Alignment and 3D Scene Change Detection for Segmentation in Autonomous Earth Moving

Julian Ryde
Computer Science and Engineering
University at Buffalo, Buffalo, NY, USA
Email: jryde@buffalo.edu

Nick Hillier
Autonomous Systems Laboratory
CSIRO ICT Centre, Brisbane, Australia
Email: nick.hillier@csiro.au

Abstract—The tasks of region or object segmentation and environment change detection in a 3D context are investigated and tested on an autonomous skid steer loader. This is achieved through a technique analogous to background subtraction utilising 3D scan data which is first aligned before a voxel subtraction operation against a prior map. We highlight the close relationships between the scan-to-map alignment, background subtraction and 3D scan-to-map matching problems.

The presented approaches take advantage of previous work on the multi-resolution occupied voxels list (MROL) representations for 3D spatial maps. This prior work is augmented to provide a mechanism for fast local 6DOF alignment with the same data structures (MROL) that have previously been shown to allow efficient global localisation. The new approach is then compared to an iterative closest point (ICP) implementation and was found to execute in a similar amount of time, but was more robust and accurate. The local alignment algorithm is inherently more amenable to the MROL representation with an associated reduction in implementation complexity and negligible parameter tuning. The hash value basis of MROL results in a map representation that can be both updated and queried in constant time regardless of mapped volume.

The approach described was tested on an autonomous skid steer loader as part of the dig-planning process by segmenting piles of material and detecting humans in the scene.

I. Introduction

It can be observed that during the alignment of 2D scans or other point-cloud information to an existing representation of the environment (the scan-to-map matching problem) that any time-variant geometric information is captured as outliers in the alignment process. It is proposed here to use such outlier information in conjunction with a defined prior representation of the world to identify environment change, and thus enable segmentation in a manner similar to that of background subtraction often employed in the segmentation of 2D image data.

Here we utilise an efficient geometric representation for 2D and 3D environments, which also lends itself to the above segmentation technique. The occupied voxel list (OVL) and its extension through the multi-resolution occupied lists (MROL) [1] class of data structures provide support for representations of N-D environments. Such representations are highly versatile for a variety of mobile robot tasks and can be formulated to be highly memory efficient, provide a means for lookups that are constant time complexity independent of mapped volume, and through the use of non-cubic voxelisations, can

be shown to provide more accurate geometric representations than traditional grid-based representations [2].

These traditional representations, such as occupancy grids or similar voxelised representations in higher dimensions, do not scale well to 3D or even large 2D environments, with large memory requirements and query times that typically increase linearly with mapped volume (or area). Optimized data structures, such as OCT-trees and KD-trees working on similar traditional representations show improved performance but have associated overheads and implementation complexity that OVL based representations lack.

Extensions of the presented technique could see application in a number of other robotics related problems including object classification by matching the outliers against other maps and removal of artifacts due to dynamic obstacles in long-term mapping applications.

In this work, the above described segmentation approach is pursued through the development and testing of an iterative alignment process which suits occupied voxel lists. The method is compared to the popular Iterative Closest Point (ICP) [3] approach, the MROL based efficient global search algorithm presented in [1] and exhaustive searching, for accuracy, execution speed and applicability to the proposed segmentation method.

II. Background

This paper builds upon the concepts of scan matching and the OVL and MROL environment representations.

A. Occupied voxel lists (OVL)

An occupied voxel list (OVL) is a list of voxels that are occupied. Such a data structure is implemented as an associative array. The keys of which are voxel integer tuples and the value associated with each key is the occupancy, i.e. number of sensor observations for that voxel, for example; $(2, 3, 4) : 2, (23, 54, 43) : 1$. For a 3D coordinate X in a Cartesian system using a cubic lattice, the integer voxel coordinates, A are given by

$$A = \lfloor X/\epsilon + 0.5 \rfloor, \tag{1}$$

with ϵ being the edge length of the cube (map resolution).

A multi-resolution occupied voxel list (MROL) is a series of occupied voxel lists maintained at varying resolutions. In

this instance voxel sizes change by a factor of two. For more details on both OVL and MROL see [1].

B. Scan matching

Whilst traditional range-sensor based survey and map building tasks have usually employed a high-end integrated navigation solution and added scan data to a map based on the provided pose, modern approaches utilise scan-matching methods to simultaneously provide the sensor pose and align the scan data in the map. The field robotics literature is replete with 2D scan-matching and less so with 3D scan matching works. At this point it is important to distinguish between scan-scan matching and scan-map matching.

In scan-to-scan matching, each scan is matched to its previous scan, possibly with an initial transformation estimate obtained through a motion model, odometry or another proprioceptive sensor such as an IMU. Unfortunately, with such approaches, not only do pose errors accumulate in an unbounded fashion, the error accumulation rate increases with the frequency that the scan-to-scan matches are performed.

In scan-to-map matching each scan is aligned to a continuously updated map and therefore requires both constant time updates and queries irrespective of map size. Once aligned the map is immediately updated with data from the new scan. The work of Wulf [4] is primarily concerned with the evaluation of 6D SLAM algorithm performance, however presents a good summary of a class of scan-to-map methods for 6D SLAM termed *metascan*. This scan-to-map approach has significant advantages. Unbounded drift errors accumulate orders of magnitude more slowly than scan-to-scan matches of the same environment, especially for trajectories involving re-traversal. The reason that this approach is usually avoided is due to the cost of computing the alignment process for each acquired scan against the entire map. However, as has been shown in [1], by using an OVL representation, the map query and update times are independent of map size, thus enabling scan-to-map matching with only a minor performance penalty versus scan-to-scan matching.

C. Background subtraction

The concept of background subtraction in computer vision was the inspiration for this work. The method of background subtraction (BGS) for use in segmenting 2D image data is a developed technique typically utilised in security or other object segmentation and/or tracking applications to monitor statically mounted video feeds. The premise of such methods is that the "background" scene, or an image of some time-invariant state is saved, and new images are compared through an operation termed subtraction. This operation extracts those items in the new images that are different to the original scene. Such a subtraction approach can be compounded by the number of time-variant elements that are inherent in most everyday scenes, such as shadows, wind-induced motion of trees and the like. An introduction and comparison of algorithms for BGS is presented in [5].

The approach for segmentation of data presented here is analogous to the BGS method for image processing in that a "background" state of the scene is stored in a prior map and through the alignment process, the difference between this prior state and the current state are highlighted as outliers. The vast majority of work with background subtraction assumes a static view point, however researchers are turning their attention to mobile view points, [6], [7].

D. Segmentation for excavation

The problem of segmenting the environment into regions to operate for the dig and dump planners on an automated skid-steer loader (see Fig. 1) was the motivating task for this work. The work of Singh [8] and Cannon [9] is still near state-of-the art for robotic excavation. More recently Schmidt et. al. [10] give a good overview of a modern approach to the dig and dump planning processes. They use a 2.5D (height-field) representation of the environment constructed from a registered point cloud and compared with a desired surface profile to obtain the material to be removed in the dig process. In this manner the premise of the afore mentioned BGS approach is introduced into a quasi 3D context: i.e. a desired surface (background) is subtracted from the current surface profile to identify the material to be removed. The dump location is typically a specified point in space, such as the tray of a waiting truck.

While this works well for well defined operations, such as digging a trench and dumping into a truck or onto a pile it lacks the flexibility required to do more significant landscaping tasks, operate in truly 3D environments (such as underground) and does not take advantage of the environment change information available during the scan matching process. It is the aspiration of the authors that such excavation systems should be able to compare the current environment state with a desired final state and determine dig and dump locations, and perform higher-order operations such as smoothing or grading to autonomously achieve the desired final environment state.

E. State of the art range-based scene change segmentation

The above presented concepts using OVL based representations and outliers of the scan matching process for change identification in 3D data is claimed to be new.

Schutz and Hugli [11] provide background in the application of scene change detection in 3D data sets through a BGS analogous method employing structured light and very close-range scenes. Key works that present arguably similar approaches or results are those of [12], whose patent describes the idea of volumetric differencing for scene change detection and [13]–[15] who are primarily concerned with segmenting moving objects from 3D range data.

In particular, the work of Wang [15] provides elements which are conceptually similar to portions of the work presented here. This work introduces the acronym SLAMMOT (simultaneous localisation, mapping and moving object tracking) to describe the task of performing SLAM whilst simultaneously tracking moving objects. Such a problem is of prime

Fig. 1. Skid steer loader (Bobcat®) autonomously depositing material onto a free standing pile. The sensor payload includes two side laser scanners in the horizontal plane and a rotating laser scanner on top of the cabin for 3D range information.

concern to many autonomous vehicle navigation tasks and is a burgeoning field of study.

III. ALIGNMENT FOR SEGMENTATION

One of the concepts behind this work is to employ the outlier data resulting from the scan-to-map alignment process to obtain scene change information. As such, the alignment process has to be robust to correspondence errors as the method inherently operates on data sets where some scan points may not have counterparts in the map. These unassociated points can be sensor errors, alignment errors, previously unseen parts of the environment or changes to the environment. Assuming that the environment has been explored well and alignment is good, then the remaining unassociated points are likely to correspond to changes in the environment. In this manner it can be observed that the alignment process for scan matching and the proposed segmentation method are intimately entwined. The MROL based pose estimation method presented in [1] provides a rapidly converging global scan-to-map matching method for localisation. It is most suitable for application when a pose estimate in a large pre-mapped area is required with poor or no a priori information on the pose. These re-localization situations are also typical for unassisted simultaneous localisation and mapping (SLAM) problems using 3D scan data, whereby, depending on the specifics of the sensor used, the data acquisition process can be relatively slow, or the data is provided in non-contiguous batches (e.g. actuated 2D scanners or mirrors may provide 3D data at less than 1Hz whilst the motions between full scans can be large and unknown).

However, when a good pose prior is available, as is typically the case with sensors that are able collect data of a 3D scene in one frame (e.g. flash LIDAR or stereo vision) and when matching 2D scans to a 3D map, the global search provided by the MROL localisation technique is unnecessarily expensive. Here, significant gains can be achieved with an iterative alignment procedure adapted for occupied voxel lists.

A. OVL optimization alignment

A mechanism for extracting those points which are outliers to the matching process and represent the segmented data is desirable. Although heuristics can be applied to the results of an ICP (or other) alignment, a method is sought which can take advantage of the OVL based representation and constant-time queries and updates afforded by such a representation.

Here, the point-to-point ICP method is used as a benchmark technique for comparison of the alignment for the scan-to-map matching process. The problem being examined here can see application in scenarios where there is a significant lack of overlap information between the prior map and the current scan data. For example if there is a significant environmental change resulting in a large portion of the prior map observed by the range sensor being obscured by the addition or removal of objects or other material. In such situations there is only a minimal overlap between the sensor's field of view and the prior map. An alignment method that is very robust to small overlaps is also useful for dynamically extending the map beyond an original seed region, giving flexibility in application to the unconstrained (no loop-closing) SLAM problem through scan-to-map matching.

Furthermore, the authors like to view the problem of map maintenance and alignment as a study of bounded loss compression. This approach allows us to define a performance criterion that is linked to an information theoretic background. Thus we define better performance in the alignment process (when an occupied voxel list or similarly derived representation is used) as an increase in the intersecting set of occupied voxels from the initial pose estimate. This captures the essence of the alignment problem when adapted to the OVL class of representations and thus it makes sense to base an objective function around this criterion. For a stationary environment, no segmentation should take place, and thus the intersection of occupied voxels from the initial pose estimate should reach a maxima when the scan data is aligned. For a segmented region, the performance criteria should still be maximised to represent the best alignment of those scan data that lie in the prior map.

While in the limit, the nearest-neighbour distance criterion employed in ICP should approach the same performance, it is operating in a different metric space and may not converge to the same solution that an alignment algorithm based on a cost function that more closely represents the performance criterion.

The proposed approach is to set up a cost function, C, that returns the number of intersections (collisions), I, between the scan data at a given sensor pose and the map voxels, (2). In practice, the value of the cost function when computed in such a manner is the sum of the individual intersections, i, found via the hash collisions in the underlying representation of the OVL and the N points from the point cloud to be aligned. This provides information on the outliers (all the non-intersecting scan points, or the data to be segmented) for a given pose as an inherent property of evaluating the cost function or performing

| Align. | Mean point | Alignment | Alignment |
Method	dist. (m)	Overlap	Time(s)
	Good Pose Prior		
Point to point ICP	0.112	1842.533	0.844
OVL	0.111	2044.0	0.313
MROL (5 levels)	0.112	2040.0	1.709
	Poor Pose Prior		
Point to point ICP	0.184	911.4	0.861
OVL	0.119	1766.0	0.67
MROL (5 levels)	0.12	1670.0	3.298

TABLE I

COMPARISON OF THE ALIGNMENT METHODS FOR A GOOD (<0.5M ERROR) AND DEGRADED (>2M ERROR) INITIAL POSE ESTIMATE FOR A SCENE WITH A HIGH PROPORTION OF CHANGE ($>25\%$ OF POINTS), AVERAGED OVER 15 RUNS. AN EXHAUSTIVE SEARCH GAVE A BEST POSSIBLE OVERLAP OF 2122 OUT OF A TOTAL OF 2726 SCAN POINTS. THE PRIOR MAP CONTAINED 32,000 CELLS AT 0.1M RESOLUTION.

the alignment process.

$$C = -I; I = \sum_{c=0}^{c=N} i_c \qquad (2)$$

The cost functional computed in this manner has varying smoothness dependant on the underlying environmental structure. This can not only lead to situations whereby a local optimization algorithm could become trapped in a non-optimal minima, but also affects the region of convergence of the alignment process. This led the authors to examine a number of alternate cost definitions. We present the results of one making use of the MROL representation by defining a cost function that is a linear summation of the scan to voxel intersections at multiple resolutions, r, (3). This provides a simple objective that requires no parameter tuning other than the selection of the appropriate voxel resolutions (which are usually dictated by sensor capabilities, the shape of the voxels, the desired alignment accuracy and the environmental structure). If desired, it is a simple matter to bias this cost for faster gross alignment or higher accuracy (at the expense of higher susceptibility to local minima) by making (3) a weighted or non-linear summation.

$$C = -\sum_{r=0}^{r \leq R} I_r \qquad (3)$$

B. Solving the cost functionals

As an evaluation of the feasibility of the method, a variant on the Gauss-Seidel gradient descent optimization algorithm was trialled, with 5 quadratically spaced step-size reductions and the gradient in each dimension estimated through a finite-difference evaluation. This approach provides a rapidly converging basis with a known resolution at convergence enabling us to objectively evaluate the alignment performance of the various approaches. More modern optimization methods may be more efficient, but were not investigated here.

C. Segmentation

After alignment, those points that lie within voxels contained in the prior map are removed. The resulting points correspond to the scene change. Depending on the alignment process employed, segmentation can be performed as part of the alignment process by selecting those points that are rejected as outliers (i.e. have no association correspondence in the prior map). Alternatively, this information is easily determined by a look-up against the prior map, for the cost of the collision look-up in the OVL based techniques and for the cost of the ANN association step otherwise.

For additional robustness in the OVL based segmentation methods, at the expense of an increase in the false negative detection rate, an adjacency check is efficiently and easily accomplished by pre-processing the prior map with a binary morphological dilation operation, effectively expanding the voxels in the prior map by one voxel in each direction. The structuring element used here (in a cubic voxelisation) was a 3 by 3 by 3 array of ones. Following alignment against the original (undilated) prior map, the segmentation comparison can be performed against the modified (dilated) prior map. Given the constant query time for OVL representations, the additional cost of this approach is limited to the off-line dilation operation on the prior map, and a small incremental dilation operation on any new data if the prior map is being grown in a SLAM type application.

Similarly, increasing the association distance in an Euclidean distance based segmentation can provide additional robustness. The additional computational cost of using any distance based segmentation method is dominated by the cost of building a search index. This is comparable to the dilation cost in the OVL based method if it is computed offline, however if the prior map is being extended with new data, the index is unable to be incrementally extended by currently available methods (a full index rebuild is required) and thus such approaches do not scale well to large maps.

Fig. 2 and Table II presents sample results of various methods when segmenting a pile of soil from a 3D scene as part of an autonomous earthmoving task.

IV. IMPLEMENTATION

A number of experiments were conducted to examine the feasibility and performance of the proposed alignment and segmentation processes.

A. An optimization to the alignment methods

Alignment speed can be significantly accelerated by reducing the number of points to be aligned through sub-sampling. Typically this may be done by voxelizing the scan before alignment, random sampling of the scan or decimation. However, volumetric sampling of the scan using the existing voxelization techniques provides a sampled data set with a more representative spatial distribution and is obtained in an efficient manner. This approach counters the multitude of range-weighting based techniques found throughout the literature to provide a more accurate alignment that is unbiased

Alignment Method	Segmentation	True Positive	True Negative	False Positive	False Negative	Accuracy	Specificity	Alignment Time (s)
					Good Pose Prior			
Point to point ICP	OVL	492.4	1798.6	380.4	54.6	0.84	0.825	0.856
Point to point ICP	OVL Dilated	418.8	2150.6	28.4	128.2	0.943	0.987	0.832
Point to point ICP	Assoc Dist	409.0	2156.2	22.8	138.0	0.941	0.99	0.844
OVL	OVL	516.0	2013.0	166.0	31.0	0.928	0.924	0.312
OVL	OVL Dilated	457.0	2159.0	20.0	90.0	0.96	0.991	0.314
OVL	Assoc Dist	442.4	2163.6	15.4	104.6	0.956	0.993	0.312
MROL (5 levels)	OVL	517.0	2010.0	169.0	30.0	0.927	0.922	1.688
MROL (5 levels)	OVL Dilated	454.0	2158.0	21.0	93.0	0.958	0.99	1.701
MROL (5 levels)	Assoc Dist	441.0	2162.8	16.2	106.0	0.955	0.993	1.737
					Poor Pose Prior			
Point to point ICP	OVL	519.6	765.8	1413.2	27.4	0.472	0.351	0.856
Point to point ICP	OVL Dilated	387.0	1533.0	646.0	160.0	0.704	0.704	0.884
Point to point ICP	Assoc Dist	376.6	1683.8	495.2	170.4	0.756	0.773	0.843
OVL	OVL	513.0	1732.0	447.0	34.0	0.824	0.795	0.667
OVL	OVL Dilated	444.0	2149.0	30.0	103.0	0.951	0.986	0.665
OVL	Assoc Dist	438.2	2154.0	25.0	108.8	0.951	0.989	0.677
MROL (5 levels)	OVL	512.0	1635.0	544.0	35.0	0.788	0.75	3.295
MROL (5 levels)	OVL Dilated	447.0	2158.0	21.0	100.0	0.956	0.99	3.285
MROL (5 levels)	Assoc Dist	438.0	2162.6	16.4	109.0	0.954	0.992	3.313

TABLE II

CONFUSION RESULTS OF THE SEGMENTATION PERFORMANCE AGAINST HAND-SEGMENTED *Truth* DATA, USING THE ALIGNMENT RESULTS AS PRESENTED IN TABLE I. THE SEGMENTATION PERFORMANCE DEPENDS HEAVILY ON ACCURATE SCAN ALIGNMENT AND THE SEGMENTATION METHOD USED.
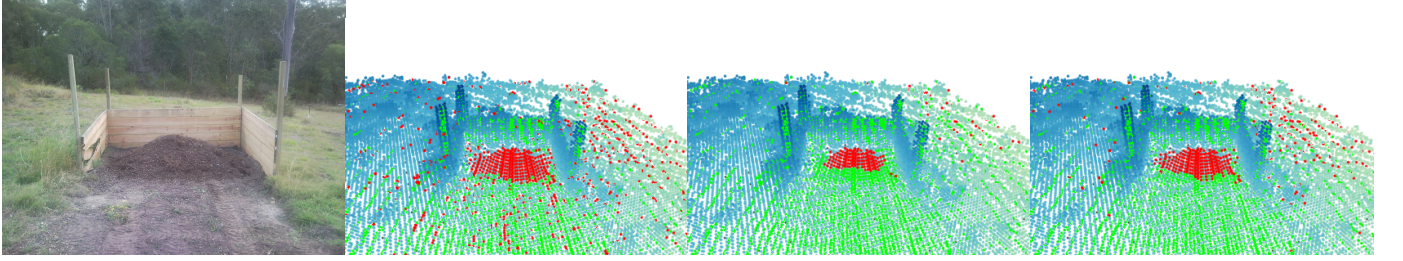

Fig. 2. Comparison of segmentation methods for the same alignment pose using a 0.1m voxel grid, with the prior map coloured by height (blue-green scale), aligned points in green and segmented points shown in red (highlighting a pile of material). From left to right: Photo of the material pile in a wooden enclosure, a direct OVL segmentation on the scanafter alignment, with a dilated prior map, and using using an ANN Euclidean distance threshold (0.1m).

by sample density and was found to provide better alignment results than any other tested sampling method. For the results presented in Table I, the alignment is performed using data subsampled by this volumetric sampling approach, and then further reduced by selecting a random sample from the volumetric sample.

B. Autonomous skid-steer loader

An autonomous skid steer loader (modified Bobcat® S185) used for technology development is displayed in Fig. 1. This vehicle is versatile and is typically deployed on construction sites. Its main defining characteristics are its compact size, ability to turn in place and the positioning of the operator between the boom arms. For this work the authors used a variety of sensors; a rotating laser mounted above the operator cabin provides 3D range data, whilst an Integrated Navigation System (INS) was used to provide initial pose estimates for the alignment process and to provide georeference information. Wheel encoders were used to detect stationary states. The INS was operated in a degraded mode (no Real Time Kinematic corrections) for this work in order to stress the alignment process robustness.

A preliminary task required of the platform is to move piles of material from one location to another, including digging and dumping into a three sided enclosure. As part of the digging process, the pile must be segmented from the surrounding environment in order to compute an extraction plan. The overall process to complete this segmentation is enumerated below.

1) Create and store a map of the empty enclosure.
2) Adopt a pose approximately in front of the enclosure.
3) Acquire a 3D scan using the rotating laser.
4) Align the scan with the empty enclosure map.
5) Scan points coincident with the empty enclosure map are removed.
6) Remaining points (the segmented data of the pile) are analysed to produce an extraction plan.

C. Prior map generation and maintenance

To generate the prior map of the empty enclosure, the Bobcat® was tele-operated to various poses around and in front of the empty material enclosure. It was paused at each pose for a short period (2 seconds) in order to get an accurate 3D point cloud from the rotating laser. Adoption of a number

of variant poses, including some from behind the enclosure, although not entirely necessary, help in making the prior map more complete. This in turn improves the efficacy of the segmentation subtraction. For enhanced accuracy the 3D scans used for prior map generation were only extracted when the encoders report no change, i.e. that the vehicle is stationary.

These scans were iteratively aligned (see Section III) to the map and co-registered to produce an empty enclosure prior map, which was then saved.

During the course of its missions, the skid steer loader adopts a pose in front of the material enclosure and obtains 3D range data of the enclosure and any contained material. This scan is aligned with the map as with the prior map creation, using the INS pose as an initial guess, to initiate the segmentation process. This alignment and segmentation could be conducted during dynamic motion if an appropriately accurate mapping solution were employed.

Another map is updated periodically during autonomous operation (for example, when leaving the enclosure following a dig or dump operation) to provide a comparison for highlighting unexpected scene changes (such as personnel entering the workspace). An example segmentation from this application is presented in Fig. 3.

V. Results and Discussion

Tables I and II present a summary of the relative alignment and segmentation performance characteristics for a sample of examined alignment and segmentation methods. As previously discussed, the maximum overlap between the scan and the prior map is used as a reference criterion for performance assessment. The typical metric used for assessment of ICP alignment quality, the thresholded mean Euclidean distance between associated points, is also listed for comparison. The presented scenario is for the scene shown in Fig. 2, with additional cropping of the point cloud increasing the proportion of segmented points in the scan, to demonstrate the algorithms' performance in situations with a high proportion of scene change.

Examination of the alignment metrics shows that the OVL and MROL based alignments are of comparable performance to ICP, with the final overlap (the performance metric for which we are primarily concerned) slightly better for the OVL and MROL alignments. The MROL approach does not appear to provide benefit (considering the additional cost) until it is utilised with a degraded initial pose estimate. The MROL method's robustness to association type errors and local minima over the single resolution OVL and ICP alignments and wider region of convergence lead to considerably better segmentation performance for poor pose priors.

Although not presented here, the ICP and OVL alignments showed near identical performance when the proportion of scene change was low (e.g. when aligning scans to static scenes), although the authors stess that the implementation and tuning simplicity of the OVL method over that of ICP is favourable.

Listed in Table II are the three segmentation methods used to identify outliers following alignment. These are: scan points that do not intersect with the OVL prior map (OVL), those that do not intersect with a dilated OVL prior map (OVL Dilated) and those that are beyond a defined Euclidean distance threshold from the prior map voxels (Assoc Dist).

Segmenting directly on the prior map is susceptible to sensor noise, particularly if the content of a voxel is near a voxel boundary or there are thin objects in the scene, such as grass. Comparing against a dilated map effectively filters noisy returns from the segmentation, at the expense of blurring the segmentation region by not classifying points as outliers when they are within one voxel distance from the prior map.

Using an Euclidean distance measure for identifying the outliers is a more traditional method of performing such associations and performs well with high flexibility through the available variation in association distance (here chosen to be the same as a voxel edge (0.1m) for simplified comparison). As per the dilated prior map method, the index building step for an ANN method can be performed offline, however the cost of rebuilding the index if the prior map is being dynamically updated does not scale and is likely to be prohibitive for large scale online use.

VI. Conclusion

We have presented a variety of alignment methods which are amenable to the OVL and MROL class of 3D scene representations and compared these methods against a conventional ICP implementation. The alignment result has been used to provide segmentation through scene change detection in a manner analogous to background subtraction in image processing. The main differences being, that occupied voxel lists rather than images are used, and alignment is done prior to subtraction. The segmentation technique is inherently related to the alignment process and so increasing the alignment performance directly improves the segmentation result.

Although the OVL and MROL alignment methods are better suited to the OVL based map representation than ICP, there is only a slight performance gain in using these approaches over the ICP method for the alignment step. However, the implementation of the OVL based alignment methods is significantly simpler than that typically used for other alignment methods and 3D environment representations whilst having few parameters that require tuning. The most import parameter being the voxel size.

It was found that a variety of approaches could be used to further filter the segmentation data to produce a cleaner segmentation. Those presented here include dilating the prior map through a morphological binary dilation operation and using an Euclidean association distance. Whilst the outright execution speed difference between the two approaches is slight, as the area and/or volume of the mapped environments increase and/or in applications where the map itself is being grown or modified by the scan data, the dilated prior map method is considered advantageous due to lower processing overheads. For instance it does not require rebuilding of
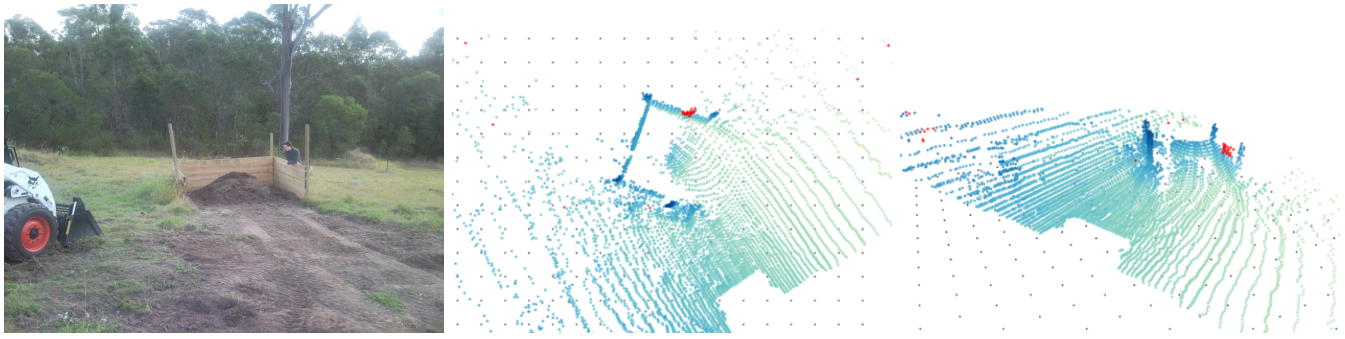
Fig. 3. Example 3D scene change results. Shown is an image of the scene with a mound of material inside a wooden enclosure, and an unexpected scene change, in the form of a person looking into the enclosure. Secondly the 3D laser range scan from different points of view with the segmented points shown in red highlighting the unexpected scene change (the person).

accelerated lookup data structures as required by ICP and simple point cloud representations.

This paper has presented an application of OVL based representations showing the advantages of the representations' memory efficient implementation and constant-time look-ups, as well as the comparatively negligible cost associated with growing the map and performing incremental morphological operations (such as dilation) on the map which can be costly for the ICP method, typically requiring whole-of-map index rebuilds to obtain the same functionality.

A. Future Work

The authors are investigating this segmentation method for wider application, including object segmentation in cluttered indoor scenes and the subsequent classification of these objects as well as object tracking and identification for industrial scenes. Additional refinements to the alignment algorithms are under investigation, including more direct recursive multi-resolution approaches.

Furthermore, the developed alignment methods are being considered for the problem of incremental map building/growing for OVL and MROL representations as well as a variety of scan-to-map localization applications.

Finally the effectiveness of 3D morphological operations such as opening should be assessed for removing isolated voxels that can be present after subtraction.

REFERENCES

[1] J. Ryde and H. Hu, "3D mapping with multi-resolution occupied voxel lists," *Autonomous Robots*, vol. 28, no. 2, pp. 169–185, February 2010.
[2] J. Ryde and M. Brünig, "Non-cubic occupied voxel lists for robot maps," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2009.
[3] P. Besl and N. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.
[4] O. Wulf, A. NÃijchter, J. Hertzberg, and B. Wagner, "Benchmarking urban six-degree-of-freedom simultaneous localization and mapping," *Journal of Field Robotics (JFR), Wiley & Son, ISSN 1556-4959*, vol. 25 issue 3, pp. 148–163, Mar. 2007.
[5] T. H. Chalidabhongse, K. Kim, D. Harwood, and L. Davis, "A perturbation method for evaluating background subtraction algorithms," in *Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS 2003)*, Nice, France, Oct 2003.
[6] Y. Sheikh, O. Javed, and T. Kanade, "Background subtraction for freely moving cameras," in *IEEE International Conference on Computer Vision (ICCV)*, 2009.
[7] Y. Sugaya and K. Kanatani, "Extracting moving objects from a moving camera video sequence," in *10th Symposium on Sensing via Image Information (SSII2004)*, Jun. 2004, pp. 279–284.
[8] S. Singh, "The state of the art in automation of earthmoving," *ASCE Journal of Aerospace Engineering*, vol. 10, no. 4, October 1997.
[9] H. Cannon, "Extended earthmoving with an autonomous excavator," Master's thesis, Carnegie Mellon Robotics Institute, 1999.
[10] D. Schmidt, M. Proetzsch, and K. Berns, "Simulation and control of an autonomous bucket excavator for landscaping tasks," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2010.
[11] C. Schütz and H. Hügli, "Change detection in range imaging for 3D scene segmentation," in *Proceedings of the International Symposium on Lasers, Optics and Vision for Productivity in Manufacturing I*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE), vol. 2786, Besancon, 10-14 June 1996.
[12] V. Sequeira, "Method and system for 3D scene change detection," Portuguese Patent WO 2006/027 339 A2, Mar 16, 2006, g06T 7/00; G06T 17/00.
[13] F. Moosmann and T. Fraichard, "Motion estimation from range images in dynamic outdoor scenes," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2010, pp. 142–147.
[14] R. Katz, J. Nieto, and E. Nebot, "Probabilistic scheme for laser based motion detection," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, september 2008, pp. 161–166.
[15] C.-C. Wang, C. Thorpe, and S. Thrun, "Online simultaneous localization and mapping with detection and tracking of moving objects: Theory and results from a ground vehicle in crowded urban areas," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2003, pp. 842–849.

ACKNOWLEDGEMENTS