTERING METHODS

A. k-means

A classic clustering algorithm is k-means, which works as follows. A set of k seed cluster centres is drawn randomly from the data points, after which the algorithm iteratively assigns points to the currently closest centre until convergence.

One weakness of k-means is that it prefers evenly-sized clusters, so it may be difficult to generate good clusters for data where some categories are over-represented. Furthermore, k-means is sensitive to the initial seeding of centres.

For some of the experiments in Sec. V, we use an alternative, deterministic, k-means seeding strategy, such that the seeds are selected at equal strides from the list of point clouds to be categorised. I. e., for n point clouds and k clusters, cluster i is initialised with the $i(n/k)$-th point cloud.

The k-means++ algorithm [1] is meant to address k-means' seeding sensitivity. After randomly selecting the first seed, k-means++ selects the remaining $k-1$ seeds by iterative sampling data, where each point is weighted by the squared distance to the closest seed so far. This strategy distributes the initial cluster centres more evenly in appearance space.

B. Hierarchical k-means++

It is also possible to use k-means hierarchically: starting with $k = 2$ clusters, and splitting clusters until all points in a cluster are "similar enough." This hierarchical scheme is particularly interesting when using clustering for place categorisation, since related sub-categories can be grouped.

We have used k-means++ with $k = 2$ clusters at each branch, terminating when the max difference between any two point clouds in the cluster is below a user-specified threshold (see Fig. 1b).

C. DBSCAN

For the DBSCAN algorithm [3], the user specifies a density threshold ε and the minimum number c of data points required to form a cluster, instead of the desired number of clusters (as for k-means) or the max allowed distance (as for hierarchical k-means). Initialised with an arbitrary seed, the neighbourhood (with radius ε) is retrieved, and if it contains more than c points, a cluster is started. Otherwise, the point is labelled as noise. One weakness of DBSCAN is that it is not generally suitable for data sets with large differences in densities, using a single ε and c parameter pair.

V. EXPERIMENTS AND RESULTS

A. Overview of results

This section details our quantitative and qualitative results of applying the methods from Sec. III and IV to three data sets: the office-like benchmark set used in Mozos et al. [10] (KyushuIndoor), one from a warehouse (ArlaWarehouse), and one from an outdoor field robot (EskilstunaField).

Our main quantitative results use the KyushuIndoor data and are presented in Sec. V-B. KyushuIndoor is the only relevant data set for 3D place categorisation from the literature that we are aware of. By comparing our results from clustering of NDT histogram descriptors to two baselines

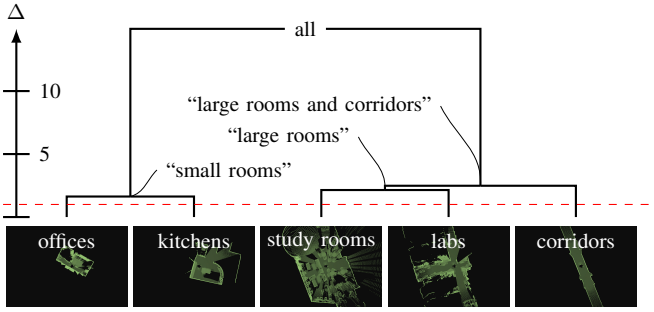


Fig. 2: Dendrogram of hierarchical k-means++ for KyushuIndoor. The termination threshold for sub-category splitting is marked with a dashed line. Max distances (3) between clusters are on the horizontal axis. Leaf nodes show the average scan found for the category.

– the state-of-the-art 3D SVM classifier of Mozos et al. [10] and clustering with a 2D descriptor used in previous works [12] – we show that k-means clustering of the NDT appearance descriptor attains high accuracy without training. By employing hierarchical k-means++, we have the added benefit of a semantically meaningful sub-categorisation of places. Using DBSCAN instead of k-means is also possible but only works well for a very narrow set of parameters.

The two other data sets (see Sec. V-C and V-D) are both collected in sequence from a mobile robot, and are used to demonstrate the usability of the proposed approach on less artificial data than the KyushuIndoor benchmark. The results from these data sets show that DBSCAN is useful for data sets without clear boundaries between places, because it allows for assigning data from transition regions to an outlier class, instead of forcing them into a category. However, DBSCAN needs carefully selected parameters, which makes it unsuitable for semi- or un-supervised categorisation.

A parameter sensitivity analysis of the NDT histogram descriptor is provided in Sec. V-E.

B. KyushuIndoor

The Kyushu University Indoor Semantic Place data set[1] [10] consists of data for five categories: corridor, kitchen, laboratory, study room and office. The data were collected with a rotating laser scanner, and the point clouds are very dense: about 2.8 million points each. The field of view is $270° \times 360°$. The data set also has reflectance images, not used in our work. In contrast to the other two data sets, this was not collected in sequence from a mobile robot. We include this data set both to demonstrate the accuracy of unsupervised clustering using NDT histograms, compared to the supervised learning and features used by Mozos et al. [10], and to demonstrate our approach in an office environment, which is relevant for many types of service robots.

1) Using k-means++ : Tab. Ia summarises the quantitative results of k-means++ clustering. Since k-means++ uses stochastic seeding, the results presented here are the mean and standard deviation of 10 runs. The overall accuracy is 88.0% $\pm$ 5.6, which can be compared to the 95.6% accuracy of the

[1] http://robotics.ait.kyushu-u.ac.jp/kurazume_lab/research-e.php?content=db#d05

state-of-the-art SVM-based classifier trained with both 3D range and reflectance images of Mozos et al. [10]. We believe that this is a strong result, which shows that even with rather simple unsupervised clustering, the NDT histogram descriptor is capable of performance close to that of a manually trained classifier — and in some cases surpasses it.

The offices and corridors are generally the easiest types (98% and 95% accuracy), while the kitchens are more difficult (73%). The main difficulty of the kitchen class in our case is that in some cases, a corridor outside the kitchen is also visible, which causes some of the kitchen point clouds to be clustered with the corridor class. If the max range for the NDT histograms is set to 9 m (instead of the baseline parameters from Tab. II) it is possible to reach much higher accuracy for the kitchens, at the cost of less discrimination between the larger rooms.

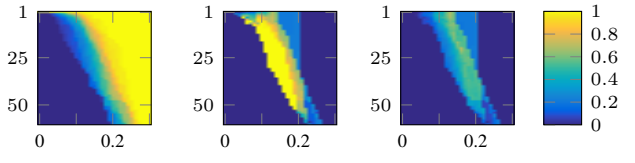
In comparison, the SVM-based system [10] reaches 100% accuracy for labs, study rooms, and corridors, and has almost identical accuracy for offices. The kitchens are challenging for the SVM-based system too, and the accuracy has a large variance: 80% $\pm$ 42, compared to our result 73.6% $\pm$ 1.1.

Fig. 2 shows a qualitative result of running hierarchical k-means++ for this data set. As can be seen from the dendrogram, the method finds the same categories as the manual labelling. It differentiates between the smaller rooms (offices and kitchens) and the larger ones. From the larger types, it finds that corridors are different from rooms, and further splits the large rooms into labs and study rooms.

Since hierarchical k-means++ uses stochastic seeding, also the topology of the tree is stochastic. Fig. 2 shows the most common tree structure. This topology is generated in 6 of the 10 cases. In three cases, the large and small rooms are still well separated, but "corridors" has the same parent as labs or study rooms. In one case, corridors are split between two leaf classes, which does not make sense semantically. This case is also the one that mainly decreases the mean accuracy (and increases the variance) for the corridor class in Tab. Ia.

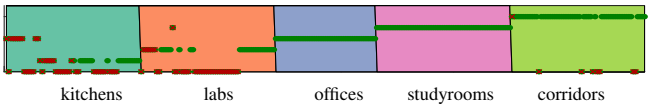
2) Using deterministic k-means: In an effort to demonstrate the descriptive power of the NDT histogram descriptor independently of clustering performance, we have also computed the confusion matrix for a deterministic k-means clustering initialised with one seed per category (see Tab. Ib). With these settings, we achieve 100% accuracy for offices and corridors, and 95% for kitchens and studyrooms. The most challenging type for this approach is "labs", which sometimes are assigned to the same category as kitchens (6/60 point clouds) or studyrooms (3/60). The overall accuracy is 93.3%, which is very similar to the 95.6% of the SVM-based classifier [10]. Comparing per-class accuracies, this method performs better than SVM [10] for kitchens (95% vs 80%) and offices (100% vs 98%), but worse for labs (78% vs 100%) and studyrooms (95% vs 100%).

3) Using DBSCAN: We have also investigated how DBSCAN performs for place categorisation with NDT histograms. The advantage of DBSCAN over k-means is that its parameters quantify the expected cluster properties (density ε and min point count c) rather than a fixed cluster number, and that it allows for an "outlier" class.

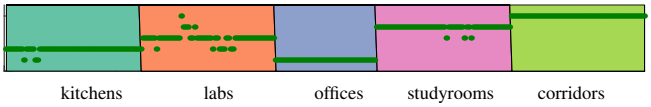


(a) Inlier ratio (b) AR (c) Score S

Fig. 3: Sensitivity analysis for DBSCAN on KyushuIndoor, with ε on the horizontal axis and c vertical. Good scores (bright/yellow) are obtained only for a very narrow band of parameters (Fig. 3c).



(a) DBSCAN. Red marks represent points labelled as outliers.



(b) Deterministic k-means.

Fig. 4: Clustering results for KyushuIndoor. The coloured fields represent ground-truth labels, and the green dots represents the result of clustering. An ideal clustering would show a continuous green line across each field.

However, selecting the parameters is challenging. Fig. 3 shows the result of a DBSCAN parameter search for KyushuIndoor. To evaluate the quality we have used the *adjusted Rand index* AR [6] to score the difference between two clusterings, and compute a score that also accounts for the outliers: $S = \mathrm{AR} \cdot \frac{\#\text{inliers}}{\#\text{all data points}}$. (Without this scaling, AR would consider a clustering that correctly labels one point cloud but discards the rest as outliers as perfect.)

Fig. 3a shows that the outlier ratio drops with increasing ε and c. However, Fig. 3b shows that AR is nonzero only in a small region. The best S is 0.7, with $\varepsilon = 0.16$ and $c = 20$. In comparison, the mean score for hierarchical k-means++ is 0.71 $\pm$ 0.06. Fig. 4 shows the result for this parameter set. Again, "kitchen" is the most difficult, but also for "lab" many point clouds are labelled as outliers. Offices and studyrooms have 100% accuracy, and corridors 91.7%.

4) 2D baseline: For comparison, we also extracted 2D grid maps from the 3D point clouds of KyushuIndoor in order to do the same k-means classification with the 2D feature descriptor of Mozos et al. [12]. The results are quite poor compared to categorisation with 3D data. Accuracy is 100% for offices and 93% for corridors, but only 30% for kitchens, 17% for studyrooms, and 33% for labs. For space reasons, we omit the confusion matrix.

C. ArlaWarehouse

The ArlaWarehouse data set was collected with a Velodyne HDL-32E lidar mounted on an AGV. The environment is a warehouse storing dairy products. The AGV moves between two adjacent halls, transitioning through an airlock with an automatic door. We include this data set to demonstrate our

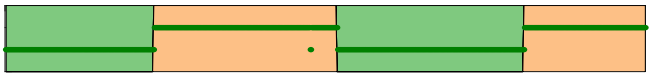
TABLE I: Confusion matrices for KyushuIndoor using k-means.

(a) Mean classification rates $\pm$ one standard deviation from 10 random trials, using hierarchical k-means++. The overall accuracy is 88.0% $\pm$ 5.6%.

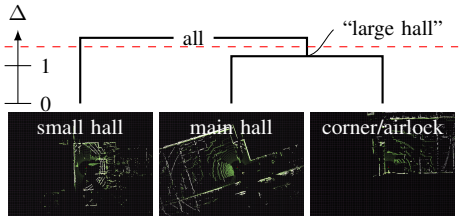
	offices	kitchens	labs	studyrooms	corridors
offices	**98.0±0.7**	2.0±0.7	0.0±0.0	0.0±0.0	0.0±0.0
kitchens	6.7±5.3	**73.6±1.1**	0.0±0.0	0.0±0.0	19.7±6.3
labs	0.0±0.0	0.2±0.5	**85.2±8.3**	6.7±3.3	8.0±9.2
studyrooms	0.0±0.0	0.0±0.0	7.2±3.9	**90.0±4.3**	2.8±2.7
corridors	0.0±0.0	4.0±12.06	0.3±0.7	0.0±0.0	**95.7±12.5**

(b) Classification rates using a deterministic k-means variant. The overall accuracy is 93.3%.

offices	kitchens	labs	studyrooms	corridors
100.00	0.00	0.00	0.00	0.00
5.00	**95.00**	0.00	0.00	0.00
0.00	13.33	**78.33**	6.67	1.67
0.00	0.00	5.00	**95.00**	0.00
0.00	0.00	0.00	0.00	**100.00**



(a) ArlaWarehouse clustering results with Δ threshold 1.5, creating two classes. The two main halls are clearly separated, and only a single scan is mislabelled.



(b) Dendrogram for ArlaWarehouse. With a Δ threshold less than 1, the "large hall" class splits in two, where the extra class corresponds to a tight passage in the airlock and a corner of the large hall.

Fig. 5: ArlaWarehouse results using hierarchical k-means++.

approach in a structured indoor environment that is markedly different from a standard office environment.

Fig. 5 shows results using hierarchical k-means++. It is more difficult to quantitative assess these results than for KyushuIndoor, since this data set includes transitions between two places, where both are visible. However, with a threshold set for generating two categories as in Fig. 5a, only 1/538 point clouds is mislabelled, which means an accuracy of 99.8%. With a lower Δ threshold (for the three categories at the bottom of Fig. 5b), 6/538 point clouds are clearly mislabelled (accuracy 98.8%), but the third category is not a single location, but rather corresponds to "tight passages".

D. EskilstunaField

The EskilstunaField data set[2] (Fig. 1a) was collected with a Velodyne HDL-64E mounted on a wheel loader. The environment is a test site for Volvo Construction Equipment in Eskilstuna, Sweden, and contains both open areas and forest. In particular, there are three main regions. One is the open gravel plain (bottom left of the figure), one is a gravel road going through the forest, and one is a narrow (bumpy) smaller forest path. The point clouds contain approx. 100 k points each, and have a 26.5° $\times$ 270° field of view. We include this data set to demonstrate performance in unstructured outdoor environments. This is in contrast to existing literature on

[2]The data set is available from http://mro.oru.se.

Fig. 6: DBSCAN results for EskilstunaField. Most of the transition areas between the more distinct zones have been labelled as outliers (red) instead of forcing an uncertain label.

place categorisation, which has always been demonstrated in structured, indoor, environments.

The results can be assessed qualitatively from Fig. 1, where a threshold resulting in three categories is displayed. Again, this environment has no sharp transitions between places, so it is not possible to compute a meaningful confusion matrix.

Examining Fig. 1a, some short sequences in the "forest path" area (blue points) are labelled as "forest road" (magenta). This mislabelling can be explained by that the road or other more open areas can also be seen from these poses. It may be possible to overcome this issue by selecting a shorter max cut-off range for the appearance descriptor. Also, in the last part of the forest area (the bend near the top of the image), the selected category alternates between "road" and "path". From the map it can be seen that this segment is indeed a wider path, somewhat between a "road" and a "path". There is also a sequence of 6 "road" point clouds mislabelled as "path" near the beginning of the main "road" segment.

Compared to a manual labelling (see 6), 30/1220 point clouds in this data set were mislabelled by k-means++, which means that the overall accuracy was 97.5%.

An interesting feature of DBSCAN, for this data set in particular, is that it avoids the misclassifications from Fig. 1a by labelling uncertain point clouds, and those in transition regions, as outliers rather than forcing them into any category. Fig. 6 shows the DBSCAN result, again with parameters selected from an exhaustive search.

E. Parameter selection for the appearance descriptor

A number of parameters govern the NDT appearance descriptor. For the results presented above, we have used parameters similar to those used in previous work for detecting loop closures [8]; see Tab. II. We have used slightly different near-range and far-range cut-off distances for the indoor and outdoor data. The minimum range should be such that the robot is not seen in the point clouds, but otherwise include as much of the scene as possible. The selection of the far-range cut-off distance relates to the question of "what is a place." Especially with outdoor lidar data, areas that are

TABLE II: Parameters of the appearance descriptor.

spherical classes	n_s	1	ambiguity thresh.	t_a	0.6
planar classes	n_p	9	e-value thresh.	t_e	0.05
linear classes	n_l	1	voxel size	B	0.4 m
ranges, indoor	$\{[1,4),[3,6),[5,8),[7,10),[9,12),[11,15)[14,\infty)\}$				
ranges, outdoor	$\{[5,8),[7,10),[9,12),[11,15),[14,20)\}$				

TABLE III: Parameter sensitivity. The table shows overall accuracy for the KyushuIndoor data set,[3] using deterministic k-means with $k = 5$. One parameter was changed for each run, and the remaining parameters were taken from the baseline shown in Tab. II.

e-value thresh. t_e	0.05 93.3%	0.10 91.9%	0.20 91.6%	0.40 88.7%
ambiguity thresh. t_a	0.4 93.7%	0.6 93.3%	0.8 94.0%	1.0 94.3%
voxel size B	0.2 m 93.0%	0.4 m 93.3%	0.8 m 91.2%	1.6 m 86.7%
max range (1–4, 3–6, etc)	10 m 81.7%	17 m 93.0%	∞ m 93.3%	
spherical classes	0 89.1%	1 93.3%	3 92.6%	9 86.7%

far from the robot position also influence the appearance descriptor, which may not be desired. Nevertheless, the NDT histogram appearance descriptor is remarkably robust to changing the parameters listed in Tab. II. A sensitivity analysis is provided in Tab. III. To measure only the influence of the appearance descriptor's parameters, and not the clustering algorithm, Tab. III was computed using the deterministic k-means seeding.

VI. SUMMARY AND CONCLUSIONS

The main contribution of this paper is to demonstrate that by leveraging the NDT histogram descriptor, and appropriate clustering techniques, place categorisation with 3D data from both structured and unstructured environments can be largely solved entirely without training. In particular, on a standard data set, we achieve 88.0% mean accuracy with stochastic k-means++, and 93.3% with a deterministic k-meansusing 3D only, compared to 95.6% for a state-of-the-art SVM-based classifier trained with both 3D and reflectance images. We take this to be a strong result, given that our approach requires no training.

This is also, to the best of our knowledge, the first paper to demonstrate and validate place categorisation with 3D data from unstructured environments.

Based on our results, we propose to use NDT histograms together with hierarchical k-means++ in general for place categorisation. We have also shown that DBSCAN can categorise NDT histograms with high accuracy, but only for a narrow band of parameters. On the other hand, the hierarchical k-means++ implementation lends itself well to a user-selected (semi-supervised) threshold selection based on the desired level of categories.

[3]Please note that for performance reasons these numbers were computed on a subsampled version of the KyushuIndoor data set (2.5 cm voxel grid). This modified data set is available on request.

We have shown that the NDT histogram appearance descriptor is remarkably robust w. r. t. parameter selection. Only a very large t_e or voxel size B, or cropping the scans has a significant effect on the overall accuracy.

Future work should include other global descriptors for 3D data (e. g., IRON [15]) and alternative (hierarchical) clustering methods. It would also be relevant to work on an on-line implementation and domain-specific seeding, also taking into account the sequential nature of data from a moving robot.

ACKNOWLEDGEMENT

Thanks to Óscar Martínez Mozos et al. for providing the KyushuIndoor data set, and assisting with conversion routines and discussions on the nature of the different sub-classes.

REFERENCES

[1] D. Arthur and S. Vassilvitskii. "k-means++: the advantages of careful seeding". In: *Proc. ACM-SIAM Symposium on Discrete Algorithms*. 2007, pp. 1027–1035.

[2] P. Biber and W. Straßer. "The Normal Distributions Transform: A New Approach to Laser Scan Matching". In: *IROS*. 2003, pp. 2743–2748.

[3] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. "A density-based algorithm for discovering clusters in large spatial databases with noise". In: *Proc. KDD*. 1996, pp. 226–231.

[4] S. Gholami Shahbandi, B. Åstrand, and R. Philippsen. "Semi-Supervised Semantic Labeling of Adaptive Cell Decomposition Maps in Well-Structured Environments". In: *ECMR*. 2015.

[5] R. Goeddel and E. Olson. "Learning Semantic Place Labels from Occupancy Grids using CNNs". In: *IROS*. 2016.

[6] L. Hubert and P. Arabie. "Comparing partitions". In: *Journal of Classification* 2.1 (1985), pp. 193–218.

[7] M. Magnusson. "The Three-Dimensional Normal-Distributions Transform — an Efficient Representation for Registration, Surface Analysis, and Loop Detection". PhD thesis. Örebro University, 2009.

[8] M. Magnusson, H. Andreasson, A. Nüchter, and A. J. Lilienthal. "Automatic Appearance-Based Loop Detection from 3D Laser Data Using the Normal Distributions Transform". In: *J. Field Robotics* 26.11–12 (2009), pp. 892–914.

[9] M. Magnusson, N. Vaskevicius, T. Stoyanov, K. Pathak, and A. Birk. "Beyond Points: Evaluating Recent 3D Scan-Matching Algorithms". In: *ICRA*. 2015, pp. 3631–3637.

[10] O. M. Mozos, H. Mizutani, H. Jung, R. Kurazume, and T. Hasegawa. "Categorization of indoor places by combining local binary pattern histograms of range and reflectance data from laser range finders". In: *Advanced Robotics* 27.18 (2013), pp. 1455–1464.

[11] O. M. Mozos, H. Mizutani, R. Kurazume, and T. Hasegawa. "Categorization of Indoor Places Using the Kinect Sensor". In: *Sensors* 12.5 (2012), p. 6695.

[12] O. M. Mozos, C. Stachniss, and W. Burgard. "Supervised Learning of Places from Range Data using AdaBoost". In: *ICRA*. 2005, pp. 1730–1735.

[13] A. Pronobis, O. M. Mozos, B. Caputo, and P. Jensfelt. "Multi-modal Semantic Place Classification". In: *IJRR* (2010).

[14] A. Ranganathan. "PLISS: Detecting and Labeling Places Using Online Change-Point Detection". In: *RSS*. 2010.

[15] T. Schmiedel, E. Einhorn, and H.-M. Gross. "IRON: A Fast Interest Point Descriptor for Robust NDT-Map Matching and its Application to Robot Localization". In: *IROS*. 2015.

[16] N. Sünderhauf et al. "Place categorization and semantic mapping on a mobile robot". In: *ICRA*. 2016.