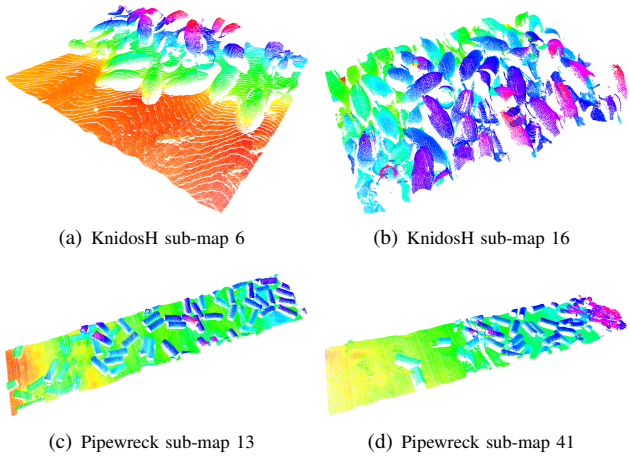
FFT-based Terrain Segmentation for Underwater Mapping

B. Douillard*, N. Nourani-Vatani*, M. Johnson-Roberson*, S. Williams*, C. Roman‡, O. Pizarro*, I. Vaughn‡, G. Inglis‡
* The Australian Centre for Field Robotics, The University of Sydney, Australia
‡ Department of Ocean Engineering, The University of Rhode Island, USA
b.douillard@acfr.usyd.edu.au

***Abstract*—A method for segmenting three-dimensional scans of underwater unstructured terrains is presented. Individual terrain scans are represented as an elevation map and analysed using fast Fourier transform (FFT). The segmentation of the ground surface is performed in the frequency domain. The lower frequency components represent the slower varying undulations of the underlying ground whose segmentation is similar to de-noising / low pass filtering. The cut-off frequency, below which ground frequency components are selected, is automatically determined using peak detection. The user can specify a maximum admissible size of objects (relative to the extent of the scan) to drive the automatic detection of the cut-off frequency. The points above the estimated ground surface are clustered via standard proximity clustering to form object segments. The approach is evaluated using ground truth hand labelled data. It is also evaluated for registration error when the segments are fed as features to an alignment algorithm. In both sets of experiments, the approach is compared to three other segmentation techniques. The results show that the approach is applicable to a range of different terrains and is able to generate features useful for navigation.**

I. Introduction

This paper presents a method for segmenting 3D point clouds representing unstructured/natural terrains. The approach is applied here to underwater bathymetric point cloud data collected as successive range/angle scans using a downward looking structured light laser system on a moving vehicle [17]. Examples of such 3D scans are shown in Fig. 1. Fig. 2 shows an overview of one of the data sets used here. The obtained scan segments are used as key areas (by analogy to key points) in a feature based alignment process to register overlapping scans. The core of the segmentation method lies in a novel ground extraction process. The terrain is represented as an elevation map which is interpreted as a 2D (discrete) signal and analysed using discrete Fourier transform (DFT). In the frequency domain, the lower frequency components represent the slower varying undulations of the underlying ground and their extraction is similar to denoising/low pass filtering of the overall signal. The cut-off frequency, below which the frequency components associated with the ground are selected, is automatically determined. This is done using peak detection on the frequency signal together with the insight that a given frequency directly maps to a feature size in the spatial domain. A user can also specify a maximum admissible objects size (relative to the extent of the scan or as a metric unit) to drive the automatic detection of the cut-off frequency. The segmentation process is fast since a single DFT represents the bulk of the calculation. It also does not require any training. The output of the process is an estimated ground surface. The notion of ground surface as used in this paper is to be understood as the background terrain undulations. Segments are formed from the non-ground points above this surface by applying a standard voxel based clustering (i.e. clustering by direct connectivity of non-empty voxels). These segments are used as features (or key areas) for alignment.



(a) KnidosH sub-map 6 (b) KnidosH sub-map 16

(c) Pipewreck sub-map 13 (d) Pipewreck sub-map 41

Fig. 1. Sub-maps which were manually labelled for the evaluation (colours mapped to height). The oblong shapes in the top row correspond to amphora on a shipwreck site. The cylindrical shape in the bottom row correspond to pipe segments found on another shipwreck site. The data sets are further detailed in Sec. V-A. The extent of the top sub-maps are $[4\,\mathrm{m} \times 4\,\mathrm{m}]$; the extent of the bottom sub-maps are $[6\,\mathrm{m} \times 1.5\,\mathrm{m}]$.

Underwater platforms are increasingly being used for high-resolution mapping tasks in scientific, archaeological, industrial and defence applications. Accurate 3D scan registration is critical to high resolution map building using dead reckoning navigation measurements. The footprint/aperture and effective range of underwater sensors is very limited relative to aerial and terrestrial equivalents. In many cases the survey areas contain a mix of man-made structures (e.g. shipwrecks, pipes) and natural/unstructured seafloor (e.g. algae, corals, sand, canyons) that do not satisfy typical assumptions such as the presence, type or density of features in 3D data observed from terrestrial and aerial robots.

This paper provides experimental evidence for the applicability of the proposed DFT based segmentation method to a range of underwater terrains and for its ability to produce key areas that allow accurate 3D alignment. The segmentation

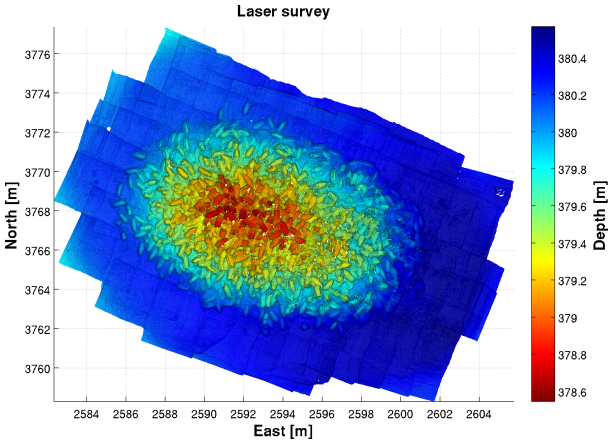


Fig. 2. Overview of the KindosH shipwreck. The site is approximately 16 meters long and 10 meters wide. The dataset is further described in V-A.

is first validated using hand-labelled underwater scans. It is then evaluated in the context of an alignment pipeline and compared to three other segmentation techniques. The proposed segmentation process has potentially a number of applications in marine robotics.

II. Related Work

Segmentation has been studied for several decades and in particular in computer vision, where it is often formulated as graph clustering, such as in Graph Cuts [4]. Graph-Cuts segmentation has been extended to 3D point clouds by Golovinskiy and Funkhouser [8] using k-nearest neighbours (KNN) to build a 3D graph with edge weights assigned according to an exponential decay in length. The method requires prior knowledge on the location of the objects to be segmented.

In ground robot applications, the recent segmentation algorithm of P. Felzenszwalb [12] (FH) for natural images has gained popularity due to its efficiency [19, 21, 22, 24]. Zhu et al. [24] build a 3D graph with KNN while assuming the ground to be flat for removal during preprocessing. 3D partitioning is then obtained with the FH algorithm. Under-segmentation is corrected via a posterior classification that includes the class "under-segmented". Triebel et al. [22] explore unsupervised probabilistic segmentation in which the FH algorithm is modified for range images and provides an over-segmentation during preprocessing. Segmentation is cast into a probabilistic inference process both in the range image and in feature space using Conditional Random Fields. Their evaluation does not involve ground truth data. Schoenberg et al. [19] and Strom et al. [21] have applied the FH algorithm to coloured 3D data obtained from a co-registered camera/laser pair. The weights on the image graph are computed based on a number of features, including point distances, pixel intensity differences and angles between surface normals estimated at each 3D point. The FH algorithm is then run on a graph representing either the range image or the colour image. In practice the evaluation is only done on road segments, or visually. On the contrary, the evaluation presented here is performed on scans with point-wise labels assigned manually. All of the mentioned algorithms reason locally and attempt to identify boundaries between objects in the scene while the approach proposed here reasons globally across the whole depth image by analysing its overall frequency content to find underlying larger scale patterns. Three segmentation techniques for 3D scans acquired in urban environments are proposed in [6]. This suite of algorithms is able to process point clouds of different densities but the algorithms all assume the platform to be located on a drivable surface so that ground seeds are available; ground seeds are not always readily available in the underwater scans processed here (for instance see Fig. 1(b)).

Segmentation of underwater terrain based on acoustic and range data typically has focused on larger scale, lower resolution applications such as generating habitat maps of the seafloor based on multi-beam sonar data [5]. The objective in these cases is to group areas of the seafloor with similar textures (derived from bathymetry or backscatter) as representing a common habitat, each segment representing hundreds or thousands of square meters of seafloor. Another use for segmentation is in acoustic side-scan sonar imagery as part of mine-detection systems [15], where segmentation is used to separate the returns into ground, target and shadow. While the focus in these cases is the detection of particular man-made structures, side-scan data usually has swaths of hundreds of square meters in width and targets appear as clusters of relatively few pixels. Segmentation is also used to aid object detection and tracking for obstacle avoidance [13]. In this case the segmentation is usually applied at the individual sensor readings as a way of dealing with noise or reducing computational demands further along the processing pipeline, rather than considering 3D surfaces observed through multiple scans. To our knowledge this paper is the first to address segmentation of ground and non-ground objects in short-range, high resolution 3D point clouds from underwater scenes.

The contribution of this work lies in the definition of a novel ground extraction method for unstructured terrain. It is based on DFTs and only requires the user to set one parameter which specifies the maximum size of objects (relative to the scan extents) beyond which objects are identified as part of the underlying ground.

III. 3D Alignment

The segmentation process presented in this work is developed as a preprocessing step to allow accurate alignment of underwater scans. The survey of 3D alignment techniques presented by Salvi et al. [18] shows that most approaches involve two main steps: coarse alignment followed by fine alignment. Coarse alignment is the step this study is concerned with since fine alignment is usually addressed with the standard Iterative Closest Point (ICP) algorithm [1]. While this particular paper focuses on dense point clouds acquired with a structured light sensor (sensors and data sets described in Sec. V-A), the long term aim of the study is to develop methods also capable of registering dense point clouds (i.e. obtained from structured light or multi-beam sonar scans) to sparser point clouds (e.g.

shipborne acoustic surveys). Point feature based alignment techniques may not be suitable since the underlying features often require homogeneous sampling across the scan pair. The Spin Image [10] and the NARF [20] features, for instance, have this requirement. An alternative to point feature based approaches are approaches which attempt to match segments of the data [7]. This makes them more directly applicable to scan pairs of different densities. The alignment method in [7] is used here and will be be referred to as S-ICP ('S' for segmentation-based).

IV. Ground-object segmentation

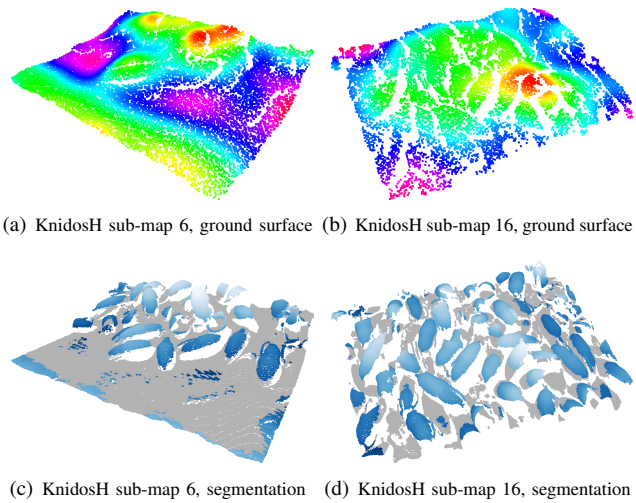
This section describes the core contribution of the paper. A mechanism to perform ground extraction in unstructured terrains is introduced. The proposed segmentation approach requires preprocessing of the point cloud to form an elevation map aligned with the DC component (the zero frequency term) of the elevation signal; this is described in Sec. IV-A. The resulting elevation map is transformed into the frequency domain via DFT and filtering operations are applied to its frequency content to extract the underlying ground surface; this is described in Sec. IV-B. Finally, ground and objects above the ground are separated following the process described in Sec. IV-C.

A. Generation of elevation maps

The aim of the preprocessing steps presented next is to generate a discrete 2D signal on a regular grid whose DC component is aligned with the horizontal plane. The obtained 2D discrete signal can then be transformed to the frequency domain with fast Fourier Transform (FFT). Note that the requirement of a regular grid is core to FFT [11]; this has implications on the type of point clouds that can be processed with the proposed approach. We consider alternative methods for applying these techniques to irregular scans at the end of this section.

To obtain a 2D discrete signal from the input point cloud, the latter is first re-aligned so that the DC component of its elevation is brought into the xy-plane. To do so, the sub-map is moved to its centre of mass. It is then axis-aligned by using a process equivalent to rotating the sub-map onto its eigen vectors and ensuring that positive elevation is along the z-axis.

In this adjusted reference frame, an elevation grid is formed by voxelisation of the point cloud. The maximum height in each column of the voxel grid is kept as the elevation. A resolution of $2\,\mathrm{cm}$ is used in all the experiments reported here. Some of the cells in the elevation grid may not contain any points. The height in these cells is estimated via nearest neighbour interpolation. Interpolation in the data sets processed here is not likely to modify the frequency content of the original elevation signal since empty cells represent at most 10% of the elevation grid. However, this is not true in more sparse scans, such as side-looking Velodyne scans [7] where the polar pattern of the scans induces a significant number of empty cells in a regular grid. Non-trivial interpolation mechanisms would be required to estimate the height in these cells. Alternatively, non-uniform DFTs or Polar FFTs could be used [11]. This is left as future work.



(a) KnidosH sub-map 6, ground surface (b) KnidosH sub-map 16, ground surface

(c) KnidosH sub-map 6, segmentation (d) KnidosH sub-map 16, segmentation

Fig. 3. Top row: examples of ground surfaces extracted with the proposed segmentation method (colours are mapped to height). The two submaps used in this example are also shown in Fig. 1; these have an extent of about $[4\,\mathrm{m} \times 4\,\mathrm{m}]$. Bottom row: resulting segmentation, grey indicates ground points, and blue non-ground points. The estimated ground surfaces (top) are sparser than the point clouds (bottom) since they are built on the elevation grid. In both cases most of the parts of the terrain belonging to the underlying seafloor are correctly identified. This is further quantified in Sec. V-C.

B. Ground surface extraction

The formulation of the ground extraction process presented in this section is based on the observation that the ground surface in a scene can be interpreted as the underlying lower frequency undulations where smaller and higher frequency shapes (e.g. objects on the ground) sit on. This "ground" surface may not physically exist in all natural scenes. For instance, larger objects such as rocks often merge with their support at their boundary to form continuous surfaces. However, the estimation of a separation layer—which we refer to as ground—defined as the lower frequency content of the elevation signal (Figs. 3(a)-3(b)) allows us to identify distinct segments above the "ground" (Figs. 3(c)-3(d)). These segments are then used as landmarks for navigation (see Sec. V-D). Intuitively, this segmentation process also relates to a saliency detection mechanism (and thus information theory); however, the definition of the theoretical relationships between the two is beyond the scope of this paper. In the remainder of this paper, the word "ground" will be used to refer to both the separation layer estimated with the mechanisms described below and the points below this layer.

1) Frequency domain filtering: The aim of performing filtering in the frequency domain is to identify those lower frequency components that form the ground surface. An example of frequency response given by an elevation map is shown in Fig. 4(a). In this section we assume that the cut-off frequency (i.e. the frequency beyond which the frequency components are considered to be objects) is given. The next section explains how this cut-off frequency is automatically determined.

Applying a hard threshold in the frequency domain creates the well known ringing effect in the reconstructed spatial signal [14]. To avoid this ringing effect a low-pass filter with smooth cut-off is used to remove the higher frequency (i.e. non-ground) components. We employ a linear phase frequency domain filter with Butterworth-like magnitude response. Another advantage of this filter is that the magnitude in the passband region is almost constant [14], hence the ratios of the lower frequencies are maintained. The equation of this Butterworth-like filter is:

$$|T(d)| = \frac{1}{\sqrt{1-\epsilon^2 \cdot \left(\frac{d}{d_c}\right)^{2 \cdot n}}}, \quad n \geq 1 \tag{1}$$

where T is the filter response, $\epsilon = \sqrt{10^{A/10} - 1}$ is the band edge value with A being the passband variation. d is the distance from DC, d_c is the cut-off point and n is the filter order. The filter order defines how sharp the damping is; $n = 2$ is used in our implementation.

The 2D frequency response of an image has the same dimensions as the image itself. The frequencies evaluated by the FFT along one dimension of the image are $[0, 1, 2, \ldots, \frac{N}{2}] \cdot \frac{1}{N}$ where N is the number of pixels along that dimension. In the case of non-square images (such as the elevation grids processed here) the range of frequencies evaluated along each dimension is different (as it depends on N). Hence, the cut-off frequency is here encoded as a ratio (a unitless number between $[0, 1]$) of the range of frequencies. This implies that consistent filtering of objects, independent of their orientation and location in the image, can be obtained in non-square images by modulating the cut-off frequency by an elliptical gain.The extent of the major and minor axis of the ellipse is given by: $a = f_c \cdot \frac{w}{2}$ and $b = f_c \cdot \frac{h}{2}$, where $f_c \subset [0, 1]$ is the cut-off frequency ratio and w/h the image width/height. The 2D filter with elliptical weighting is then obtained by replacing the ratio $\frac{d}{d_c}$ in (1) by:

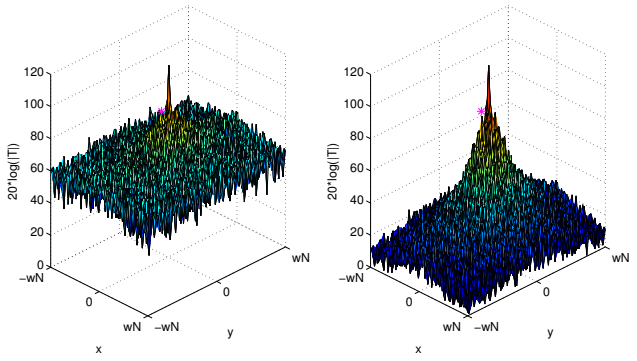
$$\left(\frac{d}{d_c}\right)^2 = \left(\frac{x - x_o}{a}\right)^2 + \left(\frac{y - y_o}{b}\right)^2 \tag{2}$$

where (x_o, y_o) is the ellipse origin, which is the centre of the shifted frequency response image (a shifted image is rearranged so that the DC is in the centre of the image). An example of filtered frequency response image is shown in Fig. 4(b).

2) Cut-off frequency estimation: dict The estimation of the cut-off frequency is based on a peak detection process. Peaks in the frequency domain result from periodic structures in the original spatial-domain image and their radius and direction correspond to the spacing and the orientation of a periodic element [23]. A peak, Π, is found as a local maximum:

$$\Pi(j, i) = \begin{cases} 1 & \text{if } (I_\omega(j, i) - I_\omega(l, k)) > c \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

where $I_\omega(j, i)$ indicates the (j, i)-th frequency image pixel, $l = \{j-n, j-n+1, ..., j+n\}$ and $k = \{i-n, i-n+1, ..., i+n\}$ are the indices of the local region. In an eight-connected region $n = 3$. The constant c is a threshold defining how much larger



(a) Original (b) Filtered

Fig. 4. Bode magnitude plots of Knidos sub-map 16. (a) The original frequency response image and (b) the filtered frequency response image. The frequencies above the cut-off (here set to 7.99% and shown with a magenta star) are suppressed by $> 40\,\mathrm{dB}$.

the peak has to be compared to the values in the neighboring cells and is set to 0 in our implementation.

If the structure of the spatial domain is known [23] this information can be used to select the correct peak and set the cut-off frequency directly. Otherwise, the most salient peak has to be searched for. The location of the closest peak to origin is used as an initial estimate of the cut-off frequency. The frequency components contained up to this peak may not be enough to reconstruct the underlying ground surface. This is for example the case when a sub-map contains slow undulating patterns or a constant curvature. Fig. 1(a) illustrates the case of an (approximately) constant curvature where the sub-map is located on the base of a mound. To assess whether more frequency components need to be included in the estimate of the ground surface, we develop a mechanism that estimates the maximum size of an object m_c for a given cut-off frequency and uses m_c to refine the cut-off frequency. Beyond m_c, an object has frequency components below the cut-off frequency and is identified as part of the ground.

The value m_c is obtained by modelling an object in the spatial domain as a step function. The frequency response of a step function is a sinc function [14]. Correspondingly, the response to a 2D step (a cube) is $J_1(z)/z$, where $J_1(z)$ is the first order Bessel function [9]. Knowing the size of the cube τ, the location of the first zero crossing of the frequency response is given by $2/\tau$. The wider the object, the closer the location of the first zero crossing to the DC; which is consistent with the analysis above identifying larger objects with lower frequency components. The first zero crossing of the first order Bessel function is followed by progressively decreasing undulations. However, the range of frequencies up to the first zero crossing contain more than 90% of the signal energy. Therefore, we assume that the overall shape of the object is maintained when applying a low-pass filter with a cut-off frequency at the first zero crossing. Given this modelling, a cut off frequency $f_c > \frac{2}{\tau}$ implies that an object of size τ is included in the ground. On the other hand, if $f_c \leq \frac{2}{\tau}$, an object of size τ is partly filtered out during ground extraction.This implies that the maximum size of objects m_c beyond which

objects are included in the ground surface can be approximated as $\frac{2}{f_c}$. Strictly speaking, objects smaller than $\frac{2}{f_c}$ are only partly filtered out, the effect being stronger for smaller objects. The latter equation transposed from the image space to the metric world gives:

$$m_c = 2/f_c \cdot \text{res}_g, \tag{4}$$

where res_g is the resolution of the elevation grid.

This formula allows the ground extraction algorithm to automatically adjust the detected cut-off frequency given a maximum object size. The ground extraction algorithm is formulated in such a way that the user provides a relative value $m_u \subset [0, 1]$ which represents the maximum object size with respect to the extent of the sub-map. The sub-map extent e is obtained as the minimum of the extents in x and y.

Once the initial peak is identified, the corresponding maximum object size is obtained using (4). If $m_c > m_u \cdot e$, f_c is set to the next closest peak from DC. This is repeated until $m_c \leq m_u \cdot e$. The outcome of this process is a cut-off frequency f_c which is selected in such a way that the maximum size of objects considered as non-ground is $m_u \cdot e$. In our implementation $m_u = 0.5$. As an example, in the case of sub-map 6 from the KnidosH data set (shown in Fig. 3), the ground extraction algorithm sets f_c to the second closest peak. Due to the overall constant upward curvature of this sub-map, additional frequency components are required to model the underlying ground. This means that the initial maximum object size given by the first peak was beyond half the sub-map extent (i.e. beyond $m_u e$).

3) Ground surface reconstruction: Once f_c is defined, and filtering is applied to the frequency domain image, the estimate of the ground surface is built by applying the inverse FFT. Examples of reconstructed ground surfaces are shown in Fig. 3.

C. Ground/object separation

Given an estimated ground surface, the objects above this surface are obtained by comparing the height of the points in the original point cloud, to the height of the surface. If a point is on or below the ground surface, it is labelled as ground; if it is above it is labelled as non-ground. Once the non-ground points are identified, they are clustered using a standard proximity voxel based clustering: non-empty voxels in contact of each other are gathered in the same cluster. The resulting set of clusters form segments of non-ground points. These can be used as landmarks for navigation as developed in Sec. V-D.

V. Experiments

This section presents two sets of experimental results. First, an evaluation of the proposed ground extraction method is performed using ground-truth hand labelled data (Sec. V-C). Second, the proposed method is evaluated in terms of the alignment it leads to when the segments are fed to the S-ICP alignment algorithm (Sec. III). These results are presented in Sec. V-D. In both sets of experiments, the proposed approach is compared to three other techniques; these are described in Sec. V-B. All experiments are repeated on two data sets which are introduced in Sec. V-A.

A. Data sets

The 3D scans used here are produced by the structured light laser profile imaging device described in [17]. The basic concept consists of a green laser sheet projected on the sea floor and a calibrated camera imaging the laser line. The shipwreck sites were surveyed at approximately an altitude of three meters with the vehicle traveling 10-15 cm/sec. Once extracted from the collected images and projected, the resulting point density is approximately 7 points per square centimeter. The individual points have a range resolution better than one centimeter. The two following data sets are processed: (1) KnidosH, which contains 101 sub-maps; (2) Pipewreck, which contains 42 sub-maps. In each set, sub-maps contain around 600,000 points.

B. Benchmark segmentation techniques

This sections presents three ground segmentation methods which will be used as benchmarks in the evaluation of the proposed approach.

1) Naive ground extraction: In this first approach, the mean height of the point cloud is averaged, and the ground is simply defined as being the set of points below the mean height. This approach will be referred to as the "Naive" method.

2) MLESAC-based plane extraction: In this second approach, a plane is fitted to the point cloud and the ground is defined as the points below this plane. The Point Cloud Library [2] implementation of a MLESAC (Maximum Likelihood Estimation SAmple Consensus) based plane fitter is used in the proposed experiments. This method will be referred to as the "Planar" method.

3) Grid-based ground extraction: In this third approach, the point cloud is voxelised, and the ground points are identified as the ones contained in the bottom voxel of each column of the voxel grid. The grid resolution relates, to some extent, to the value of the cut-off frequency automatically estimated by the proposed DFT-based method. A higher resolution (i.e. a smaller grid separation) will result in finer details of the terrain being included in the estimated ground surface, which, to some extend, corresponds to larger cut-off frequency in the context of the DFT-based method. However, as developed in Sec. IV, the Fourier formalism allows the algorithm to reason about the range of scales contained in the terrain relief, and, given a maximum admissible object size, it allows to automate the selection of the relief scales relevant to the definition of the ground. Such reasoning is not readily applicable in a grid based elevation segmentation and its resolution would have to be manually adjusted as a function of the data set. The grid resolution is set to 2cm in our experiments. This approach is referred to as the "Grid" method.

C. Comparisons to hand labelled segmentation

1) Segmentation quality metric: The segmentation methods are compared to hand labelled segmentation to determine which points are correctly or incorrectly labelled as object

and ground. The evaluation is performed by quantifying the True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) rates in the binary classification task of identifying the classes ground and object; where Positive refers to the object class and Negative to the ground class.

2) Results: From each of the KnidosH and Pipewreck data sets two sub-maps were chosen and manually labelled. For each point the labels ground / non-ground were assigned. These sub-maps (Fig. 1) were chosen since their content is representative of the two data sets.

The results are presented in Table I. For each segmentation method the true positive and true negative rates as well as the F1 score are shown. The best results in terms of the F1 score are indicated in bold.

From these results it is clear that the proposed method is outperforming all the other approaches. The Planar method generally produced the least proportion of FNs. However, this came at the expensive of having the largest FPs. On the contrary, our DFT method produced the lowest number of FPs while having very low FNs too resulting in significantly higher F1 score. These results shown that this method provides the most consistent performance in identifying the underlying ground surface compared to the tree benchmark techniques.

D. Comparisons of alignment results

This sections presents an evaluation of segmentation methods in terms of the quality of the alignment they lead to when used in conjunction with the S-ICP alignment method (Sec. III). The proposed segmentation approach plus the three benchmark techniques detailed in Sec. V-B are run on each sub-map of the two data sets. This allows to evaluate the performance of each segmentation technique in an alignment pipeline.

An initial odometry-based navigation solution is provided using a Doppler Velocity Log, Fiber-Optic Gyrocompass, and depth sensor. Sub-maps are broken before the accumulated odometry error has become significant relative to the bathymetry sensor resolution [16]. An Ultra-Short Baseline acoustic positioning system is used to reference the survey to a geodetic datum, but is otherwise too inaccurate to localize the sub-maps or correct the navigation within a sub-map. The aim of the experiments presented here is to show that pairwise alignment through accurate terrain feature segmentation can locally improve sub-map registration. The use of these local pairwise registrations in a global scan alignment process (as in [3]) or in the online navigation algorithm of [16] is left for future work.

To define pairs of sub-maps for alignment, each sub-map set is traversed as follows. The first sub-map s_1 is paired up with the closest neighbour sub-map (closest in terms of the distance between their centroid); this sub-map then becomes s_2. s_2 is paired up with the closest sub-maps amongst the sub-maps not already in a pair. Each sub-map in a data set is assigned to a pair by repeating this process.

1) Alignment quality metric: To evaluate the quality of the alignment the metric used here attempts to capture the crispness of aligned scan pairs. This metric consists of voxelising the aligned point cloud and returning the number of occupied voxels. It will be referred to as Nx. The lower the value produced by the metric, the crisper the point cloud and in turn the more accurate the alignment. The rationale for using this metric as opposed to the ICP residual is developped in [7]. The voxel resolution used in all the experiments presented here is 2cm.

2) Results: The result of the alignments are summarised in Table II. Naive, Planar and Grid refer to the techniques described in Secs. V-B1-V-B3. FFT refers to the proposed segmentation algorithm. A number of quantities are used to evaluate the quality of the alignments associated to each segmentation method. The first is the mean of the variation of the Nx metric (ΔNx) before and after the alignment is applied. This is reported as a percentage in column 1. The associated standard deviation is given in column 2. A negative value of ΔNx corresponds to an improvement of the crispness of the point cloud. A positive value, on the contrary, corresponds to the point cloud loosing in sharpness during the alignment. This may, for instance, be due to segments being mismatched across sub-maps (i.e. a data association failure). S-ICP only computes an alignment if at least N_{seg} are matched across sub-maps. N_{seg} is specified by the user and is set to 3 in our experiments. If less than N_{seg} are associated during a run of S-ICP, the alignment is not computed and ΔNx is set to 0. The number of sub-maps where the alignment has degraded ($\Delta Nx > 0$), has not been calculated ($\Delta Nx = 0$), and has improved ($\Delta Nx < 0$) is reported in columns 3-5, respectively. The mean and the standard deviation provided in columns 1 and 2 are computed using only the sub-map pairs for which the alignment is computed (i.e. $\Delta Nx > 0$ and $\Delta Nx < 0$). In the case of rather flat terrains, only small segments will be extracted which may not be large enough to be accepted for processing. The minimum segment extent is fixed to 2cm in our implementation. The absence of large enough segments in any of the sub-maps in a pair results in the pair not being processed. These cases are counted in the last column of Table II. The results for each of the two data sets are separated in each cell of the table by "/": KnidosH/Pipewreck. The best results in each column and for each data set are indicated in bold.

The first column of Table II shows that for both data sets the proposed method (last row) outperforms the benchmark methods by providing on average the best improvement in crispness when used in conjunction with S-ICP. The associated standard deviation (column 2) is of the same order of magnitude as the mean, suggesting a non negligible range of variations in ΔNx. This is due to the quality of the initial navigation solution which varies across the data set and which is on average better in the KnidosH data set. In some cases, the provided navigation is accurate (in particular along the linear segments of the ROV's trajectory) and little improvement is obtained in terms of point cloud sharpness. In other cases, in particular for areas re-visited from different transects (near loop closures), the alignment provided by S-ICP leads to a

TABLE I

COMPARISON OF SEGMENTATION METHODS WITH GROUND TRUTH (TPR/TNR/F1).

	KnidosH data set						Pipewrech data set					
	Sub-map 6			Sub-map 16			Sub-map 13			Sub-map 41		
Naive	0.57	0.79	0.70	0.60	0.60	0.58	0.77	0.83	0.82	0.74	0.87	0.82
Planar	0.48	0.99	0.64	0.21	0.84	0.31	0.68	0.92	0.78	0.76	0.89	0.83
Grid	0.82	0.83	0.87	0.65	0.80	0.68	0.82	0.77	0.84	0.69	0.86	0.78
FFT	0.90	0.85	**0.92**	0.94	0.76	**0.82**	0.98	0.95	**0.97**	0.96	0.88	**0.95**

TABLE II

DELTA CRISPNESS BEFORE/AFTER ALIGNMENT (KNIDOSH / PIPEWRECK DATA SETS)

	$\mu\,(\Delta Nx)$ [%]	$\sigma\,(\Delta Nx)$ [%]	$\#\Delta Nx > 0$	$\#\Delta Nx = 0$	$\#\Delta Nx < 0$	$\#unprocessed$
Naive	1.14 / Nan	2.37 / Nan	21 / **0**	71 / 41	8 / 0	**0 / 1**
Planar	0.43 / -2.04	2.73 / 5.87	21 / 2	59 / 35	15 / 2	6 / 3
Grid	-0.25 / -3.47	**1.20 / 2.13**	22 / **0**	**49 / 34**	26 / 5	4 / 3
FFT	**-1.28 / -4.25**	1.23 / 2.90	**6 / 0**	50 / **19**	**41 / 20**	4 / 3

more significant improvement in the point cloud crispness. Examples of alignment results are shown in Fig. 5. The other indicators confirm the relative performance of the four methods. For both data sets, the proposed approach leads to the smallest number of incorrect alignments (column 3), to the smallest number (modulo 1 sub-map pair) of non-aligned sub-map pairs (column 4), and to the largest number of improving alignments (column 5).

The percentages in the first column of Table II are small (in terms of their absolute values) due to the following reason. ΔNx (expressed in %) is obtained as: $\Delta Nx = \frac{Nx_{after} - Nx_{before}}{Nx_{after}} \cdot 100$, where the indices "before" and "after" refer to before and after applying the alignment. As can be seen in Fig. 5, an alignment results in overlapping segments being moved relative to one another compared with the original, uncorrected alignment provided by the navigation solution. The resulting change in the number of occupied cells corresponds to a small number of points compared to the number of points in the two sub-maps. Fig. 5 gives an intuition on how ΔNx values map to actual alignments.

The Naive segmentation method is not expected to provide high performance but is used here to gage the merits of a low complexity approach. The corresponding results show that accurate alignment requires more terrain modelling than a simple threshold based separation. The Pipewreck data set, which, as can be seen in Figs. 5(a)-5(b), contains large sections of flat terrain. The mean of the terrain height tends to be pulled in the flat sections, implying that large sections of flat terrain are identified as non-ground. These large segments cannot be matched across sub-maps since their extent and shape is less representative of the terrain itself than of the extents of the sub-map they were extracted from. A similar behaviour can be observed in the KnidosH data set.

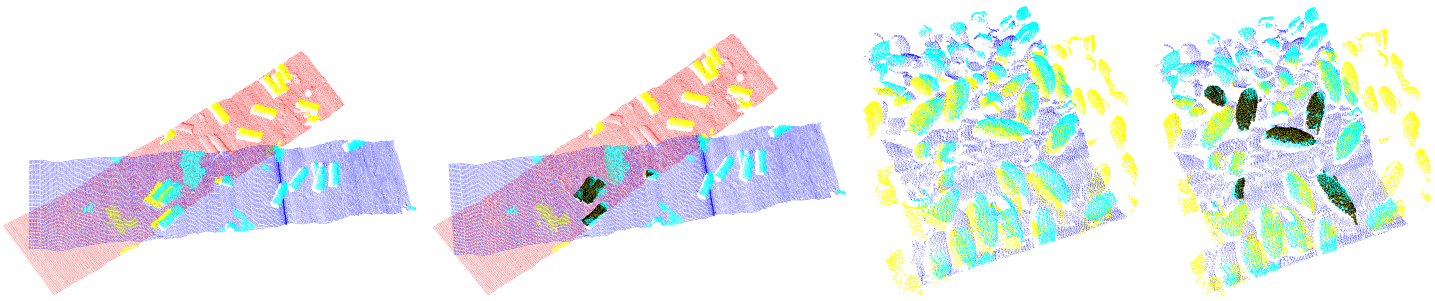
The Planar segmentation method is well suited to the Pipewreck data in which most of the sub-maps contain non-negligible visible portions of rather flat ground (Figs. 5(a)-5(b)). The extraction of a planar surface via a RANSAC method is in this case appropriate and leads to an accurate ground segmentation (see results in Sec. V-C). The application of the same method to the KnidosH data set which contains more diverse relief (Figs. 5(c)-5(d)) shows that simple plane extraction is not general enough to provide accurate segmentation / alignment in a variety of terrains. This method provides on average an improvement of the point cloud crispness in the Pipewreck data set ($\Delta Nx = -2.04$), but tends to lead to point clouds with a degraded sharpness in the KnidosH data set ($\Delta Nx = 0.43$).

The next step up in terms of complexity of the terrain model corresponds to the Grid method. A more accurate modelling of the terrain relief obtained with the Grid method allows improved performance with respect to the two previous techniques. For both data sets, the Grid method leads to an improvement of the point cloud crispness: $\Delta Nx = -0.25$ for the KnidosH data set and $\Delta Nx = -3.47$ for the Pipewreck data set. As explained in Sec. V-B3 however, automating the definition of a grid resolution which allows the larger scales of the relief defining the underlying ground to be captured is not trivial without resorting to a spectral type of analysis.

The proposed FFT based segmentation method is able to perform better than the methods making assumption of a (partially) flat terrain (Naive and Planar) and methods requiring a resolution to be fixed a priori (Grid). It can reconstruct non-planar terrains (as illustrated in Fig. 3(c)) and reason globally on the point cloud (modulo some formatting into an elevation grid) as opposed to reasoning only locally in a column of heights. The set of results presented in Table II provide experimental evidence in favour of the applicability of the proposed ground segmentation method to a range of different terrains and its ability to generate features useful for navigation.

VI. CONCLUSION

This paper has introduced a method for segmenting 3D scans of underwater unstructured terrains. The method extracts a ground surface by selecting the lower frequency components in the frequency domain that define the slower varying undulations of the terrain. The points above the estimated ground surface are clustered via standard proximity clustering to form object segments. The experimental results show that the approach is applicable to a range of different terrains and is able to generate segments which are useful landmarks for navigation. Future work will integrate segmentation based registration to a global scan alignment process, and to online navigation algorithms. The DFT-based ground extraction mechanism will also be extended to perform non-regular DFTs or polar FFTs to process polar scans that generate irregular grids, which cannot be directly processed with standard FFTs. Finally, we will



(a) before alignment (b) after alignment (c) before alignment (d) after alignment

Fig. 5. Examples of alignment results using the DFT ground segmentation method in conjunction with S-ICP. (a) and (c), pairs of sub-maps before alignment; (b) and (d), the same sub-maps after alignment. In the two left plots, a pair of sub-maps from the Pipewreck data set. The reference sub-map (i.e. the sub-map not being aligned) is shown in blue with the segments generated by the DFT-based method shown in turquoise. The sub-map being aligned is shown in red with the segments generated by the DFT-based method shown in yellow. The black segments in the aligned pair indicate the segments matched by S-ICP across the two point clouds and used to compute the alignment. As can be seen in (b), corresponding segments are brought right on top of each other after alignment. This alignment corresponds to $\Delta Nx = -5.4\%$. Note that if the alignment were to be performed using standard (dense) ICP, the two sub-maps would be brought completely on top of each other due to the ground point dominating the cost function in the ICP optimisation. (c)-(d) show a pair of sub-maps that belongs to the KnidosH data set. The same colour coding as (a) and (b) is used except that the ground points in the aligned map (in red in (a) and (b)) are not shown for the sake of clarity. In (d) it can be seen that the corresponding segments are brought right on top of each other by the alignment. This alignment corresponds to $\Delta Nx = -12.6\%$.

compare the proposed method to image segmentation based approaches such as the GraphCut algorithm.

VII. Acknowledgements

This research was supported by the Australian Research Council (ARC) through the Discovery programme, the Australian Government through the SIEF programme, and by the Australian Centre for Field Robotics at the University of Sydney. The authors would like to thank Alastair Quadros, Peter Morton and Vsevolod Vlaskine for valuable help with software, as well as James P. Underwood, Mitch Bryson and Donald Danserau for useful discussions.

References

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

[2] R. Bogdan 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.

[3] D. Borrmann, J. Elseberg, K. Lingemann, A. Nüchter, and J. Hertzberg. Globally consistent 3d mapping with scan matching. *Robotics and Autonomous Systems*, 56(2):130–142, 2008.

[4] Y. Boykov and G. Funka-Lea. Graph cuts and efficient and image segmentation. *International Journal of Computer Vision*, 70(2):109–131, 2006.

[5] C. J. Brown, S. J. Smith, P. Lawton, and J. T. Anderson. Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. *Estuarine, Coastal and Shelf Science*, 2011.

[6] B. Douillard, J. Underwood, N. Kuntz, V. Vlaskine, A. Quadros, P. Morton, and A. Frenkel. On the segmentation of 3D LIDAR point clouds. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2010.

[7] B. Douillard, A. Quadros, P. Morton, J. P. Underwood, M. De Deuge, S. Hugosson, M. Hallström, and T. Bailey. Scan segments matching for pairwise 3D alignment. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2012.

[8] A. Golovinskiy and T. Funkhouser. Min-cut based segmentation of point clouds. In *IEEE Internationa Conference on Computer Vision (ICCV) Workshops*, pages 39–46, 2009.

[9] B. Horn. *Robot vision*. The MIT Press, 1986.

[10] 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.

[11] S. Kunis. *Nonequispaced FFT - Generalisation and Inversion*. PhD thesis, Universität Lübeck, Germany, 2006.

[12] D. Huttenlocher P. Felzenszwalb. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2): 167–181, 2004.

[13] Y Petillot, I.T. Ruiz, and D.M. Lane. Underwater vehicle obstacle avoidance and path planning using a multi-beam forward looking sonar. *IEEE Jornal of Oceanic Engineering*, 26(2): 240–251, Jan 2001.

[14] J. G. Proakis and D. G. Manolakis. *Digital signal processing - Principles, Algorithms, and Applications*. Prentice Hall International, 3 edition, 1996.

[15] S. Reed, I.T. Ruiz, C. Capus, and Y. Petillot. The fusion of large scale classified side-scan sonar image mosaics. *IEEE Transactions on Image Processing*, 15(7):2049–2060, Jan 2006.

[16] C. Roman and H. Singh. A self-consistent bathymetric mapping algorithm. *Journal of Field Robotics*, 24(1-2):23–50, 2007.

[17] C. Roman, G. Inglis, and J. Rutter. Application of structured light imaging for high resolution mapping of underwater archaeological sites. In *IEEE OCEANS*, pages 1–9, Sydney, 2010.

[18] 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.

[19] J. Schoenberg, A. Nathan, and M. Campbell. Segmentation of dense range information in complex urban scenes. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2010.

[20] B. Steder, R.B. Rusu, K. Konolige, and W. Burgard. NARF: 3D range image features for object recognition. In *IEEE/RSJ International Conf. on Intelligent Robots and Systems (IROS) Workshops*, 2010.

[21] J. Strom, A. Richardson, and E. Olson. Graph-based segmentation of colored 3d laser point clouds. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010.

[22] R. Triebel, J. Shin, and R. Siegwart. Segmentation and unsupervised part-based discovery of repetitive objects. In *Robotics: Science and Systems*, Zaragoza, Spain, June 2010.

[23] B. Xu. Identifying fabric structures with fast Fourier transform techniques. *Textile Research Journal*, 66(8):496–506, 1996.

[24] X. Zhu, H. Zhao, Y. Liu, Y. Zhao, and H. Zha. Segmentation and classification of range image from an intelligent vehicle in urban environment. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1457 –1462, October 2010.