COOPERATION

Grant Agreement No. 611373

Cognitive Autonomous Diving Buddy

FP7-ICT-2013-10

D2.2 Motion compensation algorithms

Due date of deliverable: 30.6.2015 (M18)
Actual submission date: 10.7.2015

Start date of project: 01 January 2014 Duration: 36 months

Organization name of lead contractor for this deliverable: JACOBS

v1.0

Dissemination level

PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

1 Contents

2 Executive Summary

This report describes motion compensation methods aimed at stabilizing data from a moving sensor observing a partially dynamic scene. Specifically, this is done in the context of diver gesture recognition, where the background is assumed to be static and the gesturing diver constitutes the foreground. A successful motion compensation method allows to track the motion of the sensor relative to the static parts of the scene, thus facilitating a) recognition and b) segmentation of the dynamic content (i.e. the diver).

3 Problem Description

When a moving sensor observes a dynamic environment, there are two sources of change in the data: a) The motion of the sensor, and b) the motion of the dynamic parts of the environment (e.g. animals, plants, divers, etc). Motion compensation methods address the first source: Sensor motion. Here, it is explicitly used to facilitate the processing of the dynamic environment features.

Motion compensation is a necessary ingredient of any application using mobile sensors and requiring sensor data integration or fusion over time. Examples of such applications are Synthetic Aperture Radar (SAR) and Sonar (SAS), as well as hydrographic surveys using sidescan and multibeam echo sounders. The motion of the sensor over time is usually estimated using expensive inertial sensors and gyroscopes or with high-accuracy differential GPS.

Motion compensation becomes much more difficult for reasonably priced autonomous underwater vehicles (AUVs). Highly accurate inertial sensors are too expensive, and GPS is not available. Instead of relying on open loop estimates (e.g. dead reckoning), the methods described in this report use the sensor data itself to estimate the sensor motion.

4 Related Work

Motion compensation methods are specifically used here to facilitate detection and recognition of human movement in the scene. There are two strands of research related to this goal: Diver detection and sensor data registration.

The predominant approach for diver detection is the use of acoustic sensors, either in form of passive or active devices. Passive sensing methods use sound analysis and hydrophones [11] [44] [30] [43] [45] [15] [23], especially to detect the signatures of open circuit breathing systems. Active sensing systems use (imaging) sonars [2] [25] [38] [5] including sonar devices especially developed for this purpose [32] [12] [1]. Even electric-field sensors [28][29] have been used for diver detection in shorter ranges of several meters. All this work has in common that it is targeted to the special case of harbor security [34] and vessel protection [33], i.e., cases where the divers are expected to be non-cooperative and where it is desirable to detect them from a longer distance.

Sensor data registration is often used for video stabilization or (ego-)motion compensation [42, 35, 46, 22, 14]. It is well known in the land robot community that video-stabilization significantly eases vision processes on moving vehicles; see for example [20, 31]. Especially, it can be used as basis for object detection by motion extraction through differential images [24], which are also used here. A shape-analysis based on aspects ratios of the bounding boxes of the segments can then in addition be used to further support the detection [3] [48].

This report focuses on the latter aspect of using visual data to compensate sensor motion and facilitate motion segmentation. One strain of existing work deals with the most simple form of interaction, namely a robot following a human diver. Gregory Dudek etal. have for example used spectral processing of visual data to track human motion patterns - concretely the oscillating leg motions of a diver - to let an AUV follow a human [40, 41]. They also suggested in [39] the usage of visual features, concretely color cues and local temporal gait signatures in the frequency domain, to follow a diver. The tracking of divers by sonar is reported to be very difficult, mainly due to disturbances by the bubbles of breathing systems [38]; nevertheless, acoustic means like a pinger on the diver and a (ultra-)short baseline tracker [21, 47] on the robot are in principle a possible but very costly option.

Previous work by Jacobs [7, 8] also addresses the issue of motion segmentation from video sequences, and details ideas how to use detected motion to interpret gestures. However, this work has solely addressed monocular image sequences.

In the context of CADDY it is important to address the following novel problems relative to the cited state-of-the-art:

1. How can sensor motion be compensated given 3D range measurements?
2. How robust can motion compensation methods be to dynamic content in the measurements?
3. How robust can motion compensation methods be to noise in these measurements?

The work presented here extends previous work in [7] to 3D and explores the two sources of inconsistencies between two observations: Dynamic content and range measurement noise.

5 Motion Compensation via Spectral Registration with Multilayer Resampling (SRMR)

Spectral Registration with Multilayer Resampling (SRMR) is a 6 Degrees Of Freedom (DOF) registration method for noisy 3D data with partial overlap. The algorithm is based on decoupling 3D rotation from 3D translation by a corresponding resampling process of the spectral magnitude of a 3D Fast Fourier Transform (FFT) calculation on discretized 3D range data. The registration of all 6DOF is then subsequently carried out with spectral registrations using Phase Only Matched Filtering (POMF).

Through the use of all data points, instead of just certain correspondences as, e.g. in ICP [4], SRMR is robust to occlusions as well as some dynamic content.

5.1 Overview and basic notations

The motivation of this work is to find a stable process which is capable of recovering 6DOF transformations - including especially the 3D angle information - under severe interferences and occlusions between two input scans. An approach is proposed, which has two main elements for a fast and robust registration of Euler angles from spherical information. First of all, there is the permanent use of phase matching through Phase Only Matched Filtering (POMF). Second, based on a FT on a discrete Cartesian grid, not only one spherical layer but a complete stack of layers is processed in one step. This involves a resampling scheme which allows a registration on a 3D Cartesian grid that leads to better results compared to an accumulation of several layers and a subsequent corresponding correlation in 2D. The reasons for this increased robustness are motivated in Section 5.6. The subsequent phase matching based registrations determine yaw, roll/pitch, and translation by yielding peaks indicating the according solutions. Due to the key point of using a stack of layers, we dub our method Spectral Registration with Multilayer Resampling (SRMR). The entire process can be outlined as follows in three main steps with several sub-steps:

1. yaw determination:
    a. resample hemispheres (projection in spherical coordinates) on different radii from the magnitude of the 3D spectrums
    b. determine the yaw angle by a rotational registration (polar resampled) from the resampled structures (3D POMF)
2. roll-pitch determination:
    a. re-rotate the 3D spectrum according to the determined yaw angle in order to align the spectrums for yaw

   b. resample hemispheres (rectangular projection) on different radii from the magnitude of the 3D spectrums
   c. determine roll and pitch angle by translational registration from the resampled structures (3D POMF)
3. translational registration:
   a. re-rotate scan data according to all determined angles in order to align the scans for the remaining 3D translation
   b. determine the 3D translation between the rotationally aligned scans by a 3D POMF registration

So, as in many other approaches the rotation is determined first. In comparison to the spherical Fourier registration this is done by a particular resampling of points $\eta(\vartheta,\phi) \in S^2$ on the unit sphere directly from the 3D spectrum. Here, the structures on the unit sphere S^2 are taken from different radii corresponding to different 3D frequencies and they are assembled to a 3D stack. Our SRMR method exploits for the determination of the rotational orientation the fact that the information about the 3D rotation is available within the magnitude of the 3D spectrum and that it is therefore decoupled from the overall translation.
Lowercase letters indicate the time domain and uppercase letters the frequency domain. The relation between the translation and the rotation of the input voxel data and its corresponding effects within the 3D spectrum are as follows:

$$F(\mathbf{k}) = \frac{1}{N^3} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \sum_{z=0}^{N-1} f(\mathbf{x}) e^{-i2\pi(\mathbf{k}^T\mathbf{x})} \tag{1}$$

$$r(\mathbf{x}) = s(R(\alpha,\beta,\gamma)\mathbf{x} - \mathbf{t_s}) \tag{2}$$

$$R(\mathbf{k}) = S(R(\alpha,\beta,\gamma)\mathbf{k}) e^{i2\pi R(\alpha,\beta,\gamma)\mathbf{k}\mathbf{t_s}} \tag{3}$$

$$|R(\mathbf{k})| = |S(R(\alpha,\beta,\gamma)\mathbf{k})| \tag{4}$$

The 3D DFT on a Cartesian grid is given by (1). It can be shown by e.g. a coordinate transform that a rotation of a 3D structure by R(α,β,γ) orientates the magnitude of the corresponding spectrum in the same way, while a translational shift does not affect the spectrum magnitude. Having a relation between two 3D signals as in (2) with x = [x y z]T and any translational shift ts = [xs ys zs]T then the spectral relation is given by (3) with k = [u v w]T. In terms of the magnitude this relation simplifies to (4). This relation allows the decoupling of translation and rotation for the registration process.

5.2 Phase Only Matched Filtering (POMF)

The underlying signal registration in our method is computed using a Phase Only Matched Filtering (POMF) [19] within all dimensions. This correlation approach employs the fact that two shifted signals with the same spectrum magnitude carry the shift information within their phase as already indicated by (2) and (3):

$$Q(\mathbf{k}) = \frac{S(\mathbf{k})^*}{|S(\mathbf{k})|} \bullet \frac{R(\mathbf{k})}{|R(\mathbf{k})|} \tag{5}$$

$$q(\mathbf{x}) = \mathscr{F}^{-1}\{Q(\mathbf{k})\} \tag{6}$$

$$x_s, y_s, z_s = \underset{x,y,z}{\operatorname{argmax}}\, q(\mathbf{x}) \tag{7}$$

It can be applied to N-dimensional signals and it is used for the registration of the unwrapped data for the determination of the orientation as well as for the determination of the translation. (5) gives the details of the underlying transformation; as indicated by the complex conjugate the implementation of the phase matching is simply the inverse Fourier transform of the spectral angle difference. The result is a function, which contains a Dirac peak of $r(\mathbf{x}) = s(\mathbf{x}-\mathbf{t}_s)$ in the ideal case (6).

The resulting shifted Dirac pulse deteriorates with changing signal content of both signals. An alternative would be the standard cross-correlation. But cross-correlation yields several broad peaks including the main peak. This is the reason why phase matching is usually preferred when working with signals which are not predefined, e.g., in sonar or radar applications. Meanwhile a variety of methods exists to improve phase matching with respect to subpixel accuracy, e.g. [16, 17, 18]; the approach of [16] for subpixel interpolation is used here. Such interpolation is useful for improving the accuracy above the voxel discretization. The matching process requires a corresponding preprocessing of the data in order to suppress high amplitudes at the boundaries and to avoid spectral artifacts. The principle of window functions is well known and can easily be extended to multiple dimensions by convolution of single functions. A detailed description can for example be found in [36].

5.3 3D Rotation on the sphere

Any point on S^2 can be rotated according to Euler by an element of the rotation group *SO*(3). The 3D rotation within a Cartesian coordinate system is then defined as a result of a multiplication of the three matrices $R(\alpha,\beta,\gamma) \in SO(3)$ corresponding to each axis:

$$R(\alpha, \beta, \gamma) = R_x(\alpha) R_y(\beta) R_z(\gamma) \quad (8)$$

In accordance to (4) it is necessary to recover the 3D rotation from corresponding shifted spherical structures. The key point in finding a translation a by a FT is that a translation affects the Fourier coefficients in an analytical way by multiplication with $e^{i\omega a}$. To find a similar concept for rotation on a sphere, a basis must be found where a rotation also affects the corresponding coefficients in an analytical way. An exact solution for this problem is given in [27, 26] using Spherical Harmonics (SH). Note that the rotational information within the spectrum is not only available in one layer but on all radii. But the correlation approach based on SH allows only a processing on one layer. This limits the robustness when real world scans are to be registered which have only partial overlap. This is the main reasons for using another approach here. Furthermore, the computation time of the SH correlation is higher compared to phase matching, i.e., we gain faster processing as a fringe benefit.
The crucial point is how to resample the spectrum into a 3D Cartesian grid where the rotation can be determined without accumulating radial information. This is comparable to the well known decoupling of rotation and translation with a polar resampling of a 2D spectrum [13] or the polar-logarithmic resampling scheme as adopted in the Fourier Mellin transform (FMT) [10] in order to convert rotation and scaling to signal shifts. The goal is to obtain structures where the desired parameters can be found as signal shifts, which can then be determined with phase correlation.
The general idea is to resample layers of a hemisphere at different radii on S^2. Such a resampled layer of a hemisphere is intrinsically not a 2D rectangular matrix, which is the goal for an efficient registration. Therefore a two-stage algorithm is used here to deal with the inevitably structural distortions. The yaw-angle can be determined in the first step of our method over the entire possible range because of its rotational appearance within the unwrapped structure. The 3D spectrum is in the next step re-rotated according to the determined yaw angle. The 3D spectrum contains afterwards only roll and pitch as a tilt within the 3D structure. The problematic part is the mapping of the hemisphere to a square structure for roll and pitch registration, which can not be solved for a full registration from -90 to +90 degrees for these two angles. But it will be shown that the method works very precisely for a wide range of roll/pitch differences between scans, which is sufficient for most applications where 6DOF must be determined, especially in robotic mapping. The related discussion of the deviations is given in Section 5.7. Experiments showed that a reliable registration of noisy data with partial overlap is possible for yaw angles within the full range of -90 to +90 degrees with additional concurrent roll and pitch changes of up to ±35 degrees. A core point is the separation of angular information by squeezing the spherical structures on a plane, i.e., we do no preserve the properties of the sphere-like body which is usually the case when using map projections, e.g., in stereographic projection. The naive idea is to convert structures on a sphere underlying a certain

roll/pitch rotation into a 2D shift in matrix form. For single roll/pitch rotations it is evident that a perpendicular projection of the surface of the hemisphere leads to a shift within the projected structures. A straightforward extension of this approach is the assumption that a subsequent yaw rotation leads to rotation of the projected matrix structure. This is an approximation which holds for a restricted range of roll/pitch. Using this assumption, the 3D rotational registration is reduced to 2D translation plus a single rotation which can be solved by simple image registration [13]. The best results for this approximated registration are achieved using two different resampling schemes. For the yaw processing the hemisphere is resampled along spherical coordinates while the resampling for roll/pitch is done by squeezing the structures by a perpendicular projection of the hemisphere into a matrix. The details for an implementation in subsequent steps is given in the following. Since we define the sequence of 3D rotation as in (8), the registration starts by determining yaw.

5.4 Determination of yaw

The yaw-angle is determined in the first main step of our method by detecting it as rotation within a resampled structure. This square matrix is the descriptor for rotational registration denoted by $id_{rot}(v_x,v_y)$ with a length N_{rot}. The desired structure can be generated by traversing this plane in spherical coordinates, i.e., the magnitude of the spectrum $F(u,v,w)$ with a cubic size N is expressed in a spherical coordinate system with v_x and v_y as the coordinates of the resampled matrix.

$$\begin{aligned} v_x &= 1, \ \ldots \ , \ N_{rot} \\ v_y &= 1, \ \ldots \ , \ N_{rot} \end{aligned}$$

$$\varphi = arctan(\frac{v_x}{v_y}) \ ; \qquad 0 \leq \varphi \leq 2\pi \tag{9}$$

$$\theta = (v_x^2 + v_y^2)^{\frac{1}{2}} \frac{\pi}{N_{rot}} \ ; \qquad 0 \leq \theta \leq \frac{\pi}{2} \tag{10}$$

The access to the spectral magnitude data with spherical coordinates is given by:

$$u = r\sin(\theta)\cos(\varphi) + N/2 \tag{11}$$

$$v = r\sin(\theta)\sin(\varphi) + N/2 \tag{12}$$

$$w = r\cos(\theta) + N/2 \tag{13}$$

This descriptor is then resampled at a certain radius r by:

$$id_{rot}(v_x, v_y) = |F(u,v,w)| \tag{14}$$

Roll and pitch are present in this matrix as an undesirable interference. Roll and pitch are roughly shifting the matrix in x and y direction. Hence, we have two unwrapped matrices which are translated and rotated against each other. This rotation can be decoupled from translation using a polar resampling of the 2D spectrum according to the well known principle of [13].

$$ID_{rot}(u,v) = \mathscr{F}\ \{id_{rot}(v_x, v_y)\} \tag{15}$$

$$\begin{aligned} u_{mk} &= \frac{m\frac{N}{2}}{M}\cos(\frac{\pi k}{K}) + \frac{N_{rot}}{2} + 1 \\ v_{mk} &= \frac{m\frac{N}{2}}{M}\sin(\frac{\pi k}{K}) + \frac{N_{rot}}{2} + 1 \\ m &= 0, ..., M-1; k = 0, ..., K-1 \end{aligned} \tag{16}$$

$$ID_{rotpol}(m,k) = |ID_{rot}(u_{mk}, v_{mk})| \tag{17}$$

The 2D FT is then defined as shown in (15) where the polar resampling is applied according to (16). The spectral magnitude of $|ID_{rot}(u,v)|$ is then defined on a $N_{rot} \times N_{rot}$ grid which is then resampled to a $M \times K$ grid according to (17) where M is the resolution for the radial component and K is the angular resolution; a value of 128 is used for both *M* and *K* throughout all experiments presented later on. The zero frequency as the center of rotation is at $(N_{rot}/2,N_{rot}/2)$. After the polar resampling of the spectral magnitude (17) the expected shift between corresponding grids can then be determined with phase matching. Section 5.6 extends this 2D descriptor to 3D incorporating information along the radial axis.

5.5 Determination of roll and pitch

Roll and pitch are found within one single main step of the overall process. First, the yaw angle determined in the previous main step is used to re-rotate the voxel data of the magnitude of the 3D spectrum, which is then the basis for a new unwrap process. The sampling vector covers the area of the layer in a rectangular manner. The outer parts of the hemisphere are in consequence squeezed into the same number of columns, respectively rows as the center part. A rotated vector traverses the x-axis or y-axis on the hemisphere of the 3D spectrum in the same way as the resulting unwrapped matrix. The resulting matrix then contains roll and pitch roughly as a translation in x and y direction. Figure 1 illustrates this. Due to this resampling scheme it is obvious that much more voxels are resampled in the middle of the resampled hemisphere than at the edges. This is one reason that limits the range of roll and pitch in our method.

$$u = r\ \cos(\gamma)\sin(\phi) + N/2 \tag{18}$$
$$v = -r\sin(\gamma) + N/2 \tag{19}$$
$$w = r\ \cos(\gamma)\cos(\phi) + N/2 \tag{20}$$
$$id_{rect}(v_x, v_y) = |F(u,v,w)| \tag{21}$$

where the angles γ and ϕ are given by:

$$\gamma = -\frac{\pi}{2}\left(\frac{v_x - N_{rect}/2}{N_{rect}/2}\right) \qquad v_x = 1,\ \dots,\ N_{rect} \tag{22}$$
$$\phi = \frac{\pi}{2}\left(\frac{v_y - N_{rect}/2}{N_{rect}/2}\right) \qquad v_y = 1,\ \dots,\ N_{rect} \tag{23}$$

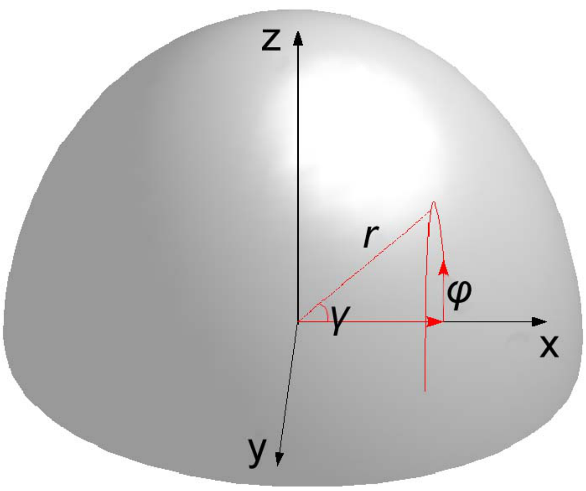


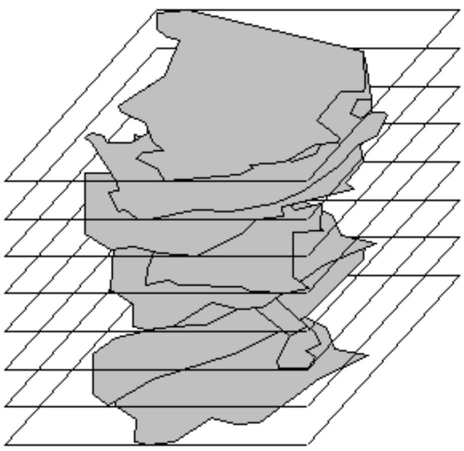
Figure 1: The angle γ (rotation around around x- or y-axis) points with the current radius *r* to positions which are projected to the y- or x-axis. From this point ϕ rotates and resamples from the 3D space.

The sampling function (18) describes the rectangle resample scheme (21) at a certain radius r within $F(u,v,w)$. This descriptor is like id_{rot} of square size with length N_{rect}. Roll and pitch can then be determined as 2D translations which again are found by phase matching.

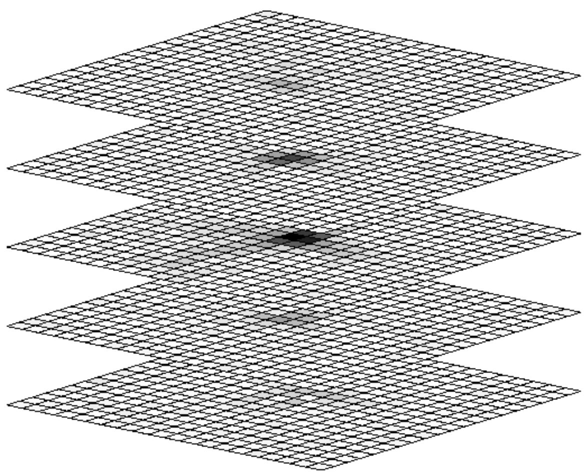
5.6 Multilayer parameter registration

The spectral structure at one certain radius is only sufficient for registration of data with significant amounts of overlap, e.g., when simulated data is used where the "scan"-pair consist of a scan and a rotated and translated counterpart that contains exactly the same data. Under more realistic conditions, especially in many applications like robot mapping, where interference, occlusions and sensor noise affect the data, registration using the spectral structure at a certain radius immediately becomes unstable. But the rotational information is present within the entire 3D spectrum at all radii. Resampling the structures according to Section 5.4 and 5.5 and by using the entire radius range can hence considerably improve the registration process. As also supported by experiments, this process becomes only more robust within a certain range. Concretely, a good range for the resampled hemisphere turns out to be from $(0.2...0.8) \cdot N/2$. This can be motivated as follows. Frequencies that are too low are simply likely to decrease the registration peak because of the insufficient available voxel data which can be resampled. Note that this effect of having a restricted number of voxels adds to the unavoidable distortions during the resampling process. Higher frequencies are obviously rather decreasing the information content caused by occluding and interfering structures.

The resulting descriptors in our method from where yaw and roll/pitch are determined are hence 3D data structures as illustrated in Figure 2, which have the intentionally redundant rotational information stacked along the z-axis.



(a) The descriptor as a stack of resampled layers.



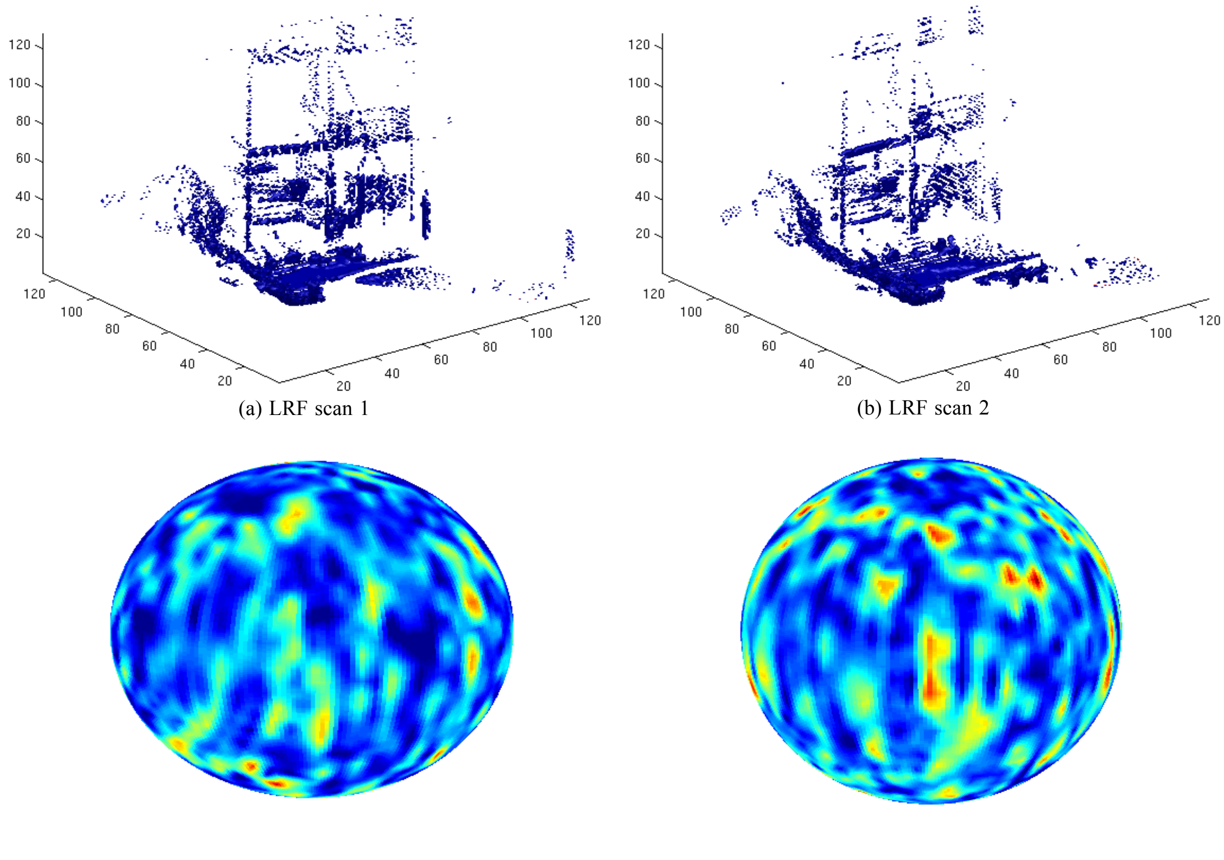
(b) Cutout of the corresponding peak after 3D POMF.

Figure 2: The 3D registration of the resampled radial information.

The relevant information is assembled within these 3D descriptors where phase matching is then applied in 3D. Note that the descriptor for yaw registration consists of preprocessed Fourier descriptors - see Section 5.4. The descriptor for roll/pitch contains the desired parameters as x and y translation, hence the POMF registration can directly be applied. Since these descriptors are resampled structures and not pixel-wise shifted signals, a postprocessing by an interpolation filter as described in [6] for an application of the FMT in 2D improves the stability of the registration.

The following example provides a better understanding of the entire sequence of descriptors which are necessary for the rotational registration. Figure 3 shows a scan pair and the corresponding surfaces of the spectrums at an arbitrary radius. These two scans are taken from a data-set

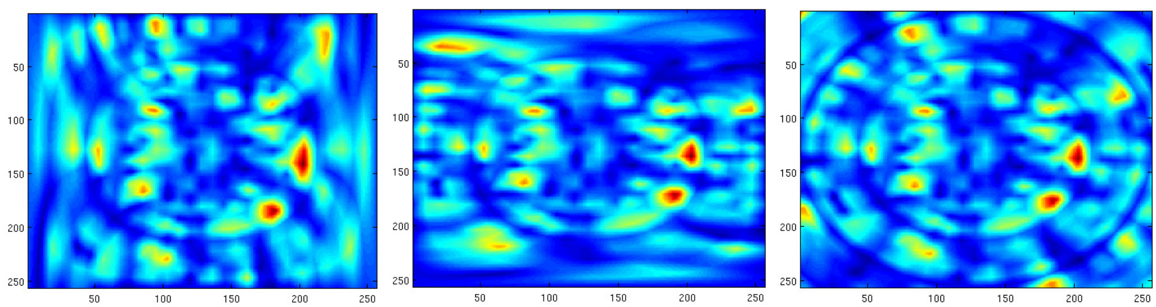
generated by a mobile robot with a Laser Range Finder (LRF) in a disaster scenario. Please note that this example pair is quite simple compared to the others in this data-set; there are relatively small interferences and occlusions between the two scans and there is a relatively small yaw rotation.
In the ideal case the corresponding yaw rotation should be visible within the structures of the spectrums magnitude; but as can be seen a match is here not feasible since it is not possible to identify common structures. This example visually motivates the need for the multilayer registration used in our method.



(a) LRF scan 1 (b) LRF scan 2

(c) Spectral spherical layer (scan 1). (d) Spectral spherical layer (scan 2).

Figure 3: An example scan pair generated by a mobile robot with a Laser Range Finder (LRF) and the corresponding 3D spectrums.



(a) xy resampling (b) yx resampling (c) rotational resampling

Figure 4: Resampling of different descriptors from the 3D spectrum (1 layer) from the example in Figure 3.

Figure 4 shows the corresponding resampled structures from the spectrum for the registration of roll/pitch and yaw. Figure 4(c) shows the rotational appearance for the yaw registration. In Figure 4(a) and Figure 4(b) two options of rectangular resampling are depicted. One is the 90∘ rotated counterpart of the other where the resampling in (18) is applied along the y axis instead of along the x axis. Ideally both results should be identical but the distortion effects as described in Section 5.5 are visible at both edges of the traversed resampling axis. Incorporating both forms of rectangular resampling into the descriptor hence improves the registration. The improvements with respect to robustness are especially observable when the roll and pitch angles are higher than 20∘ in both directions.
Concretely, the 3D descriptor for yaw is generated according to (24) and for roll/pitch according to (25):

$$\begin{aligned} id_{rot3D}(x,y,z) &= ID_{rotpol}(m,k) \\ z = 1,\ \ldots\ ,\ Z_{Nrot}\,&;\ m = x;\ k = y \end{aligned} \tag{24}$$

$$\begin{aligned} id_{rect3D}(x,y,z) &= id_{rect}(v_x,v_y) \\ z\ =\ 1,\ \ldots\ ,\ Z_{Nrect}\,&;\ v_x = x;\ v_y = y \end{aligned} \tag{25}$$

with

$$N_1 = 0.2\sqrt{N^3}$$

$$N_2 = 0.8\sqrt{N^3}$$

$$r = z \cdot N1$$

$$Z_{Nrect} = Z_{Nrot} = N_2 - N_1$$

The radius *r* is increasing giving new information of the spectral data *F*(*u,v,w*) as described in subsection 5.4 for yaw and in subsection 5.5 for roll/pitch. In case both of the alternatives of rectangular resampling - as shown in Figure 4(a) and Figure 4(b) - are incorporated, the length is $Z_{Nrect} = 2 \cdot Z_{Nrot}$. This is the recipe to process the frequency transformed scan data. Assume the spectrums of a scan pair are correlated according to (5) and the inverse FT is applied (6), the resulting function for yaw is denoted by $q_{idrot}(x,y,z)$ and for roll/pitch by $q_{idrect}(x,y,z)$. Within these functions 3D peaks are expected indicating the corresponding angle parameters. For yaw only a 1D shift along the y axis is expected which leads to:

$$\zeta(y) = q_{idrot}(N_{rot}/2, y, Z_{Nrot}/2) \tag{26}$$

For roll/pitch a 2D shift is expected, hence the 2D section is found lateral to the z-axis by:

$$\xi(x,y) = q_{idrect}(x, y, Z_{Nrect}/2) \tag{27}$$

The peak search is not in 3D because the structural shift maps the peaks to the center as given in (26) and (27). The positions $N_{rot}/2$, $Z_{Nrot}/2$ and $Z_{Nrect}/2$ correspond to the zero position for the Dirac peak when using for example the *fftshift* function in Matlab. For yaw the corresponding angle is found in 1D:

$$\gamma = (\zeta(y^*)/K) \cdot \pi \tag{28}$$

Roll and pitch can be determined from the resulting peak with:

$$(\alpha, \beta) = (\xi(x^*, y^*)/N_{rect}) \cdot \pi \tag{29}$$

where x^* and y^* are the supposed peak maximums according to (6).

5.7 The approximation effects of the resampling process

The resampling process is a projection of the spherical layer which leads inevitably to deviations from the true rotation. There are two effects which lead to deviations with increasing roll and pitch, namely effects on the yaw as well as on the roll/pitch determination. Both effects are now described in qualitative terms.

The points on the hemisphere rotated by roll and subsequently rotated by the pitch angle are subject to a mutual influence by the two angles. First, we consider the effect on yaw. A vivid illustration for the resampled and projected spherical structure is a plane that is rotated by roll/pitch and that is later on projected to the z-plane. Consider vectors pointing along the x-axis and y-axis that have rotations by pitch and roll. The resulting shifts are then cosine components. As soon as there is a rotation around both, i.e., $R_x(\alpha)R_y(\beta) \cdot [0\ 1\ 0]^T$, which yields $[\sin(\alpha)\sin(\beta)\ \cos(\alpha)]^T$, there is an undesired x-component. The z-component is here omitted due to the assumed projection. Only the rotation and the projection $R_x(\alpha)R_y(\beta) \cdot [1\ 0\ 0]^T$ directly along the x-axis is still an ideal case. But a first rotation around roll squeezes any pattern on that plane along the x-axis and a second rotation around pitch does not only squeeze it along the y-axis but it furthermore rotates the pattern away from the y-axis while the orientation on the x-axis remains constant. In case only one angle of roll/pitch is rotated or none of them, the yaw registration is hence correct. In case both roll and pitch are rotated at the same time in larger amounts, a constant offset depending on the sign is present.

A second, similar effect of mutual influence is present for roll and pitch. Assume a vector pointing along the z-axis where the sine components represent the resulting shifts. After rotation according to $R_x(\alpha)R_y(\beta) \cdot [0\ 0\ 1]^T$ the resulting projection is $[\cos(\alpha)\sin(\beta)\ \ -\sin(\beta)]^T$ where $\cos(\alpha)$ is the undesired factor.

These effects of the approximation due to the resampling involve two quite different aspects. First of all, our resampling leads to imprecisions in the determination of the angles. Second, the resampling limits the robustness of our method in case all three angles of yaw, roll, and pitch significantly change between two scans that are to be registered.

The first aspect, the systematic errors can be corrected by using a look-up table where for each rotation determined with our method the corresponding corrected values are stored. More precisely, a separate table for each angle is required which takes the roll, pitch and yaw registration results as input due to the mutual influence of all three angles and which has the compensation for the according angle as output. The simplest method to generate each table is a kind of calibration using artificially generated rotations with some arbitrary 3D data. The bunny and the dragon from the Stanford dataset are used for this purpose. The look-up values are generated from non-equidistant information and interpolated in 3D. This is comfortably done using for example a Delaunay triangulation. A convenient implementation is provided by the Matlab function *TriScatteredInterp*. The look-up operation is obviously computationally extremely fast, i.e., it does not introduce any overhead in the implementation of our method.

The second effect of the approximation due to the resampling is more substantial as it limits the amount of concurrent change in yaw, roll, and pitch. A comparison to related work in [9] shows that our method is very robust in case of small overlap between scans. This is further substantiated by experimental comparisons. This high robustness, i.e., very high success rates in case of partial overlap is due to our multilayer resampling in combination with the phase matching. This advantage is bought at the cost of the limitations in the maximum amount of concurrent rotations of yaw, roll, and pitch that can be handled. The resampling effects described before in this section also lead to a degeneration of the peaks in the phase matching up to a level where it does not succeed anymore.

5.8 Determination of translation

Once the rotation is correctly determined, the subsequent registration of the full 3D translation is straightforward. According to the determined angles (roll, pitch and yaw) the voxel data of the first scan is re-rotated and the corresponding 3D spectrum with the full phase information - see (1) - is calculated. Note that the 3D rotation of the scan data and a subsequent calculation of the spectrum is a better alternative compared to a direct rotation of the spectrum which would require an interpolation of complex data. Afterwards the 3D registration is again done by phase matching. This yields a distinguishable peak in the same way as already used for the multilayer angle registration.

6 Evaluation in Simulation

6.1 Setup

A simulation setup was developed within the UWSim [37] simulator. A wall with a rich texture was placed behind an actuated diver model. To achieve realistic movement of the diver, recorded DiverNet data was replayed and used to actuate the model. Furthermore, a generic AUV with a mounted stereo camera was also simulated. The AUV pose was perturbed significantly over time, resulting in both translational and rotational movement. The aim was to simulate a station keeping maneuver under influence of current. Figure 5 shows a screenshot of the simulation setup.

Figure 5: Screenshot of the simulation setup, including diver, AUV with mounted stereo camera, and wall segment. The stock UWSim AUV model is used for simplicity.

The experiment investigates two factors of disturbances for the motion compensation method: a) amount of dynamic content, and b) measurement noise in range readings. A total of 5 levels of dynamic content as well as 6 levels of noise were generated for a detailed analysis of both factors. Table 1 shows the exact values used for each level.

Level	Dynamic	Noise
1	None	None
2	3.5m	0.5%
3	2.5m	1.0%
4	2.0m	1.5%
5	1.5m	2.0%
6	-	3.0%

Table 1: Levels of dynamic content (diver distance to sensor, the closer the more dynamic content) and noise (percentage of range measurement, normally distributed).

Figures 6 and 7 show point clouds with varying noise levels (none, 1%, 3%) and dynamic content (3.5m, 2.5m, 1.5m), respectively. Note the significant range error present at 3%. Also, note how the occlusion at different distances of the diver to the sensor (Figure 7) is visible and shows the increasing amount of points on the moving diver.

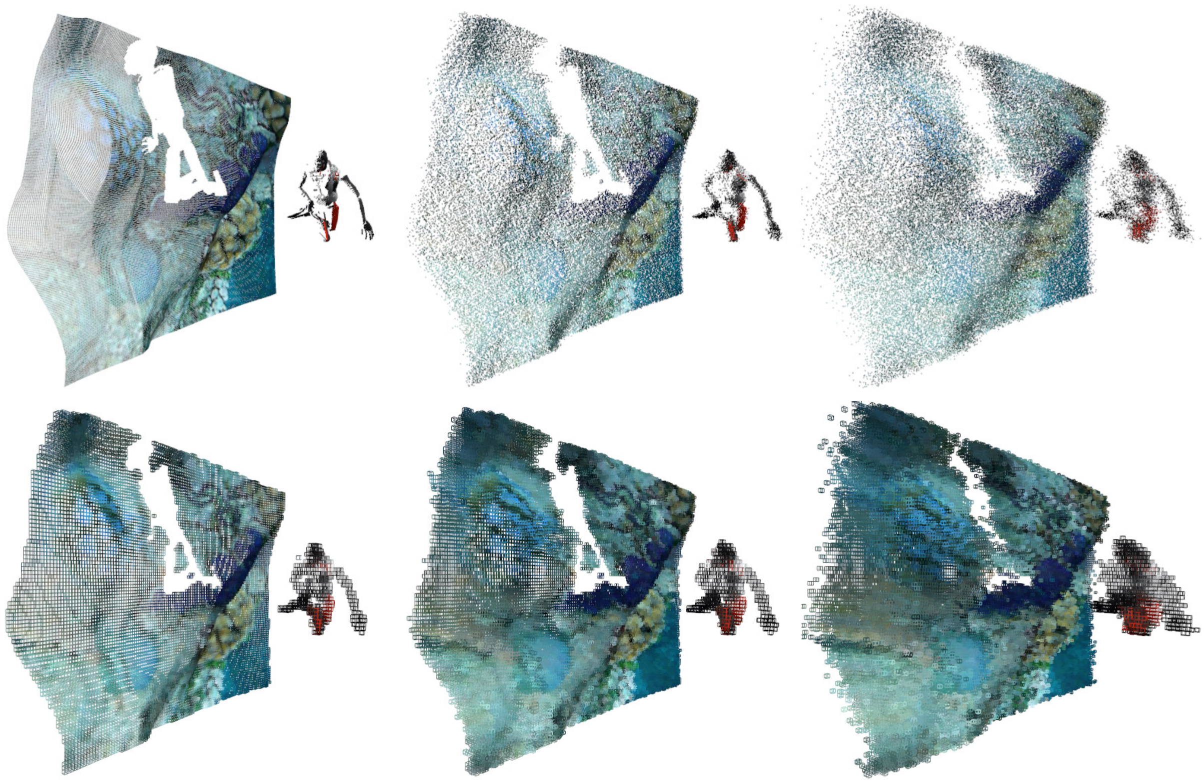


Figure 6: View No. 30 with different levels of noise. Left: No noise. Middle: 1% noise. Right: 3% noise. The top row shows plain point clouds, the bottom row shows discretized voxel grids as used in the spectral registration method. Note that the registration method did not use full color information, only intensity (greyscale).

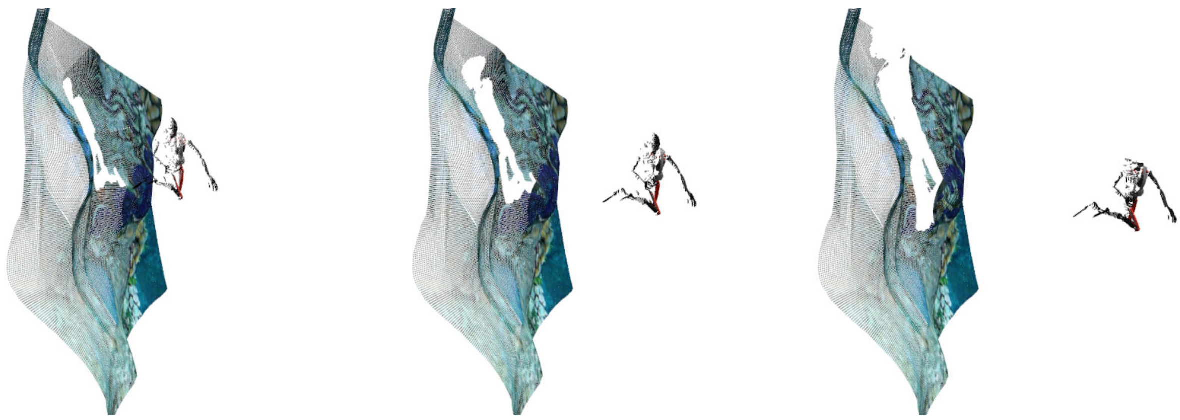


Figure 7: View No. 30 with different levels of dynamic content achieved by placing the diver at different distances from the sensor. Not the differing amounts of occlusion on the wall, showing the increased number of points on the moving diver. Left: 3.5m. Middle: 2.5m. Right: 1.5m.

A total of 233 point clouds were generated for each case, leading to a total of 5 × 6 × 233 = 6990 point clouds. The sensor trajectory as well as the diver actuation was kept exactly the same across all cases, eliminating random effects of slightly different occlusion, etc.

Point clouds were matched sequentially, simulating the motion estimation application as closely as possible. No global correction, e.g. via a Simultaneous Localization and Mapping (SLAM) technique, was performed.

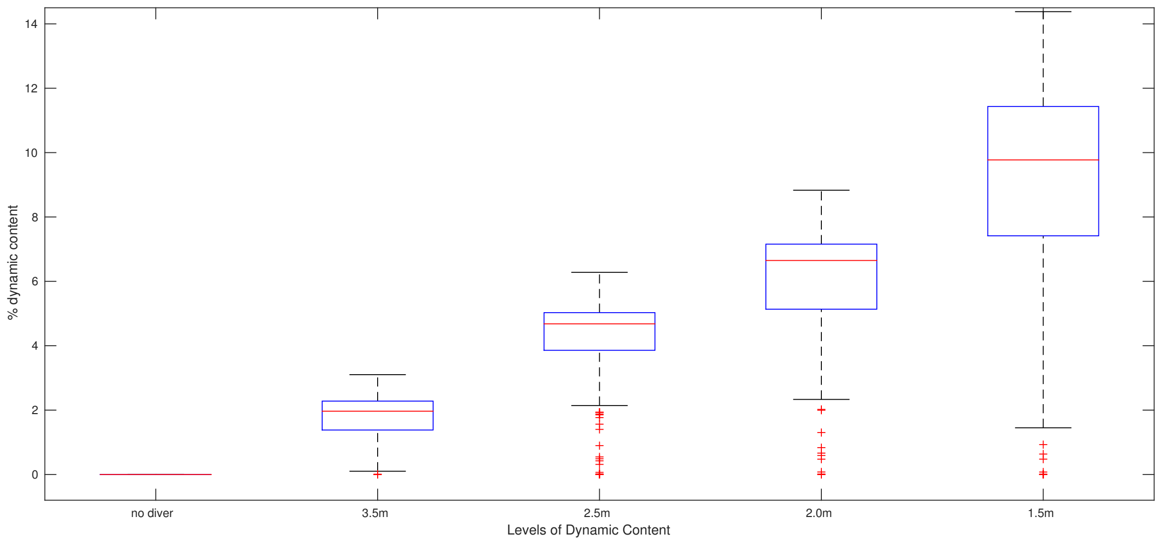


Figure 8: Box plots of number of changed range readings per level of dynamic content.

6.2 Results

Figure 9 shows the translation error relative to ground truth for each of the 30 cases. Figure 10 shows the rotation error. For each amount of dynamic content, a row of box plots shows the change of error over the different levels of range measurement noise. The box plots show the median, upper and lower quartiles, as well as outlier error readings. Note that the residual error for the case without dynamic content and without noise (top left) is due to the quantization error of converting the point cloud into a voxel grid representation. This representation is required for SRMR.

The results show that SRMR is very robust to significant levels of dynamic content. The error distribution does not change significantly along the vertical axis of plots. Motion estimation performance decreases only with very high levels of noise (last two columns).

7 Conclusion

This report described a spectral registration method applied to the problem of motion compensation using only sensor data with dynamic content. The method was shown to be robust against several levels of dynamic content as well as reasonable noise in the range data.

Further work will focus on segmentation of the moving diver, including arms and hands. The work described here facilitates this next step of processing.

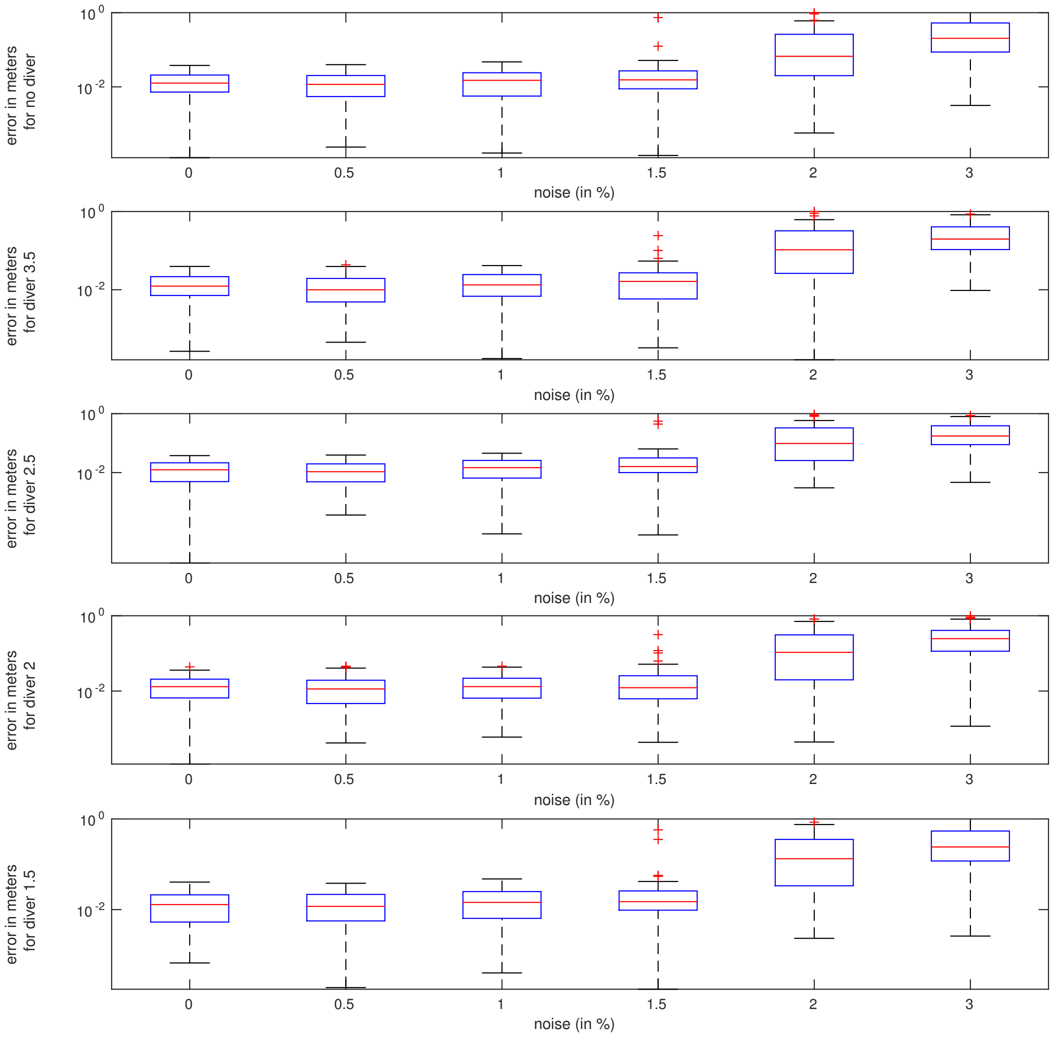


Figure 9: Box plots of translation error relative to ground truth. Each row denotes a different level of dynamic content (none at top, increasing dynamics down), and each column denotes different levels of noise (none on the left, increasing noise towards right).

8 References

[1] T. Acker. Biosonics underwater acoustic sentinel (uwacs) system for intruder detection. In *OCEANS 2009, MTS/IEEE Biloxi - Marine Technology for Our Future: Global and Local Challenges*, pages 1–4, 2009.

[2] A. Asada, F. Maeda, K. Kuramoto, Y. Kawashima, M. Nanri, and K. Hantani. Advanced surveillance technology for underwater security sonar systems. In *OCEANS 2007 - Europe*, pages 1–5, 2007.

[3] M. Bertozzi, A. Broggi, A. Fascioli, T. Graf, and M.-M. Meinecke. Pedestrian detection for driver assistance using multiresolution infrared vision. *Vehicular Technology, IEEE Transactions on*, 53(6):1666–1678, 2004.

[4] Paul J. Besl and Neil D. McKay. A method for registration of 3-d shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(2):239–256, Feb 1992.

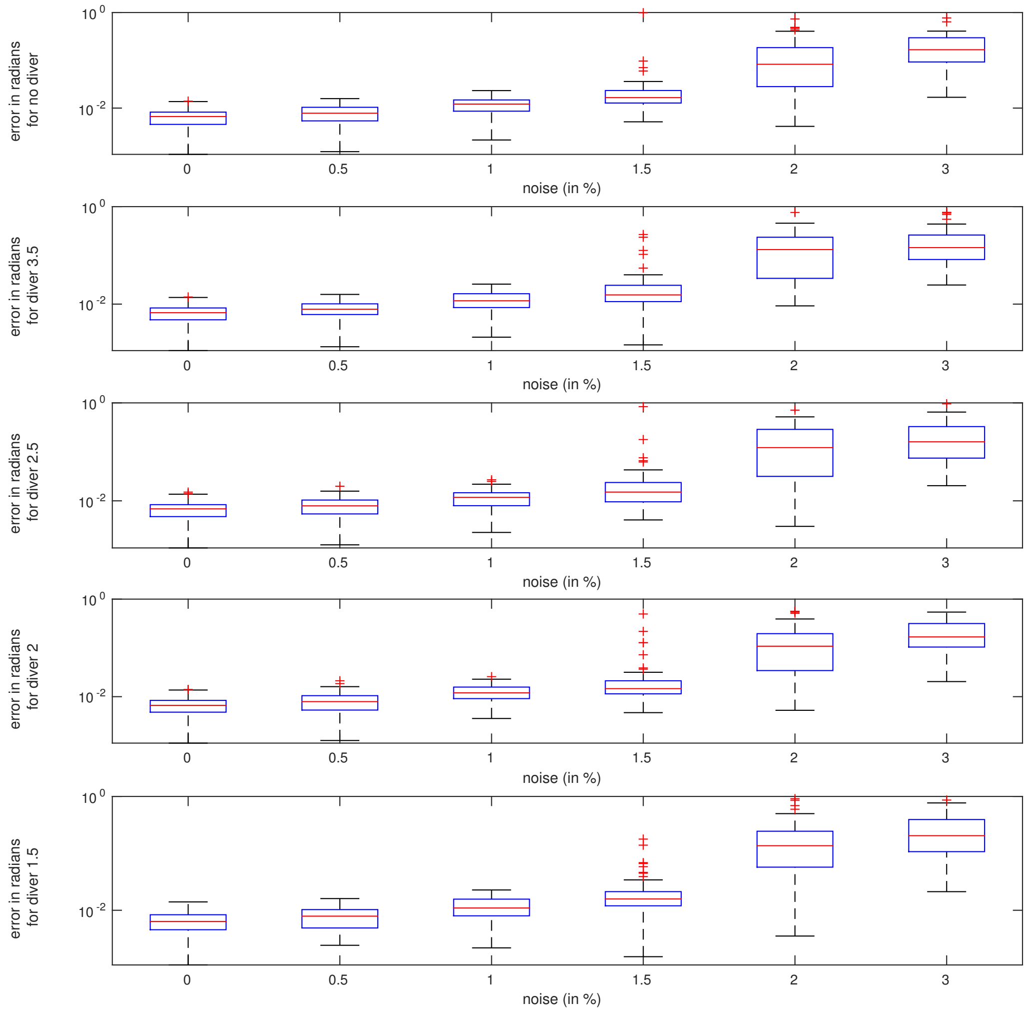


Figure 10: Box plots of rotation error relative to ground truth. Each row denotes a different level of dynamic content (none at top, increasing dynamics down), and each column denotes different levels of noise (none on the left, increasing noise towards right).

[5] E. Brekke, O. Hallingstad, and J. Glattetre. The signal-to-noise ratio of human divers. In *OCEANS 2010 IEEE - Sydney*, pages 1–10, 2010.

[6] H. Buelow, A. Birk, and V. Unnithan. Online generation of an underwater photo map with improved Fourier-Mellin based registration. In *International OCEANS Conference, IEEE Press*, 2009.

[7] Heiko B_low and Andreas Birk. Diver detection by motion-segmentation and shape-analysis from a moving vehicle. In *IEEE Oceans*, 2011.

[8] Heiko B_low and Andreas Birk. Gesture-recognition as basis for a human robot interface (hri) on a auv. In *IEEE Oceans*, 2011.

[9] Heiko B_low and Andreas Birk. Spectral 6-dof registration of noisy 3d range data with partial overlap. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 35(4):954–969, 2013.

[11] X. Chen, R. Wang, and U. Tureli. Passive acoustic detection of divers under strong interference. In *OCEANS 2006*, pages 1–6, 2006.

[12] A.M. Crawford and D. Vance Crowe. Observations from demonstrations of several commercial diver detection sonar systems. In *OCEANS 2007*, pages 1–3, 2007.

[13] E. De Castro and C. Morandi. Registration of translated and rotated images using finite Fourier transforms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-9(5):700 – 703, 1987.

[14] A. Engelsberg and G. Schmidt. A comparative review of digital image stabilising algorithms for mobile video communications. *Consumer Electronics, IEEE Transactions on*, 45(3):591–597, 1999.

[15] L. Fillinger, P. de Theije, M. Zampolli, A. Sutin, H. Salloum, N. Sedunov, and A. Sedunov. Towards a passive acoustic underwater system for protecting harbours against intruders. In *Waterside Security Conference (WSS), 2010 International*, pages 1–7, 2010.

[16] H. Foroosh, J. Zerubia, and M. Berthod. Extension of phase correlation to subpixel registration. *IEEE Transactions on Image Processing*, 11(3):188–200, March 2002.

[17] W.S. Hoge. A subspace identification extension to the phase correlation method. *IEEE Transactions on Medical Imaging*, 22(2):277–280, February 2003.

[18] W.S. Hoge and C.F. Westin. Identification of translational displacements between n-dimensional data sets using the high-order svd and phase correlation. *IEEE Transactions on Image Processing*, 14(7):884–889, July 2005.

[19] J. L. Horner and P. D. Gianino. Phase-only matched filtering. *Applied Optics*, 23:812–816, 1984.

[20] Sheng-Che Hsu, Sheng-Fu Liang, Kang-Wei Fan, and Chin-Teng Lin. A robust in-car digital image stabilization technique. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(2):234–247, 2007.

[21] F.M. Jaffre, T.C. Austin, B.G. Allen, R. Stokey, and C.J. Von Alt. Ultra short baseline acoustic receiver/processor. In *Oceans 2005 - Europe*, volume 2, pages 1382–1385 Vol. 2, 2005.

[22] J.S. Jin, Zhigang Zhu, and Guangyou Xu. A stable vision system for moving vehicles. *Intelligent Transportation Systems, IEEE Transactions on*, 1(1):32–39, 2000.

[23] A.T. Johansson, R.K. Lennartsson, E. Nolander, and S. Petrovic. Improved passive acoustic detection of divers in harbor environments using pre-whitening. In *OCEANS 2010*, pages 1–6. IEEE, 2010.

[24] Boyoon Jung and Gaurav S. Sukhatme. Detecting moving objects using a single camera on a mobile robot in an outdoor environment. In *8th Conference on Intelligent Autonomous Systems*, pages 980–987, 2004.

[25] Ronald T. Kessel and Reginald D. Hollett. Underwater intruder detection sonar for harbour protection: State of the art review and implications. In *Second IEEE International Conference on Technologies for Homeland Security and Safety*, 2006.

[26] Peter Kostelec and Daniel Rockmore. Ffts on the rotation group. *Journal of Fourier Analysis and Applications*, 14(2):145–179, 2008.

[27] P.J. Kostelec and D.N. Rockmore. FFTs on the rotation group. *Working Papers Series, Santa Fe Institute*, 2003.

[28] R.K. Lennartsson, E. Dalberg, T. Fristedt, E. Nolander, and L. Persson. Electric detection of divers in harbor environments. In *OCEANS 2009, MTS/IEEE Biloxi - Marine Technology for Our Future: Global and Local Challenges*, pages 1–8, 2009.

[29] R.K. Lennartsson, E. Dalberg, A.T. Johansson, L. Persson, S. Petrovic, and E. Rabe. Fused passive acoustic and electric detection of divers. In *Waterside Security Conference (WSS), 2010 International*, pages 1–8, 2010.

[30] R.K. Lennartsson, E. Dalberg, L. Persson, and S. Petrovic. Passive acoustic detection and classification of divers in harbor environments. In *OCEANS 2009, MTS/IEEE Biloxi - Marine Technology for Our Future: Global and Local Challenges*, pages 1–7, 2009.

[31] Yu-Ming Liang, Hsiao-Rong Tyan, Shyang-Lih Chang, H.-Y.M. Liao, and Sei-Wang Chen. Video stabilization for a camcorder mounted on a moving vehicle. *Vehicular Technology, IEEE Transactions on*, 53(6):1636–1648, 2004.

[32] A. Lovik, A.R. Bakken, J. Dybedal, T. Knudsen, and J. Kjoll. Underwater protection system. In *OCEANS 2007*, pages 1–8, 2007.

[33] R. Mueller and C. Brook. Vessel protection in expeditionary operations: At anchor and in foreign harbours. In *Waterside Security Conference (WSS), 2010 International*, pages 1–6, 2010.

[34] R. Muller and C. Brook. Net based waterside security applications: From small solutions to maritime security networks. In *Waterside Security Conference (WSS), 2010 International*, pages 1–6, 2010.

[35] A. Ollero, J. Ferruz, F. Caballero, S. Hurtado, and L. Merino. Motion compensation and object detection for autonomous helicopter visual navigation in the comets system. In *Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on*, volume 1, pages 19–24 Vol.1, 2004.

[36] A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing*. Prentice Hall Signal Processing Series, Englewood Cliffs, 1989.

[37] M. Prats, J. Perez, J.J. Fernandez, and P.J. Sanz. An open source tool for simulation and supervision of underwater intervention missions. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pages 2577–2582, Oct 2012.

[38] A. Rodningsby and Y. Bar-Shalom. Tracking of divers using a probabilistic data association filter with a bubble model. *Aerospace and Electronic Systems, IEEE Transactions on*, 45(3):1181–1193, 2009.

[39] J. Sattar and G. Dudek. Where is your dive buddy: tracking humans underwater using spatio-temporal features. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 3654–3659, 2007.

[40] Junaed Sattar and Gregory Dudek. Underwater human-robot interaction via biological motion identification. In *Robotics: Science and Systems (RSS)*, 2009.

[41] Junaed Sattar and Gregory Dudek. A vision-based control and interaction framework for a legged underwater robot. In *Proceedings of the Sixth Canadian Conference on Robot Vision (CRV)*, pages 329–336, 2009.

[42] T. Schamm, M. Strand, T. Gumpp, R. Kohlhaas, J.M. Zollner, and R. Dillmann. Vision and tof-based driving assistance for a personal transporter. In *International Conference on Advanced Robotics (ICAR)*, pages 1–6. IEEE Press, 2009.

[43] R. Stolkin and I. Florescu. Probability of detection and optimal sensor placement for threshold based detection systems. *Sensors Journal, IEEE*, 9(1):57–60, 2009.

[44] R. Stolkin, S. Radhakrishnan, A. Sutin, and R. Rountree. Passive acoustic detection of modulated underwater sounds from biological and anthropogenic sources. In *OCEANS 2007*, pages 1–8, 2007.

[45] A. Sutin, B. Bunin, A. Sedunov, N. Sedunov, L. Fillinger, M. Tsionskiy, and M. Bruno. Stevens passive acoustic system for underwater surveillance. In *Waterside Security Conference (WSS), 2010 International*, pages 1–6, 2010.

[46] F. Vella, A. Castorina, M. Mancuso, and G. Messina. Digital image stabilization by adaptive block motion vectors filtering. *Consumer Electronics, IEEE Transactions on*, 48(3):796–801, 2002.

[47] M. Watson, C. Loggins, and Y.T. Ochi. A new high accuracy super-short baseline (ssbl) system. In *Underwater Technology, 1998. Proceedings of the 1998 International Symposium on*, pages 210–215, 1998.

[48] Qiaoyun Zhou, Shiqi Yu, Xinyu Wu, Qiao Gao, Chongguo Li, and Yangsheng Xu. Hmms-based human action recognition for an intelligent household surveillance robot. In *Robotics and Biomimetics (ROBIO), 2009 IEEE International Conference on*, pages 2295–2300, 2009.