Analyzing the Quality of Matched 3D Point Clouds of Objects

Igor Bogoslavskyi Cyrill Stachniss

***Abstract*— 3D laser scanners are frequently used sensors for mobile robots or autonomous cars and they are often used to perceive the static as well as dynamic aspects in the scene. In this context, matching 3D point clouds of objects is a crucial capability. Most matching methods such as numerous flavors of ICP provide little information about the quality of the match, i.e. how well do the matched objects correspond to each other, which goes beyond point-to-point or point-to-plane distances. In this paper, we propose a projective method that yields a probabilistic measure for the quality of matched scans. It not only considers the differences in the point locations but can also take free-space information into account. Our approach provides a probabilistic measure that is meaningful enough to evaluate scans and to cluster real-world data such as scans taken with Velodyne scanner in urban scenes in an unsupervised manner.**

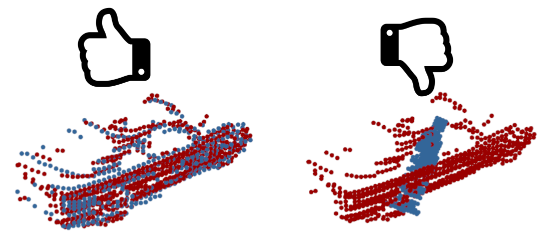

Fig. 1. It is a non-trivial task to determine the alignment quality of matched point clouds of scanned objects as can be seen from the following example. Left: Two well-aligned point clouds of a car (red and blue). Right: Two different objects (car shown in red and person shown in blue) have been aligned with each other. Although the residuals of point-to-point differences may not be too large, the objects should not be matched. This papers makes and attempt to disambiguate such situations. (Images best viewed in color.)

I. Introduction

Nowadays many applications such as autonomous driving rely on 3D data and this data is crucial for autonomous systems to understand the world and especially the moving objects around them. Many methods have been developed for interpreting such type of data. Such an interpretation, for example to estimate the speed of a moving object, requires that the scans of the individual objects are correctly registered with respect to each other. The information about the correctness and quality of the alignment is crucial when dealing with the real world. Not knowing the quality of the match can lead to serious damage of the robot or the environment, especially if the robot relies on scan matching while navigating in dynamic or hazardous environments [2]. Thus, this aspect is of great importance for automated driving as an autonomous car needs to be able to track obstacles on the road, e.g. by matching features defined on the 3D data [5]. Performing a wrong match can potentially lead to estimating speed of other cars wrongly or result in failing to recognize a pedestrian or loosing track of an object.

The most popular approach to matching point clouds, either of whole scenes or individual objects, is probably ICP [1] and its numerous flavors, we refer to [12] for an overview and a comparison. Recently, there have been ICP variants proposed that compute a globally optimal alignment using a branch and bound approach [18]. Other techniques involve correlative scan matching [11], feature-based matching, and many more. The matching can be performed on raw 3D data or by exploiting different types of features with varying complexity. While there is a zoo of features available today, much less attention has been payed to the question how to evaluate the matching quality of after the alignment has been carried out. Most approaches either use a sum of squared point-to-point distances or point-to-plane measures after alignment. We would like to stress the point here that even in case a global alignment has been found, an evaluation of the matching is beneficiary as in moving scenes, individual moving objects may have been wrongly associated to each other.

In this paper, we address the problem of evaluating the quality of 3D point clouds of objects in potentially dynamic scenes. We assume that the scan has already been segmented into objects, for example using [3] or a similar approach, and that the segmented objects have been individually aligned with an arbitrary registration method such as ICP. The objective of this paper is to provide an approach that evaluates such matches. It should provide a high score in situations in which the same objects are correctly aligned and low scores if the both objects are not aligned well or if two objects from different classes (such as car and pedestrian) are aligned. The approach proposed in this paper aims at achieving this, it is of probabilistic nature, and takes into account occupancy and free-space information, which is typically ignored when matching point clouds using ICP.

The main contribution of this work is a novel approach that can generate a probabilistically motivated quality measure for the alignment of two point clouds corresponding to individual objects. Our approach is based on the observations of two point clouds and focuses on the projections of these point clouds to virtual image planes with expected free space information around objects. We explicitly consider potentially moving objects segmented from a range scan such as people or cars, see Fig. 1 for an example. To illustrate the usefulness of the proposed evaluation, we present several experiments analyzing aligned 3D scans of objects and also

Both authors are with the Institute of Geodesy and Geoinformation, University of Bonn, Germany.

show that our measure can be computed in the order of 1 ms/object. We furthermore illustrate that the analysis of an alignment can be used to support an object-tracker estimating the trajectories of dynamic objects or to help clustering real world objects and is working better than considering a point-to-point similarity score.

II. Related Work

Finding a good scan alignment is a well studied problem and several approaches have been proposed in the past. The most popular solution is probably the ICP algorithm [1] and many variants of the original algorithm exist. Most ICP-based approaches minimize the point-to-point distances between potentially corresponding points but other variants such as point-to-plane of generalized ICP [15] exist. We recommend the paper by Pomerleau et al. [12] for an overview.

To the best of our knowledge, mostly the probabilistic evaluation of the quality of the point cloud matching has been done in correlation with the actual matching process. For example, Olson [11] proposes a system that performs correlative scan matching, maximizing the likelihood of the match between full scans. This approach is suited for 2D data and would probably be computationally expensive if extended to 3D. To the best of our knowledge, there is no efficient 3D variant that searches in the 6D transformation space. In 2D, this method provides covariance estimates, which is valuable in determining the quality of the scan alignment. We think of this covariance matrix as an orthogonal information to our approach.

Another notable example of a method that provides a covariance matrix is a method by Censi [4]. He treats scan matching as a probability distribution approximation problem and provides an estimate of the matching uncertainty. His method seems to be reliable under severe sensor occlusions and it handles gracefully in constrained situations. However, this methods has been proposed for 2D matching and, to the best of our knowledge, is mainly used as a 2D method.

A 3D variant of covariance matrix estimation can be found in modern techniques for performing ICP. A recent example of such a method is a work by Prakhya *et al.* [13]. They provide an extension of [4] to 3D data. They also show that the covariance of their method is lower in the global minimum comparing to local ones. However, it is still unclear which values of the covariance matrix should be considered good and which are not. Our method, while not forbidding to use the covariance matrix for match likelihood estimation provides additional probabilistic measure, that is a single number and therefore is relatively easy to interpret.

Recently, Yang *et al.* [18] proposed a variant of ICP that searches for the globally optimal alignment of scans using a branch and bound approach. This is an interesting techniques, with guarantees on the performance. Even if a global approach is used, however, our metric will still be beneficial, as it could help to detect a individual, wrong alignment of a moving object to a different object.

Another notable method for registering dense point clouds is NICP by Serafin and Grisetti [16] that is designed for aligning full point clouds from Kinect-style sensors. This approach considers normal information for the matching and its open source implementation makes use of projections for matching. A related form of projection is also used in our work. A further work that has a similar motivation but is targeted to 2D SLAM is the work by Mazuran *et al.* [9]. They analyze the map consistency by performing cascaded statistical tests on pairs of 2D scans and overlapping 2D polygons and use this information to determine parameters in the used SLAM backend. A further approach that bears similarities to the ideas presented here is the work by Hähnel *et al.* [8]. As part of their approach, they analyze the log-likelihood of data associations in SLAM, when searching in the space of data associations. The log-likelihood of each measurement is obtained by superimposing a scan onto a local 2D occupancy grid map built by another scan.

One more interesting method by Endres *et al.* [6] models object classes as distributions over features and use Latent Dirichlet Allocation to learn clusters of 3D objects according to similarity in shape. However, it is interesting for us to avoid using features to describe the scene and to work on pure 3D data.

To the best of our knowledge, there are not many approaches to estimating the scan alignment quality that go beyond sums of squared distances between points or planes or estimated covariance matrices in the space of transformations. Most approaches furthermore use scans as a whole and not partial scans of objects. Our work aims at contributing a new approach to this problem by combining the idea of mutual projections and the exploitation of free space information for 3D point clouds.

III. Evaluating the Alignment Quality

Our approach is supposed to evaluate the alignment of two already aligned, 3D range scans of (partially) scanned objects such as those depicted in Fig. 1. Our approach is completely agnostic to the used alignment method as long as it computes the 3D transformation between the sensor poses at scanning time. For this paper, we use a segmentation approach [3] that transforms full 3D scans into individual objects and use the ICP implementation of the point cloud library (PCL) [14]. Note that any other segmentation or alignment technique can be used instead.

Let C_1 and C_2 refer to two 3D point clouds that have been registered in a common but arbitrary reference frame. Let O_1 be the origin of cloud C_1, i.e. the pose of the sensor with respect to C_1 when recording the cloud. The same holds for O_2 with respect to C_2.

We can define two virtual image planes, one close to O_1 and one close to O_2 pointing towards the object, see Fig. 3 for an illustration. The comparison of the two clouds is performed based only on the projections of the point clouds as depth images on the virtual image planes. This leads to four projected depth images I_{11}, I_{12}, I_{21}, and I_{22}, where I_{ci} refers to the projection of the cloud C_c into the first

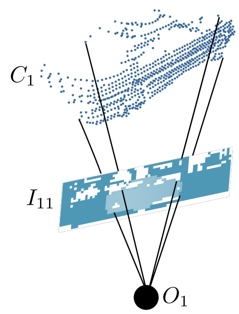

Fig. 2. A sketch of the projection with darker shades on the projection plane depicting free space information.

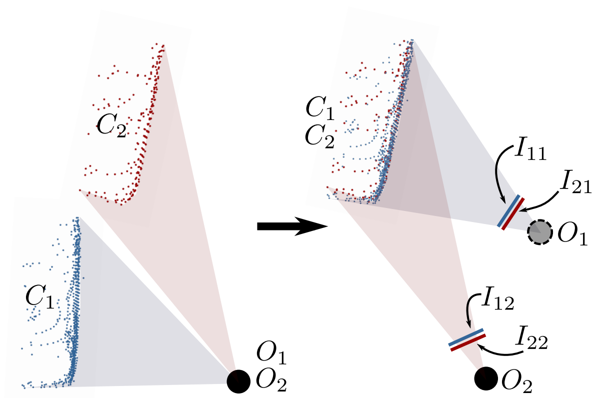

Fig. 3. This image depicts two clouds viewed from above before and after registration with ICP. Note that both clouds are projected to all four depth images: I_{11}, I_{21}, I_{12} and I_{22}. An example projection I_{11} is shown in Fig. 2. We compute a similarity measure for each pixel of corresponding projections following Eq. (1). Knowing individual probabilities from pixels we continue to combining them with Eq. (2). (Images best viewed in color.)

or second image plane (index i). Each pixel in I_{ci} stores the distance between the 3D point in C_c and the origin O_i. Is is important to note that for the projections I_{11} and I_{22}, we can also exploit negative information, i.e. knowledge about free space. We label the pixels surrounding the objects, which are generated from a scan segmentation approach (here using [3]) as free space if their depth readings are larger than the distance to the object itself. This free space information is available only for the depth images I_{11} and I_{22} as we know that there is free space around the object as seen from the origin. This information may not be available for I_{12} and I_{21} as here the projected image does not necessarily lie between the point cloud and the physical scanner location during data acquisition.

Our analysis relies on comparing the depth images I_{11} to I_{21} and I_{12} to I_{22}. Thus, we are comparing the projections of the two clouds in the same (virtual) image plane. As only I_{11} and I_{22} encode free space information, there are the following possibilities when comparing the range images pixel by pixel. Let d_j^{ci} be the depth information of pixel j in I_{ci}, then we have three possible cases:

- Both images I_{cc} and I_{ci} with $c \in \{1, 2\}, i \neq c$ have a depth value stored in pixel j. In this case, we compute the probability of the two depth values to be generated by the same object by

$$p_j = 1 - \frac{1}{\sqrt{2\pi}\sigma} \int_{-\Delta_j}^{\Delta_j} e^{\frac{-t^2}{2\sigma}} dt = 1 - \left(\Phi(\frac{\Delta_j}{\sigma}) - \Phi(\frac{-\Delta_j}{\sigma}) \right), \quad (1)$$

  where Δ_j is the distance between the depth readings in the virtual image plane at pixel j. Eq. (1) considers Gaussian measurement noise with standard deviation σ and thus the probability p_j is the area under the tails of the normal distribution. This area can be computed via a difference of the following cumulative distribution functions (CDFs) of the given normal distribution.
- Pixel j of image I_{ci} has a depth reading while a pixel with the same coordinates in image I_{cc} is marked as free space. In this case, we set the probability generated by these pixels to a low value corresponding to a value of 2σ as Δ_j in Eq. (1).
- Pixel j in either of the images I_{cc} or I_{ci} contains no value at all. In this case, we do not have enough information to make any decision and ignore the pixel.

We perform this computation for all the pixels $j = 1, \ldots, M$ in the projected depth image that have values in cloud C_1 and C_2. At this point, we have a probability p_j for each pixel $j \in \{1, M\}$, where M is the total number of pixels that have a non-zero value for both projections.

We consider the evaluation of each pixel as an expert that tells us the probability that the scans match. We can now apply a method to combine multiple expert opinions into one probability so that this probability defines degree of similarity between clouds C_1 and C_2. This problem is a well-known problem called opinion pooling. If we have no further information, we use a so-called democratic opinion pool [17], i.e. the similarity between two point clouds C_1 and C_2 is determined by a linear opinion pool:

$$p(C_1, C_2) = \sum_{j=1}^{M} \lambda_j p_j \quad (2)$$

where $\lambda_j = \frac{1}{M}, \forall j \in \{1, M\}$ and p_j are opinions reported by a corresponding expert, i.e. probabilities computed using Eq. (1). Given this approach, we finally obtain with $p(C_1, C_2)$ a similarity measure between two 3D point clouds exploiting projections and free space information.

IV. Experimental Evaluation

We propose a measure to evaluate how well two 3D point clouds of objects are aligned. Thus, this evaluation is designed to show that (i) this measure is a useful tool for quantifying the alignment quality of 3D range data of objects, (ii) our approach can be executed fast, typically below 1 ms. Furthermore, we illustrate that (iii) it can support tracking and (iv) we can even use it to cluster different object perceived in 3D scans and obtain semantically meaningful objects and perform better than an ICP-like point-to-point measure, here using the implementation of PCL.

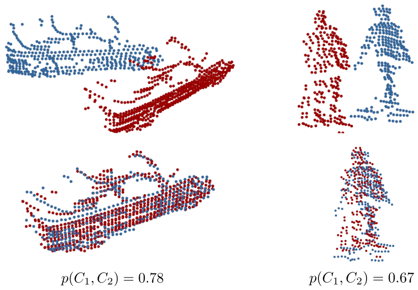

Fig. 4. Example clouds of two cars and two pedestrians aligned using ICP algorithm along with the value of the similarity measure reported by our algorithm.

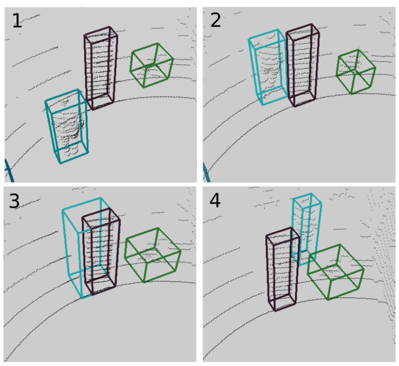

Fig. 5. Our measure supports a dynamic object tracker and can help to reject data associations generated by matching nearby point clouds. The sequence of images shows the result of an EKF-based tracker performing a cloud validation step using our measure. We only update an object track if the similarity measure between a new object and a tracked one is high. As can be seen in the sequence, the person (cyan) is not fused with the flat vertical object (brown) and so this illustrates that our approach helps to disambiguate the objects so that separate tracks can be maintained. In the Images 1 and 2, the person approaches an object. In Image 3 the person is occluded by the object and the bounding box shows an EKF prediction. Note that the data association is done correctly. Image 4 depicts that EKF is able to resume tracking of the person once the person is seen again maintaining the original track id (given by the color).

For our evaluation, we used several scans from the KITTI dataset [7] that have been recorded with a 64-beam Velodyne scanner. In this part of the KITTI dataset, typical objects are cars, people, vans, etc. We furthermore used sparser 3D data from a 16-beam Velodyne VLP-16 scanner recorded with a mobile robot on our campus in Bonn.

A. Alignment Quality

The first set of experiments is designed to illustrate that our approach for analyzing the alignment of scanned objects is a useful tool and provides meaningful scores with respect to the alignment quality. We analyze two different types of experiments here. First, the registration of the same physical object observed from different locations based on typical street scenes. Second, we evaluate how well different objects of the same class, e.g., two pedestrians or two cars can be aligned.

Fig. 4 depicts two range scans and similarity scores of an object before and after registration of ICP (using the PCL implementation). We use this implementation throughout this work, however any other registration technique can be used instead. Fig. 6 shows how the disturbance of the alignment changes the similarity score of our approach. As can be seen, the function peaks at the correct alignment. The plots illustrate how deviation from the true alignment change the score. The larger the deviation the smaller the similarity score.

TABLE I

AVERAGE RUNTIME FOR EVALUATING PAIRS OF 3D SCANS OF DIFFERENT OBJECTS ON AN INTEL I7 CPU

dataset	objects	average runtime
KITTI (64-beams)	cars	approx. 0.45 ms
KITTI (64-beams)	pedestrians	approx. 0.38 ms

B. Runtime

The next experiment is designed to show that the alignment score for a two 3D point clouds recorded with a regular laser range scanner can be computed in a efficient manner so that it can be used for online operations easily. To quantify the runtime requirements, we executed the evaluation of different objects of different size and measured the runtime on a regular desktop computer with a Intel i7 CPU. The timings are summarized in Tab. I.

As can be seen from the table, the average computation time for typical objects such as cars or pedestrians can be executed in below 0.5 ms. Thus, our approach is suitable for an online assessment of the alignment quality for up to 100+ individual objects in the scene considering a frame rate of 10 Hz of the Velodyne laser scanner.

C. Support for Tracking Dynamic Objects

The next experiment is designed to illustrate that the quality analysis of point clouds can support trackers that seek to estimate the trajectories of dynamic objects in the environment. To do this, we compute the quality measure for a made data association and subsequent point cloud alignment and reject associations that receive a low score. This approach rejects matchings in which people are fused with nearby walls or other flat objects. An example of such a situation is depicted in Fig. 5. The sequence of images from 1 to 4 shows the result of an EKF tracker aided by our similarity measure rejecting data associations that receive a low score using our evaluation. As can be seen, here the tracks are not fused and the objects get tracked correctly.

D. Support for Clustering Objects

Finally, we want to illustrate that our score is not only suited for evaluation of the alignment of scans taken from the same objects but could also be used for clustering different types of scanned objects in an unsupervised way and works better than the point-to-point score that standard ICP provides.

To illustrate that, we extracted the scanned cars, vans, and pedestrians from the KITTI dataset and computed a pair-wise ICP alignment after shifting all clouds so that the barycenter of each of them is in the origin. After the ICP alignment, we

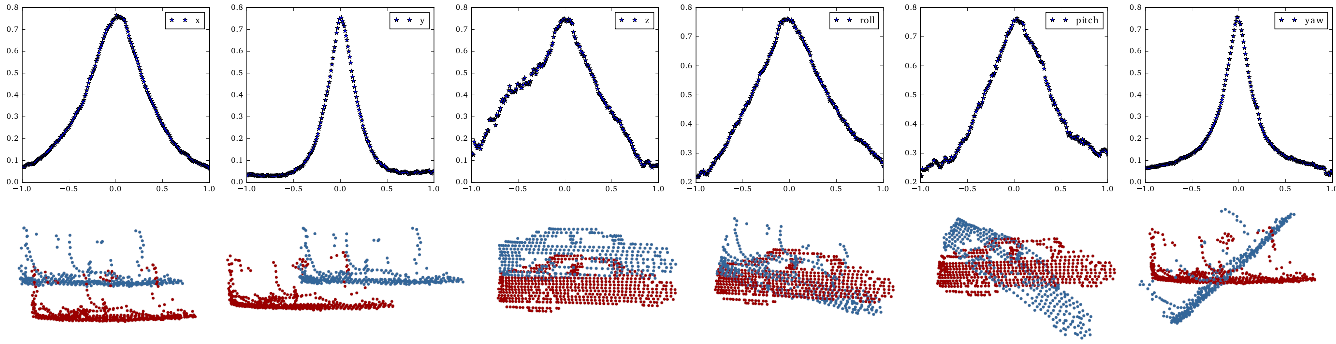

Fig. 6. This image shows the changes in proposed similarity measure with a change in the relative position of two matched clouds. We have evaluated the changes in x, y, z directions by displacing point clouds by up to 1 m as well as changing the roll, pitch and yaw of one of the clouds by an angle of up to 1 rad.

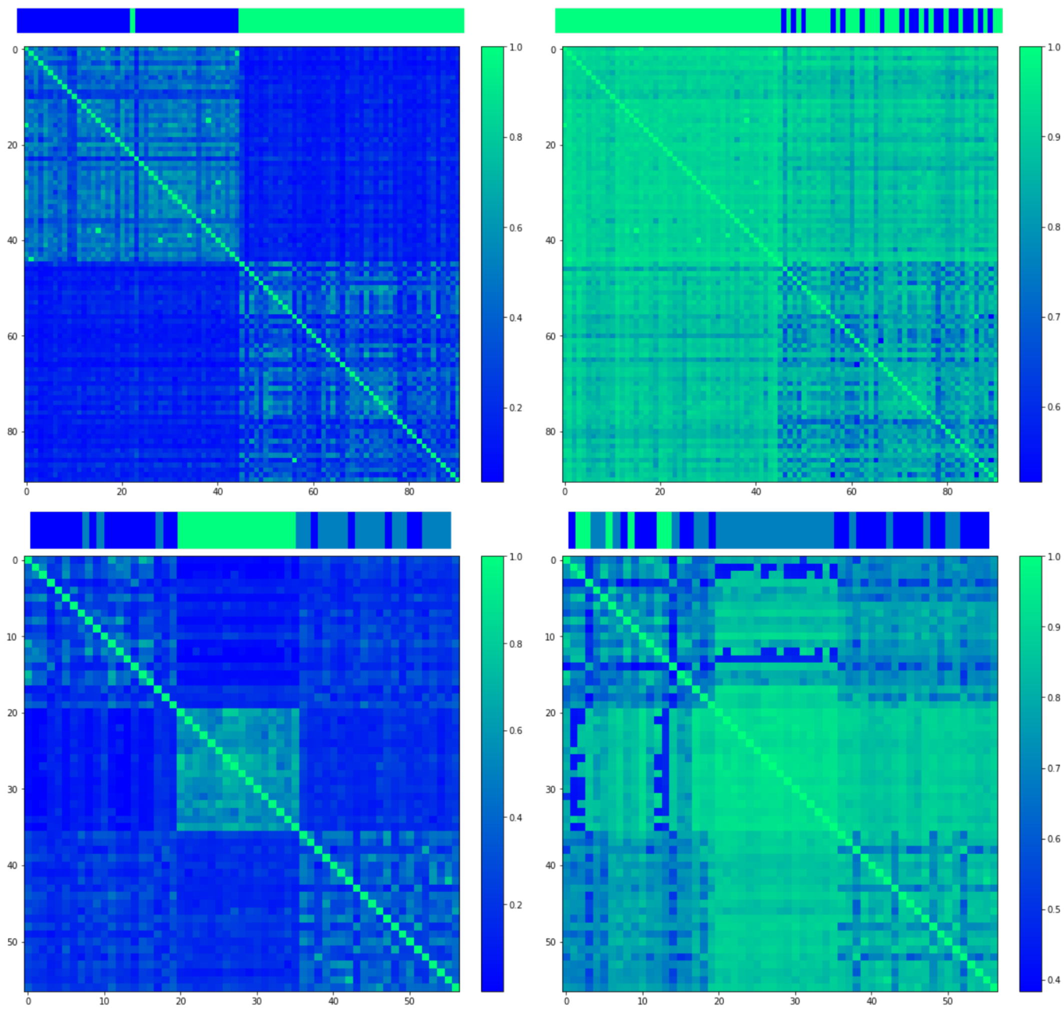

Fig. 7. Top row: Left and right top images show similarity matrices of pairwise compared between 45 people scans against 54 cars. Left image is estimated using our method, while the right image using RMSE provided by PCL (with RBF kernel on top of it to generate similarity measures). The bars on top of the matrices (blue and green refer to scans of different classes of objects, here cars and people) show the results of unsupervised spectral clustering run on the corresponding matrices. Bottom row: The images show our similarity measure on the left and RMSE put through a robust kernel on the right. The data is sampled randomly from the annotations from driving sequences 91 and 51 from the KITTI dataset. In this example, we used 20 vans, 15 pedestrians and 25 cars. They are depicted in that order on the axis of each matrix. As can be seen from the bars on top of the matrices, the clustering performs better using our measure than for the RMSE. Both algorithms segment people into a single class correctly, however, our measure provides better separation between cars and vans.

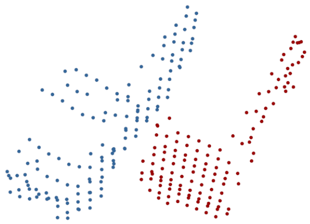

Fig. 8. This image shows two examples of vans that are hard to match properly shown in different colors. This explains bad matching score in the top-left corner of Fig. 7. The points of these vans have different density, the vans have different shape and point in different directions. A human would likely also struggle to decide that these clouds should belong to one class.

compute the similarity value using our approach between all pairs and store the values in a similarity matrix:

$$P_{N,N} = \begin{pmatrix} p(C_1, C_1) & p(C_1, C_2) & \cdots & p(C_1, C_N) \\ p(C_2, C_1) & p(C_2, C_2) & \cdots & p(C_2, C_N) \\ \vdots & \vdots & \ddots & \vdots \\ p(C_N, C_1) & p(C_N, C_2) & \cdots & p(C_N, C_N), \end{pmatrix} \quad (3)$$

where $p(C_a, C_b)$ refers to the similarity score between point clouds C_a and C_b, while N being the number of objects. For a comparison, we perform the clustering also based on the RMSE resulting from ICP's point-to-point metric. Fig. 7 illustrate two visual representations of such a matrix with two and three different types of objects. The left image matrix always corresponds to the similarity matrix of our approach while the one on the right corresponds to the RMSE-based matrix. In the images, dark blue corresponds to $p = 0$, while light green to $p = 1$.

We sorted the point clouds according to the class of scanned object and thus distinct squares in the matching matrix indicate that the measure can be used for clustering objects. To verify this, we performed spectral clustering [10] with both scores and test if the different classes of objects are correctly found given manually labeled ground truth data.

We used out-of-the-box spectral clustering as a simple approach to group objects. There may be more sophisticated ways for unsupervised clustering of 3D objects but this experiment suggests that our score serves as a better indicator than point-to-point RSME of two point clouds actually match well.

Fig. 7 shows that the classes are separated in a meaningful way when using our method. A failure case for our (and the point-to-point) approach is depicted in Fig. 8. As can be seen, these two wrongly grouped examples of two vans driving in opposite direction are actually hard to match clouds. The clouds are rather sparse, both objects have somewhat different shapes (front and back), and drive into different directions. When performing spectral clustering based on the RMSE similarity matrix, the performance drops clearly. This suggests that our method supports such similarity-based clustering of objects better than the RMSE metric and provides enough information for an unsupervised clustering algorithm to find classes from unlabeled data.

V. Conclusion

Registering 3D point clouds is a frequently performed task in various robotics applications. In this paper, we propose a novel way for analyzing the alignment quality of registered point clouds of individual objects. We provide a fast to compute, probabilistic similarity measure for any pair of registered point clouds of objects and do not rely on any specific registration or segmentation procedure. Our approach uses projections of the point clouds alongside with information about the free space that surrounds the scanned objects to evaluate a match. As our experiments suggest, the probabilistic measure is well-suited to analyze matches, supports tracking dynamic objects, and even allows us to cluster point clouds of different types of objects in an unsupervised way better than a point-to-point metric would do.

References

[1] P.J. Besl and N.D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

[2] I. Bogoslavskyi, M. Mazuran, and C. Stachniss. Robust homing for autonomous robots. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2016.

[3] I. Bogoslavskyi and C. Stachniss. Fast range image-based segmentation of sparse 3d laser scans for online operation. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2016.

[4] Andrea Censi. Scan matching in a probabilistic framework. In *ICRA*, pages 2291–2296. Citeseer, 2006.

[5] A. Dewan, T. Caselitz, G.D. Tipaldi, and W. Burgard. Rigid scene flow for 3d lidar scans. 2016.

[6] F. Endres, C. Plagemann, C. Stachniss, and W. Burgard. Unsupervised discovery of object classes from range data using latent dirichlet allocation. In *Proc. of Robotics: Science and Systems (RSS)*, 2009.

[7] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013.

[8] D. Hähnel, W. Burgard, B. Wegbreit, and S. Thrun. Towards lazy data association in SLAM. In *Proc. of the Int. Symposium of Robotics Research (ISRR)*, 2003.

[9] M. Mazuran, G.D. Tipaldi, L. Spinello, W. Burgard, and C. Stachniss. A statistical measure for map consistency in slam. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2014.

[10] Andrew Y Ng, Michael I Jordan, Yair Weiss, et al. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 2:849–856, 2002.

[11] E. Olson. Real-time correlative scan matching. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Kobe, Japan, June 2009.

[12] F. Pomerleau, F. Colas, R. Siegwart, and S. Magnenat. Comparing ICP variants on real-world data sets. *Autonomous Robots*, 34(3):133–148, 2013.

[13] S. M. Prakhya, L. Bingbing, Y. Rui, and W. Lin. A closed-form estimate of 3d icp covariance. In *Machine Vision Applications (MVA), 2015 14th IAPR International Conference on*, 2015.

[14] R.B. Rusu and S. Cousins. 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 9-13 2011.

[15] A. Segal, D. Haehnel, and S. Thrun. Generalized-icp. In *Proc. of Robotics: Science and Systems (RSS)*, volume 2, 2009.

[16] J. Serafin and G. Grisetti. NICP: Dense normal based point cloud registration and mapping. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2015.

[17] M. Stone. The opinion pool. *Ann. Math. Statist.*, 32(4):1339–1342, 12 1961.

[18] J. Yang, H. Li, D. Campbell, and Y. Jia. Go-icp: A globally optimal solution to 3d icp point-set registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(11):2241–2254, 2016.