

CorAI – Are the point clouds Correctly Aligned?

Daniel Adolfsson, Martin Magnusson, Qianfang Liao, Achim J. Lilienthal, Henrik Andreasson

Abstract—In robotics perception, numerous tasks rely on point cloud registration. However, currently there is no method that can automatically detect misaligned point clouds reliably and without environment-specific parameters. We propose “CorAI”, an alignment quality measure and alignment classifier for point cloud pairs, which facilitates the ability to introspectively assess the performance of registration. CorAI compares the joint and the separate entropy of the two point clouds. The separate entropy provides a measure of the entropy that can be expected to be inherent to the environment. The joint entropy should therefore not be substantially higher if the point clouds are properly aligned. Computing the expected entropy makes the method sensitive also to small alignment errors, which are particularly hard to detect, and applicable in a range of different environments. We found that CorAI is able to detect small alignment errors in previously unseen environments with an accuracy of 95% and achieve a substantial improvement to previous methods.

I. INTRODUCTION

In order to create safe and efficient mobile robots, introspective and reliability-aware capabilities are required to assess and recover from perception failures. Many perception tasks, including localization [1], scene understanding and sensor calibration [2], rely on point cloud registration. However, registration may provide incorrect estimates due to local minima of the registration cost function [3], uncompensated motion distortion [4], noise or when the registration problem is geometrically under-constrained [5], [6]. Consequently, it is essential to measure alignment quality and to reject or re-estimate alignment when quality is low. In the past, an extensive number of methods have been proposed to assess the alignment quality of point cloud pairs [7]–[17]. These metrics can typically be used to measure a relative alignment error in the process of registration, but provide limited information on whether the point clouds are correctly aligned once registration has been carried out [18]. Until today, few studies have targeted the measurement of alignment correctness after registration [18], [19] and previous works report that alignment correctness classification based on AdaBoost and NDT score function decrease when applied to point clouds acquired from new environments [19].

In this paper, we propose “CorAI” (Correctly Aligned?): A method to introspectively measure and detect misalignment between previously registered point cloud pairs. CorAI

The authors are with the MRO lab of the AASS research centre at Örebro University, Sweden. E-mail: Daniel.Adolfsson@oru.se

This work has received funding from the Swedish Knowledge Foundation (KKS) project “Semantic Robots” and European Union’s Horizon 2020 research and innovation programme under grant agreement No 732737 (ILIAD) and 101017274 (DARKO).

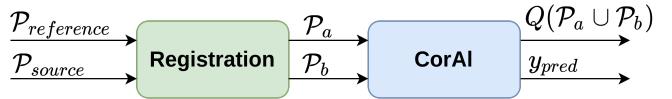
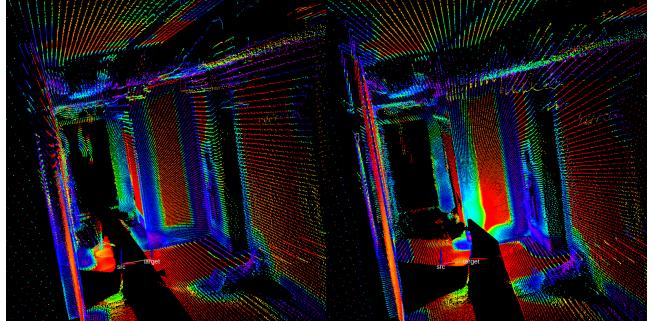
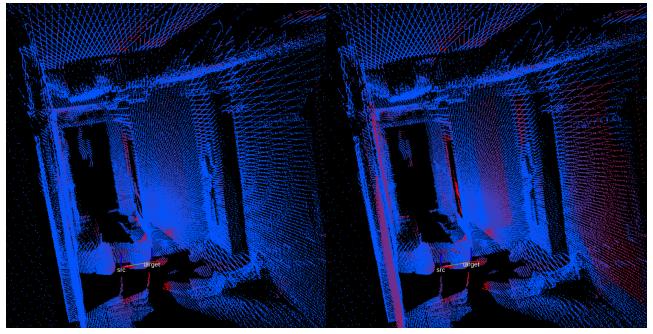


Fig. 1: CorAI depicted in blue, operates on a pair of point clouds $\mathcal{P}_a, \mathcal{P}_b$ and can classify misalignment (y_{pred}) by comparing the differential entropy in the point clouds separately and jointly. Additionally, CorAI outputs a per-point quality measure $Q(\mathcal{P}_a \cup \mathcal{P}_b)$ that highlights misaligned parts.



(a) \mathcal{P}_a colored by entropy. (b) \mathcal{P}_b colored by entropy.



(c) Correctly aligned $\mathcal{P}_a \cup \mathcal{P}_b$ colored by quality measure. (d) Misaligned $\mathcal{P}_a \cup \mathcal{P}_b$ colored by quality measure.

Fig. 2: Top: Differential entropy in point clouds separately. Bottom: The joint point cloud ($\mathcal{P}_a \cup \mathcal{P}_b$) colored by per-point quality measure $q_k(\mathcal{P}_a, \mathcal{P}_b)$ when aligned (c) and when misaligned (d). Blue and red indicate alignment and misalignment respectively.

specifically aims to bridge the gap in classification performance when applied to new unseen environments.

Our method is well grounded in information theory and gives an intuitive alignment correctness measure. CorAI measures the difference between the average differential entropy in the joint and separate point clouds. For well-aligned point clouds, the joint and the separate point clouds have similar entropy. In contrast, misaligned point clouds tend to “blur” the scene which can be measured as an increase in joint entropy as depicted in fig. 3. By using the separate point clouds to estimate the entropy inherent in the scene, our proposed method can assess quality in a range of different environments.

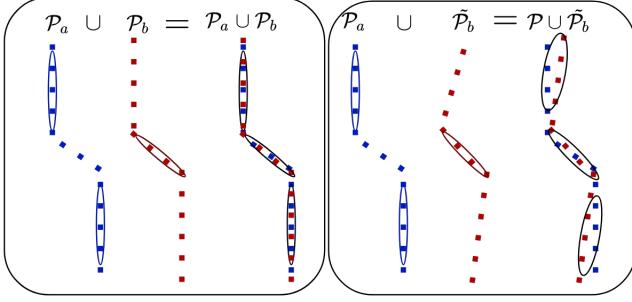


Fig. 3: Example how uncertainty (entropy) is preserved when joining aligned $\mathcal{P}_a \cup \mathcal{P}_b$ (left), but increases when joining misaligned point clouds (right). The entropy for aligned point clouds should be similar to the entropy in the separate point clouds and can be used when quantifying alignment quality.

The contribution of this paper is an intuitive and simple measure of alignment correctness between point cloud pairs. We demonstrate how to use this quality measure to train a simple model that detects small alignment errors between point clouds, large errors are not considered in this paper. To train our model, we use previously corrected scans that are assumed to have no alignment error.

We make the following claims: (i) Our proposed method CorAl measures the correctness of point cloud alignment by accounting for the expected scene entropy. (ii) Our method is accurate in a wide range of environments and can generalize well to new environments without retraining.

II. RELATED WORK

Several methods have been used to assess alignment quality in the literature. However, in most cases these methods are used in an ad-hoc manner. Few systematic evaluations of their general ability to be used as a classifier to detect aligned vs. misaligned point clouds have been made.

One well-used alignment measure is the root-mean-squared (RMS) point-to-point distance, truncated by some outlier rejection threshold. This is also the function that is minimized by iterative closest point registration [7]. However, this measure has been shown to be highly sensitive to the environment and the choice of the outlier threshold [19], [20]. Consequently, this is a poor measure for alignment correctness classification.

One family of methods instead attempts to estimate the alignment uncertainty between point cloud pairs in the form of a covariance matrix [21]–[24]. Some use Monte Carlo strategies to estimate uncertainty by sampling registrations in a region [22]. This exhaustive search is unpractical in mobile robotics. Others attempt to estimate uncertainty in closed form using the Hessian [19], [25], [26], representing the steepness of the alignment score function around a minimum. These methods assume that the registration has reached a global maximum, which is not necessarily true. Until today, alignment classification based on uncertainty covariance have been less accurate compared to matching score [19].

Almqvist et al. [19] explored alignment classifiers based on RMS as well as other existing methods [16], [20], [26]–[29], including the NDT score function [26], and investigated how to combine the measures with AdaBoost into a stronger

classifier. The classifiers were evaluated on two outdoor data sets, and although their classifiers reached almost 90 % accuracy for the hardest cases on each data set individually, accuracy drops to around 80 % when cross-evaluating between the data sets. In their evaluations, the NDT score function proved to be the best individual measure for alignment assessment. The combined AdaBoost classifier did not have significantly higher accuracy, but reduced parameter sensitivity.

Liao et al. [11] recently proposed a registration method based on fuzzy clusters, which involves a registration quality assessment. This fuzzy cluster-based quality assessment (FuzzyQA) compares the similarity of dispersion and disposition of points around fuzzy cluster centers. It has been used to detect if the point clouds are coarsely aligned, Coral instead attempts to detect small alignment errors.

Nobili et al. [5] proposed a method to predict alignment risk prior to registration by combining overlap information and an alignment metric. The alignment metric quantifies the geometric constraints in the registration problem. The alignment metric is based on point-to-plane residuals and has been evaluated in structured scenes with planar surfaces, while our method can operate well even in unstructured environments. Additionally, our method seeks to estimate the alignment after registration has been completed to introspectively measure the registration success, as opposed to predicting the risk prior to registration.

Bogoslavskyi et al. [18] defined a quality metric based on positive and negative point information, and used it to measure alignment error and cluster three known object types in a controlled experiment. Rather than focusing on objects, our method aims to classify alignment quality of observed scenes in different environments. Additionally, their method operates on range images, which might not be available, while our method operates on unorganized point clouds.

To the best of our knowledge, there is no method for binary point cloud alignment classification that performs accurately and transfer well to new environments without parameter tuning or retraining.

III. CORAL METHOD

Our work is inspired by the Mean-Map-Entropy (MME) measure proposed by Droschel and Behnke [30] for map quality assessment. MME is based on differential entropy [31] and measures the randomness of multivariate Gaussian distributions. Droschel and Behnke used MME in absence of accurate ground truth when evaluating map refinement. As shown in our evaluation, MME cannot be used as a general alignment quality measure as it is also affected by measurement noise, sample density and environment geometry. MME is more affected by changes in the environment compared to CorAl. Hence, the measure is not expected to generalize between, e.g., a structured warehouse and an unstructured outdoor forest environment. We overcome this effect using dual entropy measurements computed 1) in both point clouds separately and 2) in the joint point cloud. The intuition is that joining two well-aligned point clouds should not introduce additional uncertainty.

tainty and entropy should remain constant if the point clouds overlap sufficiently.

A. Computing joint and separate entropy

Our method operates on the dense point clouds \mathcal{P}_a , \mathcal{P}_b , given in a common fixed world frame, that contain a set of points in the Cartesian space $\mathbf{p}_k = [x \ y \ z]$. For later use, we define the joint point cloud $\mathcal{P}_j = \mathcal{P}_a \cup \mathcal{P}_b$; i.e., all points in \mathcal{P}_a and \mathcal{P}_b together.

From all points within a radius r around each point \mathbf{p}_k , we compute the sample covariance $\Sigma(\mathbf{p}_k)$. From the determinant of the sample covariance $\det(\Sigma(\mathbf{p}_k))$ we can then compute the differential entropy as:

$$h_i(\mathbf{p}_k) = \frac{1}{2} \ln(2\pi e \det(\Sigma(\mathbf{p}_k))) \quad (1)$$

for the point cloud $i = a, b$ that contains \mathbf{p}_k . An example of point clouds colored according to eq. (1) can be seen in Fig. 2a and 2b. The sum of differential entropy for \mathcal{P}_i can then be computed as

$$H_i(\mathcal{P}_i) = \sum_{k=1}^{|\mathcal{P}_i|} h_i(\mathbf{p}_k), \quad (2)$$

where $|\mathcal{P}_i|$ is the number of points in the point cloud \mathcal{P}_i .

Using Eq. 2 we can derive measures of the *separate* and *joint* average differential entropy of two point clouds \mathcal{P}_a , \mathcal{P}_b .

$$H_{\text{sep}} = \frac{H_a(\mathcal{P}_a) + H_b(\mathcal{P}_b)}{|\mathcal{P}_a| + |\mathcal{P}_b|}, \quad (3)$$

$$H_{\text{joint}} = \frac{H_j(\mathcal{P}_j)}{|\mathcal{P}_j|} = \frac{H(\mathcal{P}_a \cup \mathcal{P}_b)}{|\mathcal{P}_a| + |\mathcal{P}_b|}. \quad (4)$$

Our first alignment quality measure uses the difference between the joint and the separate average differential entropy:

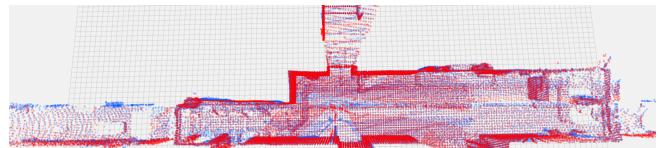
$$Q(\mathcal{P}_a, \mathcal{P}_b) = H_{\text{joint}}(\mathcal{P}_j) - H_{\text{sep}}(\mathcal{P}_a, \mathcal{P}_b), \quad (5)$$

which can also be given per-point by

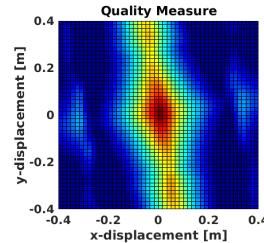
$$q_k(\mathbf{p}_k) = h_j(\mathbf{p}_k) - h_i(\mathbf{p}_k), \quad (6)$$

where the point entropy is evaluated on the joint point cloud j and the separate point cloud $i = (a \text{ or } b)$ where \mathbf{p}_k originates from. An example of point clouds colored by per-point entropy difference according to eq. (6) is depicted in Fig. 2c and 2d. Typically, $Q(\mathcal{P}_a, \mathcal{P}_b)$ is close to zero for well-aligned point clouds and increases with the alignment error as depicted in fig. 4, which visualizes the function's surface for position and angular alignment errors around the correct alignment.

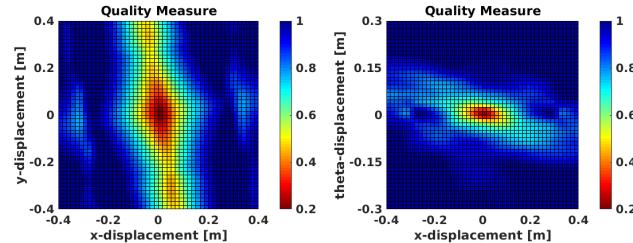
Well-aligned point clouds $\mathcal{P}_a \cup \mathcal{P}_b$ acquired in structured environments have low differential entropy for most query points \mathbf{p}_k . This is reflected by low values for the determinant of the sample covariance. As the determinant can be expressed as the product of the eigenvalues of the sample covariance $\det(\Sigma(\mathbf{p}_k)) = \lambda_1 \lambda_2 \lambda_3$, we see that the measure is sensitive to an increase in the lowest of the eigenvalues when larger eigenvalues are constant. For example, the entropy of points on a planar surface is represented with



(a) Top view of first two aligned point clouds \mathcal{P}_a (blue) and \mathcal{P}_b (red) in the ETH "stairs" dataset.

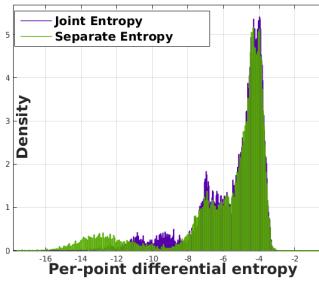


(b) CorAI measure $Q(\mathcal{P}_a, \mathcal{P}_b)$ eq. (5) visualized by color for various (x,y) displacements.

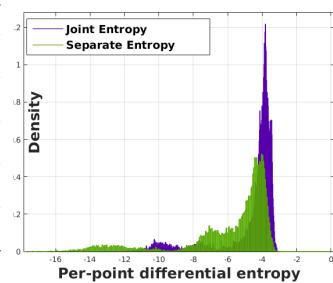


(c) CorAI measure $Q(\mathcal{P}_a, \mathcal{P}_b)$ for various (x, θ) displacements.

Fig. 4: Example of CorAI measure $Q(\mathcal{P}_a, \mathcal{P}_b)$ for various induced (x, y, θ) alignment errors. Q has a minimum at the true position. The steepness of the surface around the true alignment indicates that CorAI is sensitive to small misalignments.



(a) Aligned point clouds: differential entropy distributions are similar.



(b) Misaligned point clouds: joint differential entropy is higher.

Fig. 5: Probability distribution of per-point entropy eq. (1) for joint and separate point clouds when (a) aligned and (b) misaligned. Aligned point clouds have similar joint and separate entropy distributions, while joint entropy is higher compared to separate entropy when point clouds are misaligned. Joining misaligned point clouds blurs the scene which can be observed by an entropy increase. An exception can be seen in (a) region $(-12, -8)$ where the entropy increases by joining aligned point clouds.

a flat distribution with two large (λ_1, λ_2) and one small (λ_3) eigenvalue. Misalignment changes the point distribution in the joint point cloud from flat to ellipsoidal which can be observed as an increase of the smallest eigenvalue λ_3 . This makes the measure sensitive to misalignment of planar surfaces, but generalizes well to other geometries. As shown in the evaluation, the measure can capture discrepancies between point clouds regardless of whether these are due to rigid misalignments or distortions which can occur when scanning while moving, e.g. because of vibrations or sensor velocity estimation errors. That means that the method can be overly sensitive when used together with a registration method or odometry framework that does not compensate movement distortion or has a low accuracy.

Overlap is required between point clouds to produce evidence of alignment. For that reason, we classify point clouds with less than 10% overlap as misaligned. By defining the overlap as all points with a neighbor within r in the other

point cloud, non overlapping points have no effect on the quality measure in eq. (5).

B. Dynamic radius selection and outlier rejection

For well aligned point clouds, the quality measure is close to zero, meaning that the joint and separate point clouds have similar mean and probability distributions of per-point entropy as depicted in Fig. 5a. Unfortunately, the entropy in eq. (1) is ill-posed when the determinant $\det(\Sigma(\mathbf{p}_k))$ is close to zero and a small increase of the determinant causes a large increase of the entropy. Accordingly, the lowest measured entropies can increase (which indicates misalignment) even when joining well aligned point clouds as depicted in fig. 5a. The ill-posed entropies are found where point density is low, typically for solitary points or far from the sensor where the radius r is not large enough to include points that represent the geometry in the environment. The effect of the problem with entropies can be mitigated by maximizing the ratio $Q_s = Q_{\text{misaligned}}(\mathcal{P}_a, \mathcal{P}_b)/Q_{\text{aligned}}(\mathcal{P}_a, \mathcal{P}_b)$. A larger ratio indicates that the measure is able to discriminate between aligned and misaligned point clouds. We propose three strategies to address the ill-posed entropies due to variations in sampling density originating from the sensor.

(1): Eq. 1 is modified to $h_i(\mathbf{p}_k) = \frac{1}{2} \ln(2\pi e \det(\Sigma(\mathbf{p}_k)) + \epsilon)$ where ϵ limits the lowest possible entropy. This make sure that entropy is similar for points distributed along a line and a plane. The improvement can be seen by comparing fig. 6(a-b).

(2): Radius r is chosen based on the distance d between the point \mathbf{p}_k and the sensor location, to account for that point density decrease over distance. The radius is hence selected as: $r = d \sin(\alpha)$ in the range $r_{\min} < r < r_{\max}$ where α is the vertical resolution of the sensor. For other sensor types e.g. RGB-D, the resolution could be chosen similarly according to the angular sensing resolution. A dynamic radius enables the quality measure to include more points far from the sensor and correctly detect alignment and misalignment for these as seen in fig. 6(c).

(3): Remove E_{reject} percent of points \mathbf{p}_k with the lowest entropies. The effect is depicted in fig. 6(d).

C. Classification

We use logistic regression as a model for classification:

$$p = \frac{1}{1 + e^{-z}}, \quad (7)$$

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$y_{\text{pred}} = \begin{cases} \text{aligned} & \text{if } p \geq t_h \\ \text{misaligned} & \text{if } p < t_h, \end{cases} \quad (8)$$

where x_1, x_2 are input variables (described for each method in section IV-A). Instead of passing the quality measure $x_1 = Q(\mathcal{P}_a, \mathcal{P}_b)$, H_{sep} and H_{joint} are passed separately to $x_1 = H_{\text{joint}}$ and $x_2 = H_{\text{sep}}$. $\beta_0, \beta_1, \beta_2$ are learned model parameters, p is the class probability and t_h is a class probability threshold and can be adjusted to the application needs. For example, in mobile robotics, it is desired that

misaligned point clouds are not accidentally reported as aligned (false positives), potentially causing a system failure. In contrast, aligned point clouds classified as misaligned are typically harmless. For that reason, t_h can be increased to reject false positives and hence improve robustness. We used the default threshold $t_h = 0.5$.

IV. EVALUATION

We evaluate an equal portion of aligned and misaligned point clouds. Misaligned point clouds are created by adding an offset for each point cloud pair: an angular offset ($e_\theta = 0.57^\circ$) around the sensor's vertical axis and a random translational (x, y) offset at a distance ($e_d = 0.1\text{m}$) from the ground truth. These errors are large enough to be meaningful to detect in various environments, yet challenging to classify.

A. Evaluated methods

The evaluated methods are summarized here together with their most important parameters.

a) *MME*: Mean Map Entropy as proposed by Droseschel and Behnke [30] summarized in eq. (2). The parameter is the radius r for associating points.

b) *CorAl (proposed in the paper)*: Separate and registered entropy H_s, H_j as described in eqs. (3) and (4). Parameters are r_{\min}, r_{\max} and α to determine nearby points radius, and E_{reject} to set outlier rejection ratio and ϵ

c) *CorAl-Median (proposed in the paper)*: H_s, H_j are modified to calculate the median entropy rather than the mean entropy, we hypothesize that this modification can be more robust. The parameters are unchanged.

d) *NDT (point-to-distribution normal-distributions transform)*: The method uses the 3D NDT [32] representation similarly to Almqvist [19] (NDT3), which constructs a voxel grid over one point cloud, and computes a Gaussian function based on the points in each voxel. The likelihood of finding the points in \mathcal{P}_b , given the NDT representation of \mathcal{P}_a , is computed as

$$s = \frac{\sum_{k=1}^n \tilde{p}(\mathbf{p}_k)}{n}, \quad (9)$$

where n the number of overlapping points, defined as those points (which fall in an occupied NDT voxel, or in a voxel that is a direct neighbor of an occupied voxel) and \tilde{p} is the probability density function associated with the nearest overlapping NDT-cell. The most important parameter for NDT is the voxel size v which is set equal to $2 * r$ in our evaluation as this makes the sample covariance of NDT cells and entropy computed from points in a similarly large volume.

e) *Rel-NDT (proposed in the paper)*: We wanted to investigate if entropy can be used to improve generalization of NDT to different environments. The idea is that environment type is reflected in the average entropy of the scene and can be combined with NDT score to improve classification. We did this by computing the average entropy of all NDT-covariances associated with \mathbf{p}_k in the point-likelihood terms and feed that together with the NDT score (9) to the classifier. No additional parameters to NDT are required.

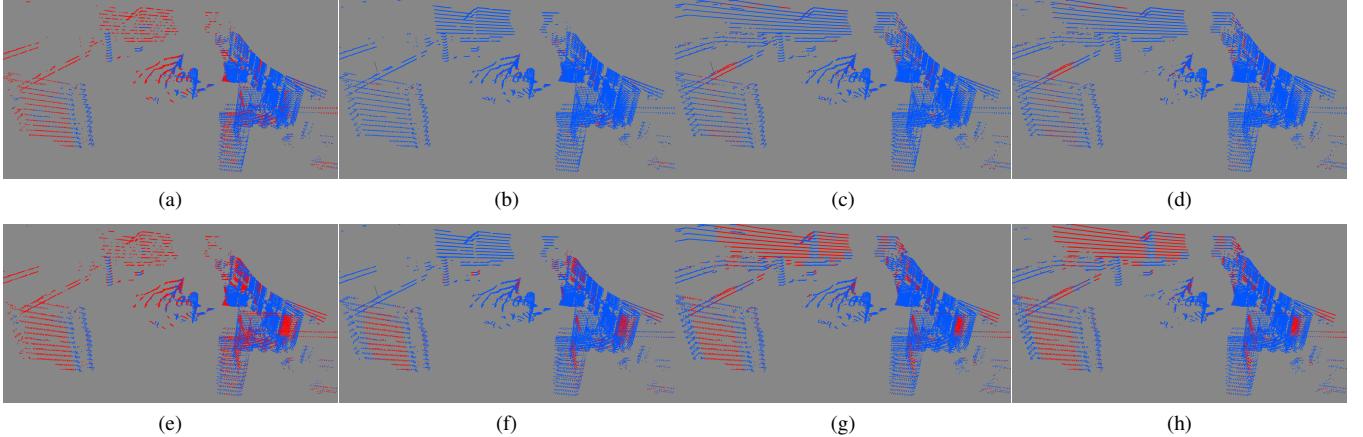


Fig. 6: Joint point cloud colored by per-point quality, q_k ranging from blue (aligned) to red (misaligned). The location of the point cloud origin is highlighted in fig. 7. Columns depict the same parameters for aligned (top) and misaligned (bottom) point clouds. (a,e): A fixed radius $r = 0.3m$ gives ($Q_s = 1.46$). (b,f): radius dynamically adjusted $r_{min} = 0.3m, \alpha = 1.33, r_{max} = 0.7$ gives ($Q_s = 2.93$). (c,g): $\epsilon = 10^{-8}$ is added ($Q_s = 4.06$). (d,h): $E_{reject} = 10\%$ is added ($Q_s = 4.3$).

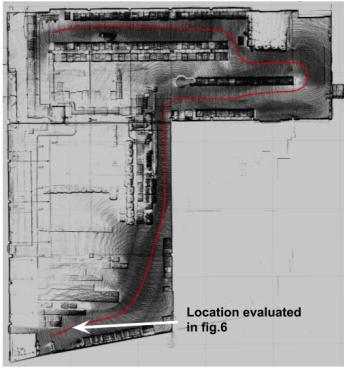


Fig. 7: Data acquired by a truck in a warehouse environment. The sensor trajectory is drawn in red. The environment in the figure is $50 m \times 50 m$ and the sequence length is 102 m. In the first segment of the trajectory, starting at the bottom left, the walls are clearly visible. The final segment is located between aisles where walls are typically out of sight and the sensor observes complex structures such as shelves. The truck traverses over a rough floor with height differences causing vibrations on the sensor.

f) FuzzyQA: FuzzyQA [11] measures the alignment quality by a ratio $\rho = \frac{\text{AFCCD}}{\text{AFPCD}}$, where AFCCD and AFPCD are two indexes describing the points' disposition and dispersion around fuzzy cluster centers. The two point clouds are coarsely aligned if $\rho < 1$. However, AFCCD and AFPCD are passed separately to the classifier input x_1, x_2 .

g) Input to the classifier: CorAl, FuzzyQA and Rel-NDT output two decision variables that are passed as input variables x_1, x_2 to the classifier (III-C). The other evaluated methods output a single variable x_1 , and $x_2 = 0$ is fixed.

B. Qualitative evaluation, live robot data

First, we present qualitative results from real-world data in a structured warehouse environment. A forklift equipped with a Velodyne HDL-32E spinning laser scanner was manually driven at fast walking speed in the environment depicted in fig. 7. The environment in the sequence varies from large and open with visible walls, to small and narrow between aisles of pallets. To generate ground-truth alignments for the

warehouse dataset, we first aligned the point clouds using a scan-to-map approach [6]. We then inspected the alignment between subsequent scans and found that at least 40/484 (8.3%) point clouds were impaired by rigid misalignments or non-rigid distortions from vibrations and motion to the extent that these could be easily visually located.

Alignment classification was then performed on the remaining scans by inducing errors as described in section IV. We used the following parameters as they provided a relatively high value of Q_s for the first scan pair in the dataset: $\alpha = 0.92^\circ$, $E_{reject} = 0.2$, $r_{min} = 0.2$, $r_{max} = 1.0$ and voxel size $v = 2r_{min} = 0.4$. We found that CorAl-mean, MME and NDT reached an accuracy of 96%, 70% and 99% respectively. In this case, NDT performs slightly better than CorAl. We believe that CorAl is more sensitive to the typical alignment noise that is still present in the aligned scans. This typical alignment noise introduces a variance in the CorAl score and makes it hard to train a classifier that is sensitive to small misalignment's. Whether this is desired behavior depends on the application.

C. Quantitative evaluation, ETH benchmark data set

Our main quantitative evaluation is done using the public ETH registration dataset [33]. This dataset includes 3 sequences in structured (blue) environments (Apartments, ETH Hauptgebaude, Stairs), 3 sequences in semi-structured (brown) environments (Gazebo in summer, Gazebo in winter, Mountain plain) and 2 challenging sequences in unstructured (green) environments (Wood in summer, Wood in autumn). Each sequence contains between 31 and 47 scans acquired from stationary positions. The dataset contains accurate ground truth positions, required to evaluate the different methods. In order to make the evaluation fairer, more realistic and applicable to real applications, we downsample the original, dense, point clouds using a voxel grid of 0.08 m. As the dataset has less variation in sampling density compared to the warehouse dataset, we used a fixed radius $r = 0.3$ and

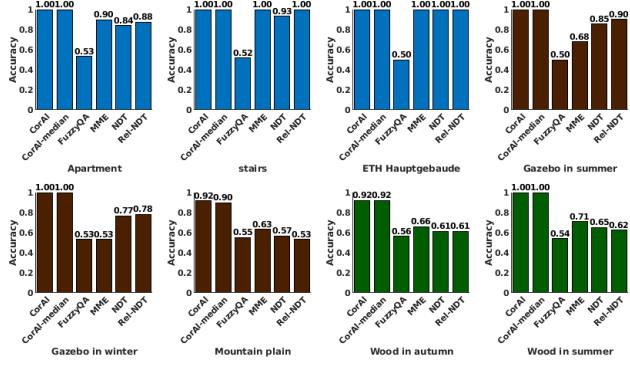


Fig. 8: Separate training. The overall accuracy was CorAl: 98%, CorAl-median: 98% FuzzyQA: 53% MME: 77% NDT: 78% Rel-NDT: 80%

set $E_{\text{reject}} = 20\%$, $\epsilon = 0$. NDT voxel size was set equal to the diameter $v = 2r = 0.6$ to create a fair comparison.

a) *Performance:* CorAl has an overall run-time of 0.246 ± 0.095 seconds per point cloud pair on an Intel Core i7 and depends on the point cloud density.

1) *Separate training:* The first test evaluates the capability to learn classification in a specific type of environment and serves as a reference for further evaluations. The classifiers were trained and evaluated on each sequence separately, using 5-fold cross validation.

Results are shown in fig. 8. We found that all methods except FuzzyQA performed well on the structured environments. We did not expect that FuzzyQA would handle this as it is specifically designed to classify coarse alignment. Surprisingly, even MME scored 90–100% on the structured environment. This indicates that even naive methods can assess alignment quality in a highly structured environment. In the semi-structured and unstructured sequences, only CorAl and CorAl-median performed well, with consistently >90% accuracy, even in the most challenging sequences. All other methods are only slightly better than random, except for the gazebo sequences. Rel-NDT improves NDT in most cases, however not consistently. We believe this is because entropy alone provides little information about the environment. This is supported by the low overall accuracy of MME. Both NDT methods performed decently (77–90%) in the gazebo sequence, indicating that NDT requires at least some structure or surfaces free from foliage to be effective as an alignment correctness measure.

2) *Joint training:* The second test evaluates how the methods are able to learn alignment classification when trained in a variety of environments. To do that, the methods need to be versatile. Training was performed on all the ETH sequences, evaluation was then performed on each sequence individually. The results are shown in fig. 9. The accuracy of all classifiers decreased compared to the previous test. CorAl performed best, with accuracy 85–100% in all cases. CorAl-median reached a slightly lower accuracy compared to CorAl. Rel-NDT performed better than NDT in most cases, however not consistently. The generally high accuracy of CorAl indicates that it is possible to find general parameters that makes the method valid in various environments.

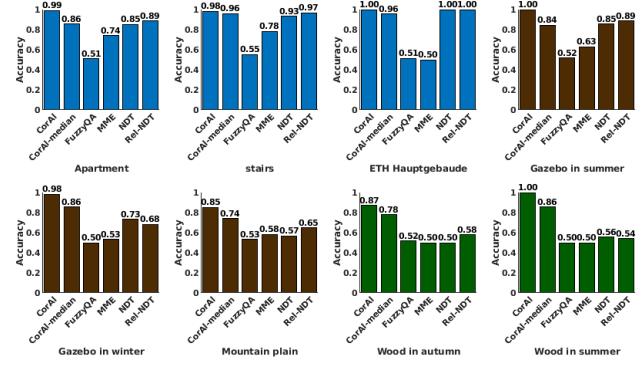


Fig. 9: Joint training. Overall accuracy CorAl: 96%, Coral-median: 95% FuzzyQA: 52% MME: 60% NDT: 75% Rel-NDT: 78%

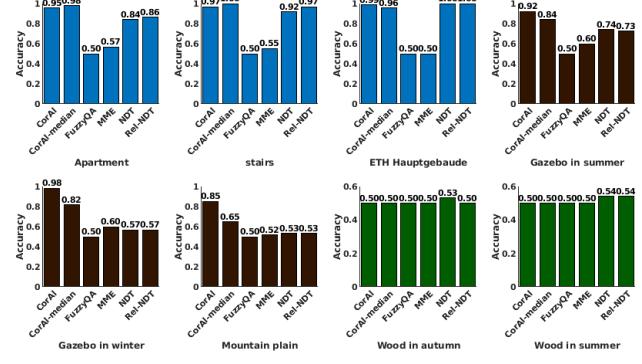


Fig. 10: Evaluation on unseen environments. Overall accuracy: 83%, CorAl-median: 79% FuzzyQA: 50% MME: 54% NDT: 72% Rel-NDT: 72%, In structured and semi-structured environments: 95%, Coral-median: 88% FuzzyQA: 50% MME: 56% NDT: 78% Rel-NDT: 79%.

3) *Generalization to unseen environments:* The final test evaluates how classifiers perform in environments with different characteristics than those observed in the training set. We trained and evaluated on different sequences and environments. The 3 structured environments were used for training and the remaining 5 (semi-structured and unstructured) were used for evaluation and vice versa. The classification accuracy is depicted in fig. 10. When trained on structured and evaluated on semi-structured environments, CorAl performed accurately(85–98%) and other methods performed close to random except NDT for Gazebo summer (74%). No method generalized well from structured to unstructured environments. On the other hand, learning from semi-structured and unstructured environments was enough to afford very high accuracy in structured environments with CorAl – very close to what was attained with joint training on all sequences. The previous joint evaluation show that it's possible to train a model that is simultaneously accurate in all environment types. For that reason, we believe that the reason the classifier trained in a structured environment does not generalize to an unstructured environment is that the model overfits when not using sufficiently diverse and challenging data.

V. CONCLUSIONS

In this paper we introduced CorAl, a principled and intuitive measure of alignment correctness between point

clouds. Using dual entropy measurement that compares the expected entropy found in the separate point clouds with the actual entropy, CorAI can measure point cloud alignment correctness and substantially outperforms previous methods when evaluated on a public data set. Specifically, we were able to use CorAI to train a classifier based on logistic regression that is simultaneously accurate in a diverse range of environments. Our experiments shows that our method generalizes well from (i) unstructured and semi-structured to structured environments, and (ii) from structured to semi-structured. None of the evaluated methods generalized well from structured to unstructured environments. Therefore we conclude it is possible to train a general and accurate alignment classifier given that training data is sufficiently diverse. Relatively modest results 96% was achieved on live data. We think that the poor quality of the ground truth (obtained by lidar odometry and manual inspection) causes high variance in the CorAI score. The score is sensitive to small misalignment's, therefore a higher quality ground truth is required to make a fair evaluation. We believe that CorAI per-point quality and classification can be a useful tool for alignment evaluation and can improve robustness in various perception tasks by serving as a fault detection step.

In the future we will investigate how to automatically learn sensor specific parameters or use the range image to find neighbouring points for covariance computation. This could address variations in point density owed to different sensors and environment scales.

REFERENCES

- [1] D. Adolfsson, S. Lowry, M. Magnusson, A. Lilienthal, and H. Andreasson, "A Submap per Perspective - Selecting Subsets for SuPer Mapping that Afford Superior Localization Quality," in *2019 European Conference on Mobile Robots (ECMR)*, Sept. 2019, pp. 1–7.
- [2] B. Della Corte, H. Andreasson, T. Stoyanov, and G. Grisetti, "Unified Motion-Based Calibration of Mobile Multi-Sensor Platforms With Time Delay Estimation," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 902–909, Apr. 2019.
- [3] A. C. M. Tavares, F. J. Lawin, and P. Forssén, "Assessing losses for point set registration," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3360–3367, 2020.
- [4] J. Zhang and S. Singh, "LOAM: Lidar odometry and mapping in real-time," in *Robotics: Science and Systems*, 2014.
- [5] S. Nobili, G. Tinchev, and M. Fallon, "Predicting alignment risk to prevent localization failure," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 1003–1010.
- [6] H. Andreasson, D. Adolfsson, T. Stoyanov, M. Magnusson, and A. J. Lilienthal, "Incorporating ego-motion uncertainty estimates in range data registration," in *2017 (IROS)*, Sep. 2017, pp. 1389–1395.
- [7] P. J. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE TPAMI*, vol. 14, no. 2, pp. 239–256, Feb 1992.
- [8] A. Segal, D. Haehnel, and S. Thrun, "Generalized-ICP," in *Robotics: Science and Systems V*. Robotics: Science and Systems Foundation, jun 2009. [Online]. Available: <https://doi.org/10.15607%2Frss.2009.v.021>
- [9] S. Billings and R. Taylor, "Generalized iterative most likely oriented-point (g-imlop) registration," *International journal of computer assisted radiology and surgery*, vol. 10, 05 2015.
- [10] N. Tustison, S. Awate, G. Song, T. Cook, and J. Gee, "Point set registration using havrda-charvat-tsallis entropy measures," *IEEE transactions on medical imaging*, vol. 30, pp. 451–60, 10 2010.
- [11] Q. Liao, D. Sun, and H. Andreasson, "Point set registration for 3d range scans using fuzzy cluster-based metric and efficient global optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 9, pp. 3229–3246, 2021.
- [12] Y. Aoki, H. Goforth, R. A. Srivatsan, and S. Lucey, "Pointnetlk: Robust & efficient point cloud registration using pointnet," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 7163–7172.
- [13] G. D. Evangelidis and R. Horaud, "Joint alignment of multiple point sets with batch and incremental expectation-maximization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 6, pp. 1397–1410, 2018.
- [14] S. Bouaziz, A. Tagliasacchi, and M. Pauly, "Sparse iterative closest point," in *Proceedings of the Eleventh Eurographics/ACM SIGGRAPH Symposium on Geometry Processing*. Eurographics Association, 2013, pp. 113–123.
- [15] T. Stoyanov, M. Magnusson, H. Andreasson, and A. J. Lilienthal, "Fast and accurate scan registration through minimization of the distance between compact 3D NDT representations," *The International Journal of Robotics Research*, vol. 31, no. 12, pp. 1377–1393, 2012.
- [16] S. M. Rusinkiewicz, "Efficient variants of the ICP algorithm," in *The Third International Conference on 3D Digital Imaging and Modeling*, 2001, pp. 145–152.
- [17] B. Eckart, K. Kim, A. Troccoli, A. Kelly, and J. Kautz, "MLmd: Maximum likelihood mixture decoupling for fast and accurate point cloud registration," in *2015 International Conference on 3D Vision*, 2015, pp. 241–249.
- [18] I. Bogoslavskyi and C. Stachniss, "Analyzing the quality of matched 3d point clouds of objects," *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6685–6690, 2017.
- [19] H. Almqvist, M. Magnusson, T. P. Kucner, and A. J. Lilienthal, "Learning to detect misaligned point clouds," *Journal of Field Robotics*, vol. 35, no. 5, pp. 662–677, 2018.
- [20] L. Silva, O. R. Bellon, and K. L. Boyer, "Precision range image registration using a robust surface interpenetration measure and enhanced genetic algorithms," *TPAMI*, vol. 27, no. 5, pp. 762–776, May 2005.
- [21] D. Landry, F. Pomerleau, and P. Giguère, "CELLO-3D: Estimating the covariance of ICP in the real world," in *IEEE (ICRA)*, May 2019, pp. 8190–8196.
- [22] O. Bengtsson and A.-J. Baerveldt, "Robot localization based on scan-matching—estimating the covariance matrix for the IDC algorithm," *Robotics and Autonomous Systems*, vol. 44, no. 1, pp. 29–40, 2003.
- [23] J. Nieto, T. Bailey, and E. Nebot, "Scan-SLAM: Combining EKF-SLAM and scan correlation," in *Field and Service Robotics*, ser. Springer Tracts in Advanced Robotics, P. Corke and S. Sukkariah, Eds. Springer, 2006, pp. 167–178.
- [24] S. M. Prakhyा, L. Bingbing, Y. Rui, and W. Lin, "A closed-form estimate of 3d ICP covariance," in *2015 14th IAPR (MVA)*, 2015, pp. 526–529.
- [25] A. Censi, "An accurate closed-form estimate of ICP's covariance," in *IEEE (ICRA)*, apr 2007, pp. 3167–3172.
- [26] M. Magnusson, "The three-dimensional normal-distributions transform — an efficient representation for registration, surface analysis, and loop detection," Ph.D. dissertation, Örebro University, Dec. 2009, Örebro Studies in Technology 36.
- [27] P. Biber and W. Strasser, "The normal distributions transform: a new approach to laser scan matching," in *Proceedings 2003 IEEE/RSJ (IROS 2003)*, vol. 3, Oct. 2003, pp. 2743–2748 vol.3.
- [28] M. Chandran-Ramesh and P. Newman, "Assessing map quality and error causation using conditional random fields," *IFAC Proceedings Volumes*, vol. 40, no. 15, pp. 463–468, Jan. 2007.
- [29] A. Makadia, A. Patterson, and K. Daniilidis, "Fully automatic registration of 3d point clouds," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, June 2006, pp. 1297–1304.
- [30] D. Droseschel and S. Behnke, "Efficient continuous-time slam for 3d lidar-based online mapping," in *ICRA*, May 2018, pp. 1–9.
- [31] G. Darbellay and I. Vajda, "Entropy expressions for multivariate continuous distributions," *Information Theory, IEEE Transactions on*, vol. 46, pp. 709 – 712, 04 2000.
- [32] M. Magnusson, A. J. Lilienthal, and T. Duckett, "Scan registration for autonomous mining vehicles using 3D-NDT," *Journal of Field Robotics*, vol. 24, no. 10, pp. 803–827, Oct. 2007.
- [33] F. Pomerleau, M. Liu, F. Colas, and R. Siegwart, "Challenging data sets for point cloud registration algorithms," *IJRR*, vol. 31, no. 14, pp. 1705–1711, Dec. 2012.