

Scan Segments Matching for Pairwise 3D Alignment

B. Douillard, A. Quadros, P. Morton, J. P. Underwood, M. De Deuge, S. Hugosson, M. Hallström, T. Bailey
The Australian Centre for Field Robotics, The University of Sydney, Australia

Abstract—This paper presents a method for pairwise 3D alignment which solves data association by matching scan segments across scans. Generating accurate segment associations allows to run a modified version of the Iterative Closest Point (ICP) algorithm where the search for point-to-point correspondences is constrained to associated segments. The novelty of the proposed approach is in the segment matching process which takes into account the proximity of segments, their shape, and the consistency of their relative locations in each scan. Scan segmentation is here assumed to be given (recent studies provide various alternatives [10], [19]). The method is tested on seven sequences of Velodyne scans acquired in urban environments. Unlike various other standard versions of ICP, which fail to recover correct alignment when the displacement between scans increases, the proposed method is shown to be robust to displacements of several meters. In addition, it is shown to lead to savings in computational times which are potentially critical in real-time applications.

I. INTRODUCTION

This paper investigates various ways of exploiting segmentation (illustrated in Fig. 1) to improve the pairwise alignment of 3D scans. One core issue in accurately aligning 3D scans is data association. Standard alignment techniques use the Iterative Closest Point (ICP) algorithm [22] which solves data association at a point-to-point level¹. However, as noted in the literature [23] and as shown in the proposed experiments, this method fails when nearest points do not correspond to correct association (illustrated in Fig. 3). This paper proposes to solve data association by reasoning at the object level rather than at the point-to-point level. Recently proposed segmentation techniques [10], [19] allow the separation of objects in 3D scans. Using such segmentation as an input, the proposed methods explicitly match objects (i.e. segments) across scans as opposed to reasoning at the –possibly myopic– point-to-point level. The experiments demonstrate that standard nearest neighbour matching provides incorrect data association as the distance between scans increases, while object level matching is robust to displacement of a few meters. This leads in turn to sharper scan alignments that are not only able to recover larger displacements but also allows computational savings that are potentially critical in real-time applications.

In recent years, a number of consumer grade 3D imaging devices have been released. These include time-of-flight based sensors such as the D-Imager [1] and the Swiss-Ranger [4] for indoor environments and the Velodyne [2] for outdoor deployments. Structured light based technology has

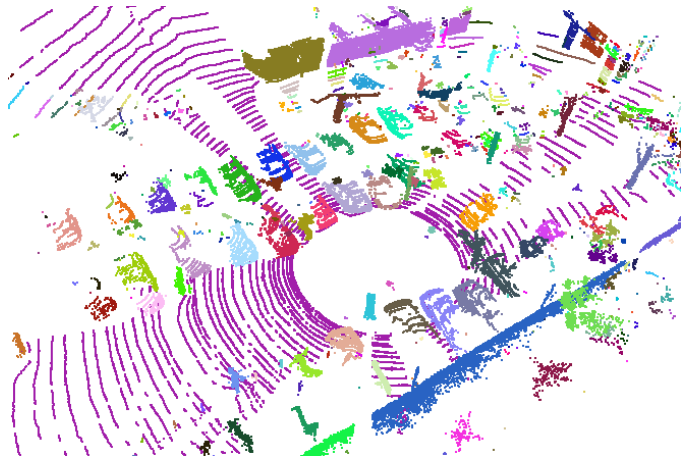


Fig. 1. Example of segmentation of a Velodyne scan acquired in a parking lot. This scan belongs to the dataset Campus car park; the datasets are described in Table I. Segment IDs are mapped to colours. It can be seen that all the cars and trees around the centre of the scan are correctly separated while some over-segmentation can be observed in the background of the scene, where the data are sparser. The extracted ground surface is indicated in purple.

also reached the consumer market with the Kinect sensor [3]. These sensors have contributed to a wealth of research in a variety of domains including autonomous driving [17], 3D modelling [12], grasping and manipulation [21], [13] and object recognition [11], to only mention a few. The work presented here belongs to this overall effort attempting to exploit the richness of 3D data to build more accurate perception systems. Specifically, we concentrate on the pairwise registration of Velodyne scans acquired in outdoor urban environments with the aim of building sharp 3D alignments of sequences of scans. The module developed here has applications to map building and object modelling via 3D surface registration.

The problem of 3D scan alignment has been studied for over two decades. A recent survey [23] shows that 3D alignment techniques have converged around one main pipeline which involves two broad steps: coarse alignment followed by fine alignment. Coarse alignment uses for instance Principal Component Analysis [15] or feature based matching with features such as the Spin Image [14] to compute an initial coarse registration. Fine alignment then generally involves the Iterative Closest Point (ICP) algorithm [6] with the few exceptions of techniques using Genetic Algorithms [24] or Signed Distance Fields [18] for instance. In the last two years techniques searching efficiently and exhaustively for a global optimum (as opposed to implementing a greedy optimisation as in ICP) by means of Fourier transforms were intro-

¹This includes nearest neighbour search and point-to-plane matching, the latter also requiring point-to-point associations [22].

duced [8]. These have the potential of addressing a number of the shortcomings of ICP and will be investigated in future work. The pipeline for 3D alignment that is proposed here implements coarse alignment by matching segments across scans. Matched segments are then used to constrain point-to-point associations in regular ICP. Specifically, nearest neighbour search is performed in matched segments only. To the best of the authors knowledge, a pipeline combining 3D object segmentation, segment matching across scans, and alignment, had not been proposed before. This pipeline is able to maintain alignment accuracy when the displacement between scan increases unlike more standard variants of ICP, while also being more computationally efficient. It can also be thought of as a pre-processing step for ICP. Since these more standard versions of ICP are often used in ground robotics systems [17], [20], the proposed module, used as an alternative or complementary method for alignment, can have an impact on a variety of robotics platforms.

In summary, the contribution of this work is a technique for scan alignment based on ICP which solves data association at the object level by explicitly matching segments across scans rather than using standard point-to-point type of search. It is shown to provide improved performance with respect to point-to-point matching when the displacement between scans increases, while in addition being more computationally efficient.

II. PAIRWISE SCAN MATCHING

Sec. II-A introduces the segmentation method used in this work. Sec. II-B to II-D describe the proposed alignment technique.

A. Scan Segmentation

Each velodyne scan is first segmented using the technique described in [10]. The output of this technique is illustrated in Fig. 1. In short, this method builds a terrain mesh from a Velodyne scan, identifies ground points via gradient computations in the terrain mesh and partitions the remaining points via clustering in a voxel grid. The method does not assume the terrain to be flat. It is applied to each scan. The resulting segments are further processed to speed up the computation of the alignments. The segments are voxelised using a grid of resolution 20cm and only the segments containing at least n voxels and with a minimum height of z are kept. In our implementation n is set to 100 and z to 1m. The resulting set of N segments is referred to as $S = \{s^i\}_{i=1\dots N}$ in the algorithms below.

<p>Input : $S_1 = \{s_1^i\}_{i=1\dots N}, S_2 = \{s_2^j\}_{j=1\dots M}$ Output: <i>transform</i> 1 $\{(s_1^p, s_2^q)_{(p,q) \in N \times M}\} = \text{MatchSegments}(S_1, S_2)$; 2 $\text{transform} = \text{Icp}(\{(s_1^p, s_2^q)_{(p,q) \in N \times M}\}, S_1, S_2)$;</p>
--

Algorithm 1: Scan Alignment

B. Scan Alignment Overview

The two main components of the proposed scan alignment process are described in Algorithm 1. The core contribution

of this work lies in the operation `MatchSegments` which defines pairwise segment associations. This operation is described in Sec. II-C. Once segment associations are defined, a slightly modified version of ICP algorithm is run to estimate the 6 DOF transformation specifying the alignment of the two scans. This is described in Sec. II-D. The parameters required by each function are detailed in the corresponding sections.

C. Segment Associations

The overall segment association process is defined in Algorithm 2. The main operations are: (1) the definition of a distance matrix (operation `GenerateDistMatrix`), (2) the iteration of the two following steps: (a) definition of segment associations (function `Hungarian`) and (b) the filtering of possibly wrong associations based on the spatial arrangement of segments (function `GeoConsistency`). The thresholds t_d and t_c involved in step 2 are progressively increased to allow more segments to be selected in case an insufficient number of segment associations is initially obtained while keeping the tolerance as tight as possible. These operations are further detailed in the following sections. C_1 and C_2 (line 6) refer to the positions of the centres of mass of the segments in S_1 and S_2 , respectively.

The minimum number of accepted segment associations is set to 4 (after manual adjustment). This is implemented on line 7 of Algorithm 2 where the function `Card` returns the cardinality of set A , i.e. the number of segment associations. The upper bound for t_d is set to 0.7 (line 9). The upper bound for t_c is set to 2m (line 11). These upper bounds are defined experimentally by computing statistics on hand-labelled correct and incorrect segment associations. Due to space constraints, the computation of these bounds is not described here.

<p>Input : $S_1 = \{s_1^i\}_{i=1\dots N}, S_2 = \{s_2^j\}_{j=1\dots M}$ Output : $A = \{(s_1^p, s_2^q)_{(p,q) \in N \times M}\}$ Parameters: <i>gate</i> 1 $t_d = 0.2$; 2 $t_c = 0.2$; 3 $D = \text{GenerateDistMatrix}(S_1, S_2, \text{gate})$; 4 while <i>True</i> do 5 $A = \text{Hungarian}(D, t_d)$; 6 $A = \text{GeoConsistency}(A, C_1, C_2, t_c)$; 7 if $\text{Card}(D) \geq 4$ then 8 break; 9 if $t_d \leq 0.7$ then 10 $t_d = t_d * 1.1$; 11 if $t_c \leq 2.0$ then 12 $t_c = t_c * 1.1$; 13 end</p>

Algorithm 2: Function `MatchSegments`

1) *Distance Matrix Generation*: The matrix D (line 3) is obtained by computing distances from all segments of one scan to all segments of the other scan. Various distances can be used. Here we use the shape distance metric defined in [11], which is also referred to by the authors of the latter

publication as the Symmetric Shape Distance. It evaluates the shape similarity between two object scans. The computation of the Symmetric Shape Distance requires the two 3D surfaces to be first aligned, which is achieved here using ICP, as described in [11].

To accelerate this process which is $O(NM)$, the distance between segments whose centres of mass are further away than *gate* is not computed. *gate* is set to 2m in the experiments corresponding to table II, and 15m in the experiments corresponding to table III (since in the latter case the displacement between each scan is larger).

2) *Bipartite Assignment*: Given the set of distances in D , the Hungarian algorithm [16] is used here to solve the segment association problem. Segment association is interpreted as a bipartite assignment problem where a segment in one scan is uniquely assigned to a segment in the other scan (some segments may be left not assigned). Only segment pairs whose corresponding distance in D is below t_d are processed. t_d is progressively increased (line 11) until enough segment associations are found. This allows to maintain a tight tolerance on the shape difference of associated segments (i.e. a tight tolerance on the corresponding distance in D) while accepting a minimum number of segment associations (line 7) to enable the computation of the alignment. Examples of values of t_d computed by the algorithm are indicated in Fig. 3.

Depending on the displacement of the platform between the two scans to be aligned, the point of view from which the scene is observed may be different enough that objects are seen from different sides (for instance, with the platform moving forward, a car may be seen from the front in scan 1 and then from the side in scan 2). Increasing t_d caters for this type of situation since segment associations are searched by progressively tolerating larger shape differences (i.e. larger distances in D).

3) *Geometric Consistency*: Geometric consistency refers here to a mechanism used to reject possibly wrong segment associations. It exploits the spatial arrangement of segments by verifying that the set of associated segments have consistent relative locations in each scan. For instance the distance between two trees observed in two successive scans should be constant; if this is not the case, the trees observed in the two scans are not the same and their matching across scan is not correct. Enforcing geometric consistency is a way to implement this type of check. Pairs of associated segments not having consistent relative positions in each scan are discarded; the resulting set of segment associations is then qualified as geometrically consistent.

The procedure used here to check for geometric consistency is detailed in Algorithm 3. It takes as an input the set of segment associations A found by bipartite matching (Sec. II-C.2) as well as the position of the centre of mass of the segments in S_1 and S_2 , noted C_1 and C_2 , respectively. Each association in A is encoded as a pair $(p, q)_{p \in N, q \in M}$, where p and q are segment IDs from the current and previous scan, respectively. For concision, the following notation is used with respect to accessing pairs (p, q) in A : " $l \in A$ " means

that l is a pair of the form (p, q) , with p and q accessed as $l[0]$ and $l[1]$, respectively (see line 1 and 2); " $A[p]$ " means that the pair (p, q) is accessed via p (see line 20).

For each possible pair of associations (line 1 and 2) the distances d_0 and d_1 between the centres of mass in the current scan (d_0) and the previous scan (d_1) are computed (line 5 and 6). A check is performed to avoid processing each association pair more than once (line 3).

If the associations l_a and l_b are correct, the difference *diff* between d_0 and d_1 is close to zero (line 7). As mentioned in the previous section, changes in view point, imply changes in the shape of the observed segments since different sides of a given segment may be scanned. This also implies a variation in the position of the centre of mass of the segments and thus requires t_c to be adjusted to allow for more tolerance when the change in view point is more important. The adjustment of t_c is managed by the parent algorithm, Algorithm 2 (line 12). In Algorithm 2, t_c and t_d are increased at each iteration until they respectively reached their upper bound. Our implementation in fact alternates increasing t_c and t_d (so that the tolerances are maintained as tight as possible) but Algorithm 2 presents a slightly different version for concision. Examples of values of t_c computed by the algorithm are indicated in Fig. 3.

If the difference between d_0 and d_1 is below the threshold t_c (line 7), the positions of the two centres of mass in each scan is considered consistent and a counter representing the number of consistent relative positions (in the current scan) is incremented for each segment (line 8 and 9). Otherwise, a counter representing the number of inconsistent relative positions is incremented (line 11 and 12).

Finally, if for a given segment (in the current scan) the counter of consistent relative positions is less than or equal to the counter of inconsistent relative positions (line 19), the corresponding segment association is removed from the original set of segment associations (line 20). Thus, potentially wrong segment associations are filtered out based on the consistency of the relative positions of the segments in each scan.

This method does not require the scene to be entirely static but the majority of its elements to be so. The method has been observed to be robust in situations where a few cars are moving in an otherwise static scene. In this case, the method tends to remove associations involving moving segments. On the other hand, in the case of traffic on a highway for instance, correct associations of the static parts of the scene may be removed since the majority of the scene is consistently moving. This results in incorrectly filtering out some of the initially correct associations. Defining the point at which the method starts breaking between these two types of situations is left for future work.

Other approaches for enforcing geometric consistency in feature matching have been proposed in particular in the context of SLAM [5]. Available procedures include Branch and Bound algorithms. These build a graph where nodes represent feature associations (across scans); each node being associated to sets of constraints representing distances to

```

Input      :  $A, C_1, C_2$ 
Output     :  $A'$ 
Parameters:  $t_c$ 
1 for  $l_a \in A$  do
2   for  $l_b \in A$  do
3     if  $\{l_a, l_b\}$  already processed then
4       continue;
5      $d_0 = \text{Eucli}(C_1[l_a[0]], C_1[l_b[0]]);$ 
6      $d_1 = \text{Eucli}(C_2[l_a[1]], C_2[l_b[1]]);$ 
7     if  $|d_0 - d_1| \leq t_c$  then
8        $\text{consistentCount}[l_a[0]] + = 1;$ 
9        $\text{consistentCount}[l_b[0]] + = 1;$ 
10    else
11       $\text{nonConsistentCount}[l_a[0]] + = 1;$ 
12       $\text{nonConsistentCount}[l_b[0]] + = 1;$ 
13    end
14  end
15   $A' = A;$ 
16  for  $e \in \text{nonConsistentCount}$  do
17     $ncc = \text{nonConsistentCount}[e];$ 
18     $cc = \text{consistentCount}[e];$ 
19    if  $ncc \geq cc$  then
20       $\text{Del}(A'[e]);$ 
21  end

```

Algorithm 3: Function GeoConsistency

neighbouring features (in the same scan). Nodes with consistent sets of constraints are linked and the maximum clique is searched for; the latter corresponding to the largest set of geometrically consistent feature associations. This approach was tested with an implementation relying on a randomised gradient descent for the maximum clique search [5]. It was observed to occasionally return a clique that was not the largest one, and did not contain enough associations to compute the alignment while the proposed approach would return the largest set of associations. Exact solutions to the maximum clique problem are available [7]; these are left for future work. RANSAC based methods [9] represent another alternative for enforcing geometric consistency. These would be computationally expensive since they would require re-running ICP on the set of segments at each RANSAC iteration. A faster implementation of a RANSAC based alignment may use segments' centroid only. However, the drift in the centroids' position due to the changes in view point (see Sec. IV-B.1) may result in false associations. This approach would share a number of similarities with the one developed in [25] with the difference that it is applied to 3D data instead of 2D scans and that segmentation is performed prior to alignment as opposed to combining the two processes. RANSAC based alignment will be further investigated in future work. In conclusion, the sequence of operations implemented by the successive application of Algorithm 2 and 3 allows to associate features based on their proximity, on their shape and on the consistency of their relative positions. To the best of the authors' knowledge, this had not been proposed before.

D. Alignment of Sets of Segments

Once the set of segment associations is defined, a 6 DOF point-to-point ICP [22] routine is used to compute the alignment. In order to exploit the now known segment associations the nearest neighbour search in ICP is constrained to be performed only in associated segments. This provides a crucial improvement on a standard ICP procedure on the full scans; in particular when the distance between these two scans increases. This is illustrated in Fig. 3. The intuition behind this improvement is the following: when the two scans are separated by a small displacement, nearest neighbour as implemented by ICP is enough to solve data association. However, when the distance between the two scans increases, nearest neighbour search is less likely to associate corresponding objects, hence the need to explicitly reason at the object level in the data association process. The proposed segment matching algorithm (Algorithm 2) is an attempt to perform such object level reasoning. Its performance is quantified in Sec. IV. All versions of ICP tested here are run for 20 iterations (which in all cases was observed to be enough for the ICP residual to stabilise).

III. EVALUATION METRIC

The metric used here to evaluate the quality of the alignment attempts to capture the crispness of the corresponding point cloud. This metric simply consists of voxelising the point cloud and returning the number of occupied voxels. The lower this number, the more crisp the point cloud and in turn the more accurate the alignment. This is illustrated in Fig. 4.

All algorithms tested here are variants of ICP which provides a measure of alignment quality expressed as a residual (i.e. the root mean square of the nearest neighbour distances between scan points). This measure was not used in this study as an evaluation metric for the two following reasons. (1) The points used to compute the ICP residual are not necessarily all the points of both scans but only the ones matched via nearest neighbour association. This implies that some alignments may correspond to a relatively low ICP residual but may at the same time be incorrect. This was in particular observed when processing scan segments (as opposed to whole scans) in the experiments presented in Sec. IV. (2) The ICP residual cannot be directly computed for a set of aligned scans at once but would have to be averaged in a pairwise manner. This may mask occasional incorrect mis-alignments, which is not desirable in this study and is avoided by the proposed metric.

IV. EXPERIMENTS

This section provides an evaluation of the proposed techniques against standard alignment methods. The datasets chosen for evaluation are first described. The techniques we compare and the way the comparison is performed are then detailed. Finally, experimental results are presented.

Name	ID	Num scans	Distribution of Objects	Platform Speed	Description
ACFR lawn	1	100	intermediate	5 km/h	turns on University lawn
Homebush	2	100	sparse	30 km/h	straight road through the Olympic park
Campus car park	3	200	dense	5 km/h	straight line and turn in a car park
Camperdown	4	200	dense	5 km/h	going around a roundabout
Royal	5	100	dense	30 km/h	straight line past the Royal pub
Usyd entrance	6	100	sparse	20 km/h	bent street past Sydney University entrance
Abercombie	7	200	dense	30 km/h	straight line in residential area

TABLE I DATASETS

A. Evaluation Protocol

The seven sequences used for evaluation are described in Table I. They were acquired using a Velodyne sensor mounted on a moving platform. The sequences are chosen to cover a variety of platform speeds, levels of clutter, and so that they contain a limited number of moving objects. In the configuration used during data acquisition, the Velodyne produced scans at 10Hz. Since the sequences chosen here contain either 100 or 200 scans, they correspond to a duration of 10 or 20 secs.

Four variants of ICP are used as benchmarks in this evaluation. ICP point-to-point and ICP point-to-plane [22], are applied to either full scans or scans from which the ground points have been removed. The resulting four variants correspond to columns 1 to 4 in Tables II and III. ICP point-to-point and ICP point-to-plane are referred to as “p2pt” and “p2pl” in Tables II and III. “grd rm” refers to ground point removal.

Two different processes are applied to each dataset: (1) alignment computed without skipping scans between two scans aligned (Table II), (2) alignment computed by skipping four scans between the current scan and the one it is aligned to (Table III). In the latter case, the displacement between two pairs of scans is larger implying that data association is more challenging. This second configuration enables testing of the performance as the displacement between scans increases.

For both types of processing, the quality of the computed alignments is quantified as follows. The set of aligned scans is projected in a global voxel grid to compute the crispness metric described in Sec. III. In the case of alignments computed without skipping scans, only every fifth scan is projected into the voxel grid so that the set of points used to compute the crispness metric is exactly the same as for the evaluation of the alignments computed every fifth scan. This implies that the numbers in Table II and III are directly comparable. As explained in Sec. III, the smaller this number, the better the alignment (i.e. the more crisp). The ground points are not included when evaluating the crispness of an alignment since well aligned ground segments may not correspond to a reduced number of occupied voxels (and result in an improvement of the alignment quality metric) due to the scanning pattern of the Velodyne sensor. Segmentation is computed beforehand, individually for each scan.

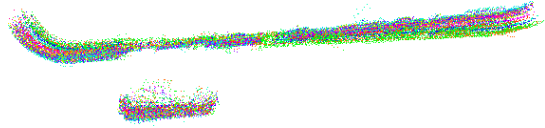


Fig. 2. Example of drift observed in the alignment. This figure shows a top view of a wall and a car in twenty of the scans of the Homebush sequence after alignment with p2pt + seg match (every fifth scan displayed). One colour corresponds to one scan. While segment-to-segment associations are correct, local changes in shape (due to variations in view point) cause a drift in the alignment.

B. Results

1) *Alignment Without Skipping Scans:* The alignment results obtained without skipping scans between two scans aligned are presented in Table II. The label “seg match” refers to the segment association mechanism in Algorithm 2, function `Hungarian`. The label “geo consis” refers to function `GeoConsistency` in Algorithm 3. The column “p2pt + seg match” refers to Algorithm 2 being run without lines 6, 11, 12. The column “p2pt + seg match + geo consis” refers to Algorithm 2. The numbers in bold indicate the best results for each dataset.

It can be seen that ICP p2pl (point-to-plane) systematically provides the best result. The alignment quality reported in the last two columns (corresponding to the proposed methods) is less than the one produced by ICP p2pl. This comes from a drift accumulated at each pairwise alignment (illustrated in Fig. 2). This drift is due to the segmentation not being exactly the same from one scan to the next; variations in segmentation being caused by view point changes across the sequence. Sets of slightly different segments are being aligned which artificially pulls them onto each other in areas where their surfaces have no actual overlap. Segment associations are not incorrect (as illustrated in Fig. 2) but the corresponding alignment is drifting due to variations in the segments shapes.

This problem is less significant when processing full scans with ICP p2pl for the two following reasons. On one hand, the additional (structural) information contained in a full scan (as opposed to a sub-set of its segments) allows to better constrain ICP and leads to a sharper alignment. On the other hand, the displacement between each scan is small enough that data association can be correctly solved with standard nearest neighbour search. Both aspects combined allow ICP p2pl to perform best. In the following section, we show that this is no longer true when the displacement between aligned scans increases. The best results are then obtained by combining segment matching for better data association and ICP p2pl on full scans for refining the alignment. ICP p2pl here also performs better than the p2pt version since most of the scans in the seven sequences contain a majority of planes and geometrical shapes, which ICP p2pl is better suited for.

2) *Alignment With Skipped Scans:* The alignment results obtained when skipping four scans between two scans aligned are presented in Table III. The numbers in bold indicate the best results for each dataset (not including

Set ID	p2pt	p2pl	p2pt grd rm	p2pl grd rm	p2pt seg match	p2pt seg match geo consis
1	38,657	28,926	31,398	29,142	41,404	39,689
2	82,480	61,431	65,834	64,231	70,518	71,224
3	140,616	99,548	125,029	102,675	122,265	123,648
4	186,564	117,677	120,748	120,972	139,803	143,874
5	86,369	80,420	81,223	83,172	84,711	85,334
6	170,337	74,703	170,405	79,058	88,178	87,608
7	435,214	207,207	434,214	214,253	245,467	245,231

TABLE II ALIGNMENT RESULTS: CRISPNESS METRIC; SKIP 0.

the last column which is discussed below). This set of results shows that when the displacement between two scans increases to a point where nearest neighbour search no longer provides correct data association, alignment by reasoning at the object level by explicitly matching objects is significantly more accurate. This is verified with the Homebush, Royal and Abercombie datasets (sets 2, 5, 7) since as indicated in Table I these three datasets correspond to a higher speed of the platform implying that the displacement between two scans is larger. The way nearest neighbour search fails in these datasets and the way the proposed approach correctly recovers the transform are illustrated in Fig. 3.

ICP p2pl provides the best results in the case of the Camperdown and Usyd entrance datasets (sets 4 and 6). In the former the vehicle is moving slowly since it is going around a roundabout. In the latter, the distribution of objects in the scene is sparse making data association less challenging. Both situations are favorable to obtaining a crisp alignment with ICP p2pl, which is confirmed by row 4 and 6 of Table III. In other areas, where the vehicle is moving slowly but where the distribution of objects is denser (ACFR lawn and Campus car park datasets, sets 1 and 3), removing the ground points provides enough separation between landmarks to decrease data association ambiguities and obtain improved alignments. As indicated in Table III, ICP p2pt with ground removal provides alignments with the best quality in the case of the ACFR lawn and Campus car park datasets. However, when the vehicle is moving faster (Homebush, Royal and Abercombie datasets, sets 2, 5, 7), explicitly matching segments becomes necessary to accurately recover the transformation between two scans. As shown in Table III and as illustrated in Fig. 3, ICP p2pt with segment matching and geometric consistency test leads to the best quality alignments for the Homebush, Royal and Abercombie datasets. These results provide an experimental demonstration of the main claim of this paper: reasoning at the object level rather than the point-to-point level allows to significantly constrain data association when the displacement between scans becomes larger, which in turn enables the generation of sharper alignments. A visualisation further illustrating this point is presented in Fig 4.

The alignment provided by ICP p2pt + seg match + geo consis can also be considered as an initial alignment which can be further improved with the additional (structural) information contained in the *full* scan (compared to using only a few scan segments). This is the pipeline quantified in the last column of Table III: ICP p2pl is run (on full scans) with an initial alignment provided by ICP p2pt + seg match + geo consis. This last combination of algorithms leads to

Set ID	p2pt	p2pl	p2pt grd rm	p2pl grd rm	p2pt seg match	p2pt seg match geo consis	p2pt seg match geo consis p2pl
1	49,497	38,677	32,875	36,286	35,037	39,341	29,483
2	85,792	90,190	87,044	87,196	89,430	75,041	66,927
3	120,829	130,631	114,727	131,660	137,655	123,126	90,152
4	185,381	122,367	132,320	145,595	169,708	129,083	110,820
5	103,436	115,489	103,867	114,699	125,867	90,502	81,289
6	149,466	77,351	151,073	85,255	106,253	81,024	70,281
7	369,039	285,248	368,894	291,428	316,228	254,083	224,733

TABLE III ALIGNMENT RESULTS: CRISPNESS METRIC; SKIP 4.

Set ID	p2pt	p2pl	p2pt grd rm	p2pl grd rm	p2pt seg match	p2pt seg match geo consis
1	4.5	9.5	3.3	5.2	6.1	7.0
2	10.7	15.8	3.5	4.9	5.5	5.0
3	6.1	12.0	4.3	6.5	11.3	13.0
4	4.4	7.1	2.6	4.4	4.7	8.0
5	4.1	6.1	3.1	5.2	7.6	8.1
6	3.0	7.1	4.5	5.7	4.6	5.6
7	3.5	7.9	2.9	6.1	4.6	4.7
Average	5.2	9.4	3.5	5.4	6.3	7.3

TABLE IV COMPUTATION TIMES AVERAGED PER SCAN PAIR (IN SECS).

the best alignments for *all* datasets. To explain this result we consider both Table II and III, which contain directly comparable values, as explained in Sec. IV-A. Comparing these values shows that the alignment quality provided by this last pipeline are either close to or better than the best values in Table II². This implies that this last pipeline corresponds to a system requiring only every fifth scan while maintaining the original quality of alignment. More generally, this pipeline is able to maintain alignment quality while the displacement between two aligned scans increases, which is made possible by the segment matching procedure developed in Algorithm 2.

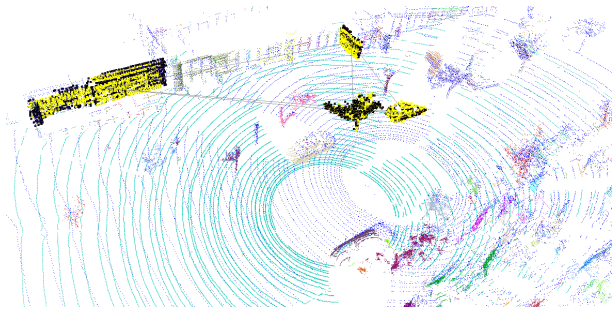
Considering the timing estimates in Table IV³ (bottom row), the pipeline in the last column of table III is computed in about 7.3 sec (ICP p2pt + seg match + geo consis) + 9.4 sec (ICP p2pl), that is 16.7 sec. On the other hand, generating an alignment of the same accuracy with ICP p2pl only requires processing every scan (as opposed to processing every fifth) that is 5×9.4 sec, which is 47 sec. Not only does the proposed method allow the alignment quality to be maintained while the displacement between scans increases, but it also leads to savings in computation times (potentially critical in real-time applications) since scans can be skipped between two processed scans. In the experiments presented here, a 64% savings in computation time is achieved.

V. CONCLUSION

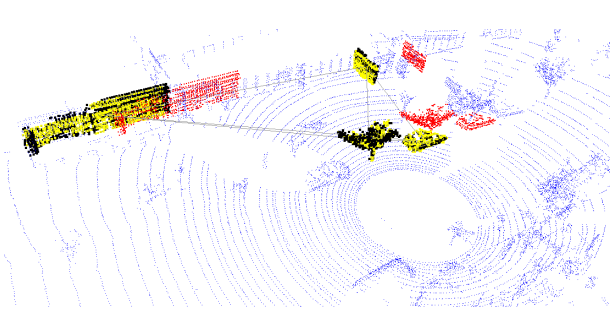
This paper has introduced a method to constraint data association prior to applying ICP to a pair of 3D scans.

²Some results may be better due the maximum nearest neighbour distance used in ICP p2pl which is 20cm here, while it was 2m in all experiments in Table II. A smaller maximum nearest neighbour distance limits the number of possibly incorrect point-to-point associations and may result in a better alignment. Note that a small threshold can only be used after accurate a pre-alignment.

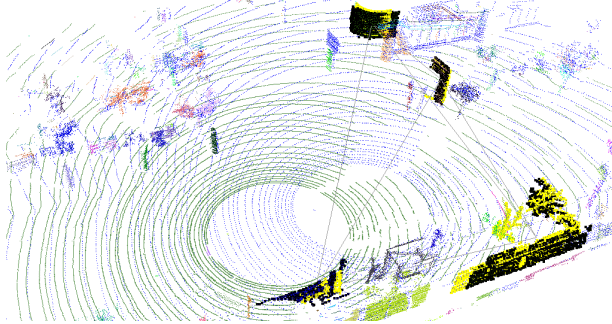
³All computation times are estimated by running Python implementations (with nearest neighbour search performed via KD-trees and implemented in C++) on a laptop Intel Core Duo 2.6GHz. With full C++ implementations the reported computation times can be expected to be one to two orders magnitude smaller. Segmentation time is included in the last two columns; it is on average 0.3 sec per scan.



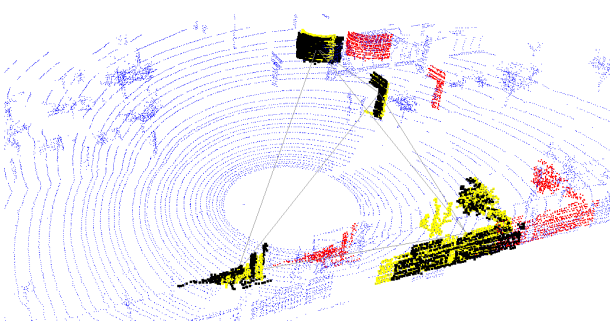
(a) Homebush dataset scan 310; ICP p2pt + seg match + geo consis; $t_d = 0.35$, $t_c = 0.35$; 3 segment associations were removed via geo consis



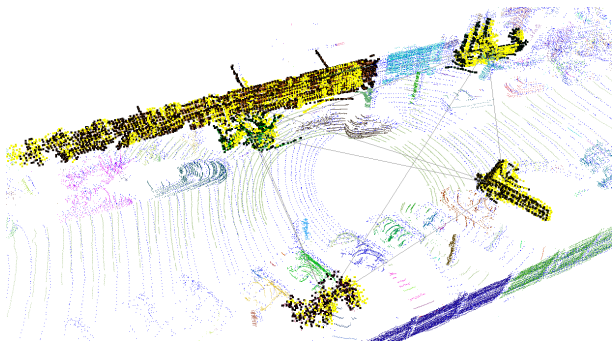
(b) Homebush dataset scan 310; ICP p2pl



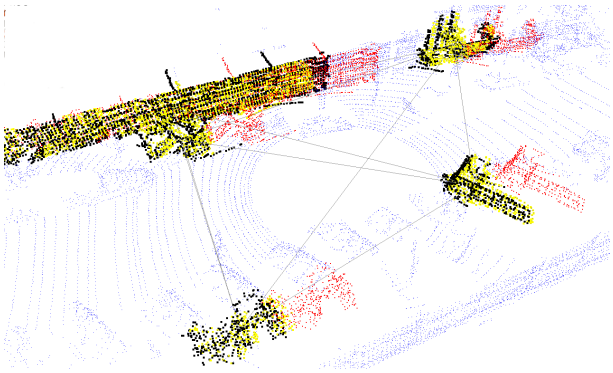
(c) Homebush dataset scan 355; ICP p2pt + seg match + geo consis; $t_d = 0.63$, $t_c = 0.63$; 5 segment associations were removed via geo consis



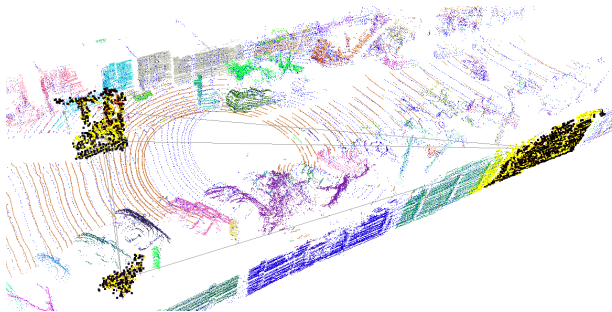
(d) Homebush dataset scan 355; ICP p2pl



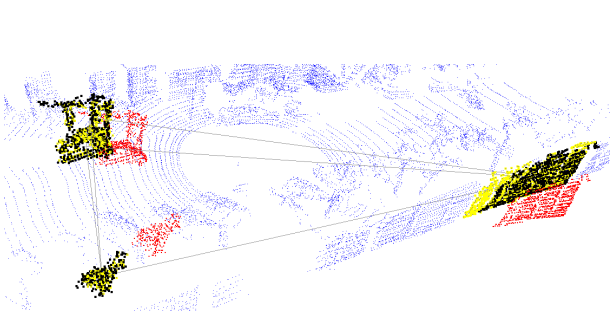
(e) Royal dataset scan 1780; ICP p2pt + seg match + geo consis; $t_d = 0.35$, $t_c = 0.35$; 4 segment associations were removed via geo consis



(f) Royal dataset scan 1780; ICP p2pl

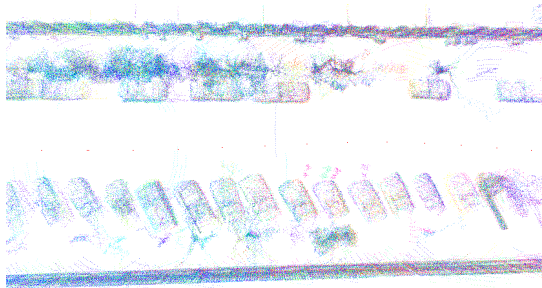


(g) Royal dataset scan 1840; ICP p2pt + seg match + geo consis; $t_d = 0.43$, $t_c = 0.39$; 9 segment associations were removed via geo consis

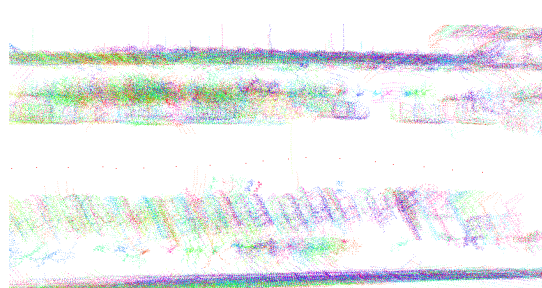


(h) Royal dataset scan 1840; ICP p2pl

Fig. 3. Examples of alignments computed with ICP seg match + geo consis (first column) and ICP p2pl (second column). In the first column, the current scan is coloured by its segmentation and the previous scan is displayed in blue. The segments selected to compute the alignment are indicated by yellow and black points (these two colours respectively indicating the segment selected in the current and previous scan to compute the alignment, i.e. the output of the function `MatchSegments`, Algorithm 2). The links plotted between these segments represent the segment-to-segment distances which were found to be consistent across the scan pair and passed the geometric consistency test (Algorithm 3). The blind spots of the two scans on the ground indicate that the alignment method was able to recover a transformation of several meters. The value of t_d and t_c computed by Algorithm 2 are indicated below each figure as well as the number of segment associations removed by the geometric consistency test (Algorithm 3). In the second column, the previous scan is also displayed in blue, and the red segments indicate the transformation estimated by ICP p2pl (these segments are the same as the ones highlighted in the first column). The segments and the links from the first column are left in this second set of figures to facilitate the visual comparison between the two. It can be seen that the transform estimated by ICP p2pl is off by a few meters while the alignment estimated by the proposed method puts corresponding segments right on top of each others.



(a) Royal dataset; ICP p2pt + seg match + geo consis + p2pl; nvov = 81,289



(b) Royal dataset; ICP p2pt + grd rm; nvov = 103,436

Fig. 4. Visualisation (top view) of the alignment computed by (a) ICP p2pt + seg match + geo consis + p2pl and (b) ICP p2pt + grd rm for the Royal dataset when skipping 4 scans between each scan aligned. One colour corresponds to one scan; a total of 20 scans is displayed; ground points are removed for clarity. The number “nvov” indicates the value of the alignment quality metric and it can be seen that a sharper point cloud corresponds to a smaller value of nvov. While the combination seg match + geo consis allows the generation of a crisp alignment (a), which can be seen for instance in the neat delineation of the cars on the side of the road, ICP p2pt + grd rm (which corresponds to the best result amongst the techniques not using segment matching) generates a much fuzzier point cloud due to data association issues. In both figures, the thickness of the bottom wall on the side of the street is not due to mis-alignment but to the perspective of the display.

This method exploits the segmentation of each scan by explicitly matching segments across scans. Segments are matched based on their proximity, their shape and the consistency of their relative positions. The method was evaluated using seven sequences of Velodyne scans acquired in urban environments. Unlike various other standard versions of ICP which fail to recover correct alignment when the displacement between scans increases, the proposed method is shown to be robust to displacements of several meters. In addition it is shown to lead to savings in computational times which are potentially critical in real-time applications. Future work will concentrate on comparing more extensively various approaches for enforcing geometric consistency, including RANSAC based methods and Branch and Bound algorithms. The proposed approach will also be applied to 3D object modelling via scan registration.

VI. ACKNOWLEDGEMENTS

This research was supported by the Centre for Intelligent Mobile Systems (CIMS) funded by BAE Systems as part of an ongoing partnership with the University of Sydney, by the New South Wales State Government, by the Australian Research Council (ARC) through the Discovery programme, and by the Australian Centre for Field Robotics at the University of Sydney.

REFERENCES

- [1] D-Imager. panasonic-electric-works.net/D-Imager.
- [2] High Definition LIDAR Velodyne. www.velodyne.com/lidar/.
- [3] Kinect sensor. www.xbox.com/en-us/kinect.
- [4] SwissRanger. www.mesa-imaging.ch.
- [5] T. Bailey. *Mobile Robot Localisation and Mapping in Extensive Outdoor Environments*. PhD thesis, Australian Centre for Field Robotics, University of Sydney, 2002.
- [6] P.J. Besl and H.D. McKay. A method for registration of 3-D shapes. *IEEE Transactions on pattern analysis and machine intelligence*, 14(2):239–256, 1992.
- [7] I.M. Bomze, M. Budinich, P.M. Pardalos, and M. Pelillo. The maximum clique problem. *Handbook of combinatorial optimization*, 4(1):1–74, 1999.
- [8] Heiko Blow and Andreas Birk 0002. Spectral registration of volume data for 6-dof spatial transformations plus scale. In *ICRA*, pages 3078–3083. IEEE, 2011.
- [9] C.S. Chen, Y.P. Hung, and J.B. Cheng. A fast automatic method for registration of partially-overlapping range images. In *Computer Vision, 1998. Sixth International Conference on*, pages 242–248. IEEE, 1998.
- [10] B. Douillard, J. Underwood, N. Kuntz, V. Vlaskine, A. Quadros, P. Morton, and A. Frenkel. On the Segmentation of 3D LIDAR Point Clouds. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2010.
- [11] B. Douillard, J. Underwood, V. Vlaskine, A. Quadros, and S. Singh. A Pipeline for the Segmentation and Classification of 3D Point Clouds. In *Proc. of the International Symposium on Experimental Robotics (ISER)*, 2010.
- [12] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox. Rgb-d mapping: Using depth cameras for dense 3d modeling of indoor environments. In *the 12th International Symposium on Experimental Robotics (ISER)*, 2010.
- [13] Y. Jiang, S. Moseson, and A. Saxena. Efficient grasping from rgb-d images: Learning using a new rectangle representation. *ICRA*, 2011.
- [14] A. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 25(1):433–449, 1999.
- [15] S. Kim, C. Jho, and H. Hong. Automatic registration of 3d data sets from unknown viewpoints. In *Workshop on Frontiers of Computer Vision*, pages 155–159.
- [16] H.W. Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.
- [17] J. Levinson, J. Askeland, J. Becker, J. Dolson, D. Held, S. Kammel, J. Kolter, D. Langer, O. Pink, V. Pratt, M. Sokolsky, G. Stanek, D. Stavens, A. Teichman, M. Werling, and S. Thrun. Towards fully autonomous driving: Systems and algorithms. In *IEEE Intelligent Vehicles Symposium*, 2011.
- [18] T. Masuda. Object shape modelling from multiple range images by matching signed distance fields. In *3D Data Processing Visualization and Transmission, 2002. Proceedings. First International Symposium on*, pages 439–448. IEEE, 2002.
- [19] F. Moosmann, O. Pink, and C. Stiller. Segmentation of 3D Lidar Data in non-flat Urban Environments using a Local Convexity Criterion. In *IEEE Intelligent Vehicles Symposium*, pages 215–220, 2009.
- [20] Paul Newman, Gabe Sibley, Mike Smith, Mark Cummins, Alastair Harrison, Christopher Mei, Ingmar Posner, Robbie Shade, Derik Schrter, Liz Murphy, Winston Churchill, Dave Cole, and Ian Reid. Navigating, recognising and describing urban spaces with vision and laser. *The International Journal of Robotics Research*, 28, October 2009.
- [21] D. Rao, Q.V. Le, T. Phoka, M. Quigley, A. Sudsang, and A.Y. Ng. Grasping novel objects with depth segmentation. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 2578–2585. IEEE.
- [22] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In *Proc. 3DIM*, pages 145–152, 2001.
- [23] J. Salvi, C. Matabosch, D. Fofi, and J. Forest. A review of recent range image registration methods with accuracy evaluation. *Image and Vision Computing*, 25(5):578–596, 2007.
- [24] L. Silva, O.R.P. Bellon, and K.L. Boyer. Enhanced, robust genetic algorithms for multiview range image registration. In *3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on*, pages 268–275. IEEE, 2003.
- [25] S.W. Yang, C.C. Wang, and C.H. Chang. Ransac matching: Simultaneous registration and segmentation. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, pages 1905–1912. IEEE, 2010.