Grant Agreement No. 611373

FP7-ICT-2013-10

D2.5 Algorithms for diver pose estimation through remote and local sensing

Due date of deliverable: 31/05/2016
Actual submission date: 09/06/2016

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

Organization name of lead contractor for this deliverable: JACOBS

Revision (draft, version 1,2,3…)

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 Diver behaviour

2.1 Dataset Description

Diver data was collected in the Y-40 pool in Padova in June 2014, February 2015 and June 2015. Divers were equipped with the current version of DiverNet and asked to perform a series of tasks underwater. In total, 56 people participated in the experiment, 36 sessions yielded reliable motion data.

	Men (mean age)	Women (mean age)
Padova 06.2014	12 (48,08yrs)	4 (39,25yrs)
Padova 02.2015	11 (44,0yrs)	8 (46,75yrs)
Padova 06.2015	4 (52,0yrs)	17 (44,1yrs)
Total	27 (47,0yrs)	29 (44,34yrs)

Table 2.1. *Gender and age distribution of the recorded divers during all data collection trials.*

2.2 Data annotation

We used ANVIL to code low-level behavior of part of the dataset according to a low-level Behavior repertoire. The repertoire data was incorporated in the first analysis tool of diver behavior – the DiverControlCenter.

As a higher level of annotation proved more useful for behavior interpretation and segmentation, we developed a real-time annotation tool based on the diver tasks. We collected time-synchronized video material for all data collections for comparison with DiverNet data. The DiverNet motion recording is consistent with the motion displayed on the video material for all 36 cases mentioned above.

DiverNet data and Video data was segmented according to these real-time annotations. The behavior segments were uploaded to the CADDY server. The DiverControlCenter can segment and analyze all DiverNet data.

Task list:

- breathing with regulator
- breathing without regulator
- taking off the mask
- taking off the regulator
- controlling buoyancy
- moving up and down between 2 targets
- moving in the horizontal plane at a slow pace
- swimming quickly in a horizontal plane
- displacing an object
- free behaviour task
- T-posture

The low-level behaviour repertoire for all diver limbs and breathing activity are shown in the next tables.

Behaviour	Definition
hand signal	hand performs defined diving signal
manipulating own diving equipment	hand touches, grabs, moves or picks up object related to the diver's equipment
manipulating other diver's equipment	hand touches, grabs, moves or picks up object related to another diver's equipment
touching object	hand touches object (except equipment)
picking up object	hand grabs and picks up object (except equipment)
carrying object	upper limbs carry object
touching person	hand touches another person's body
touching own body	hand touches part of the own body
touching floor	hand touches floor
hanging on to something	hand hangs on to object, floor or person
still	upper limbs still without contact with equipment, object, person or own body parts
moving	upper limbs moving without contact with equipment, object, person or own body parts
none	

Table 2.2. *Low-level behaviour for upper limbs right.*

Behaviour	Definition
hand signal	hand performs defined diving signal
manipulating own diving equipment	hand touches, grabs, moves or picks up object related to the diver's equipment
manipulating other diver's equipment	hand touches, grabs, moves or picks up object related to another diver's equipment
touching object	hand touches object (except equipment)
picking up object	hand grabs and picks up object (except equipment)
carrying object	upper limbs carry object
touching person	hand touches another person's body

touching own body	hand touches part of the own body
touching floor	hand touches floor
hanging on to something	hand hangs on to object, floor or person
still	upper limbs still without contact with equipment, object, person or own body parts
moving	upper limbs moving without contact with equipment, object, person or own body parts
none	

Table 2.3. *Low-level behavior for upper limbs left.*

Behaviour	**Definition**
still bent	lower limbs do not move, while they are located in open water (no contact with floor)
kneeling/touching floor	parts of the lower limbs (except the bottom of the foot) touch the floor
standing	body upright on the feet which touch the floor
paddling	lower limbs move smoothly, while angle between shank and thigh changes
none	

Table 2.4. *Low-level behaviour for lower limbs right.*

Behaviour	**Definition**
still bent	lower limbs do not move, while they are located in open water (no contact with floor)
kneeling/touching floor	parts of the lower limbs (except the bottom of the foot) touch the floor
standing	body upright on the feet which touch the floor
paddling	lower limbs move smoothly, while angle between shank and thigh changes
none	

Table 2.5. *Low-level behaviour for lower limbs left.*

Behaviour	Definition
expiration	air bubbles visible near the diving regulator
none	

Table 2.6. *Low-level behaviour for diver breathing.*

3 Diver pose estimation

3.1 DiverNet Model

DiverNet is a network of 17 9-axis inertial measurement units (IMU) mounted on the diver. Each IMU is mounted on one of the body parts - 3 on each arm and leg, 1 for each shoulder, 1 on torso, 1 on lower back and 1 on head - and is used to calculate the orientation of that body part. The individual orientations are fused together in a complete human diver kinematic model. Fig. 3.1 shows the positions of IMUs on diver's body.

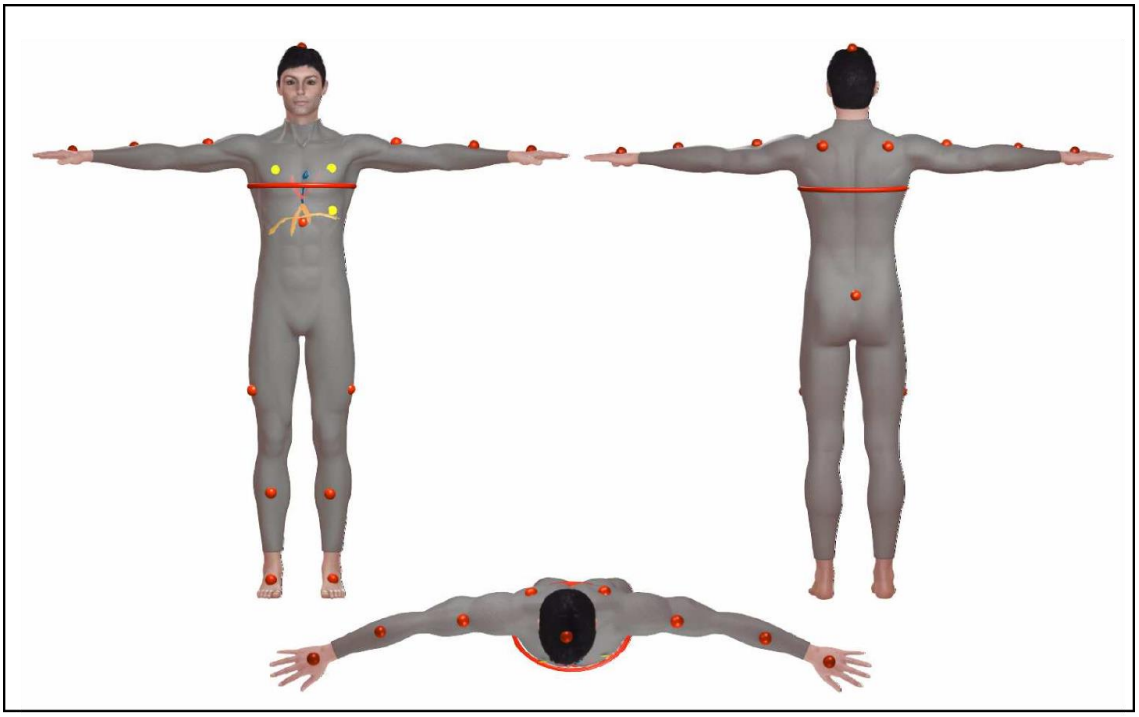

Figure 3.1. *IMU positions on diver's body*

The mounting of the sensors is done with a special diver suit with sewn-in velcro patches, where the sensors have a small piece of velcro glued on their back. This is displayed in Fig. 3.2. However, it is impossible to have the sensor perfectly aligned and firmly in place, and sometimes a sensor can rotate. Because of that, a calibration procedure is performed, where the diver makes a T-posture (like in Fig. 3.1), and the coordinate frame of each sensor is rotated so that it matches the expected orientation of that body part in the T-posture.

Figure 3.2. *DiverNet mounted on the diver*

3.1.1 Diver posture estimation and visualization

In order to estimate the orientation of each body part, a complementary filter described in [1] is used. This filter was chosen over often used Kalman filter-based solutions as it has shown to perform with similar accuracy, but has a much lower computational cost.

After the absolute orientation of each individual body part is calculated, this data is fused in a complete model, where the orientations are translated from absolute to relative, and this data is used to visualize the model. The entire processing is done in ROS in C++, and Rviz is used for visualization. Fig. 3.3 displays the visualized diver model.

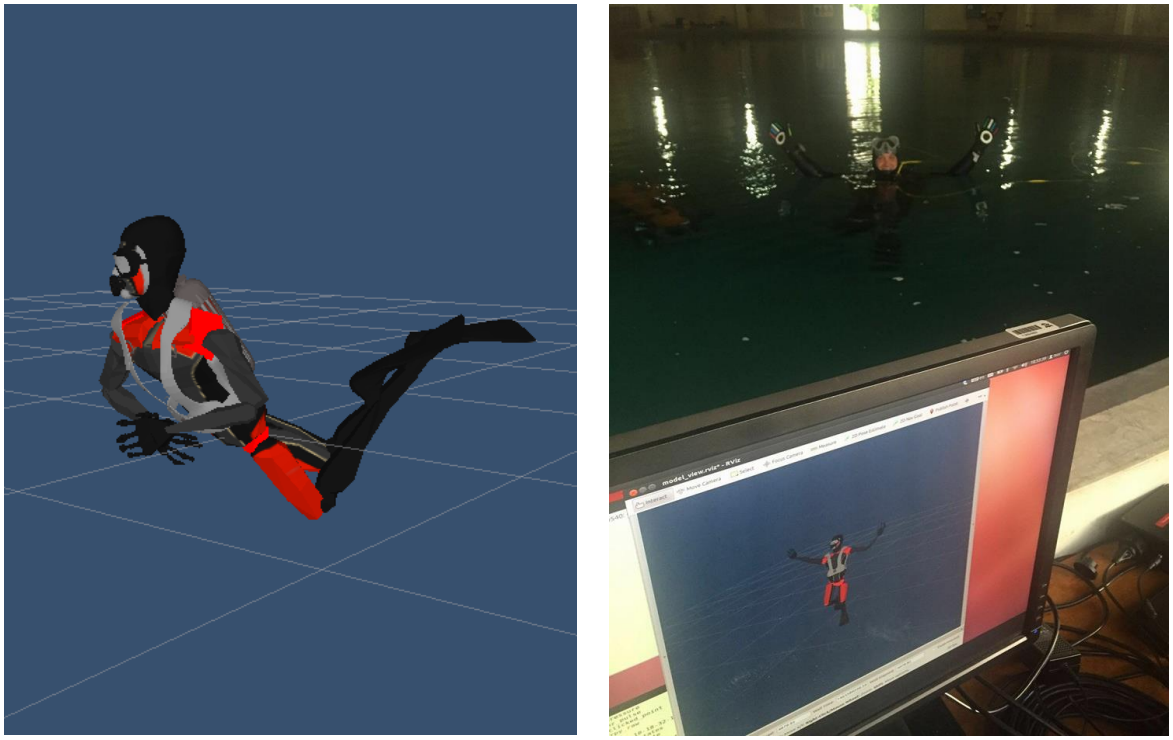

Figure 3.3. *(left) Diver model displayed in Rviz; (right) Diver in the background*

3.1.2 Automatic diver activity classification

The reconstructed diver posture obtained from DiverNet is used for automatic activity classification, where the system would know what the diver is doing even without a human operator observing. Two approaches are currently being tested. The first one uses a Dynamic time warping algorithm to align the live data with recorded training data. It classifies current activity based on the distance to aligned training set, choosing the activity with the smallest distance. This method showed good results with static postures, but is not as appropriate for dynamic activity which is important for most diving activities. Tests were conducted with 7 poses shown in Figure 3.4. Confusion matrix for initial tests is shown in Figure 3.5.

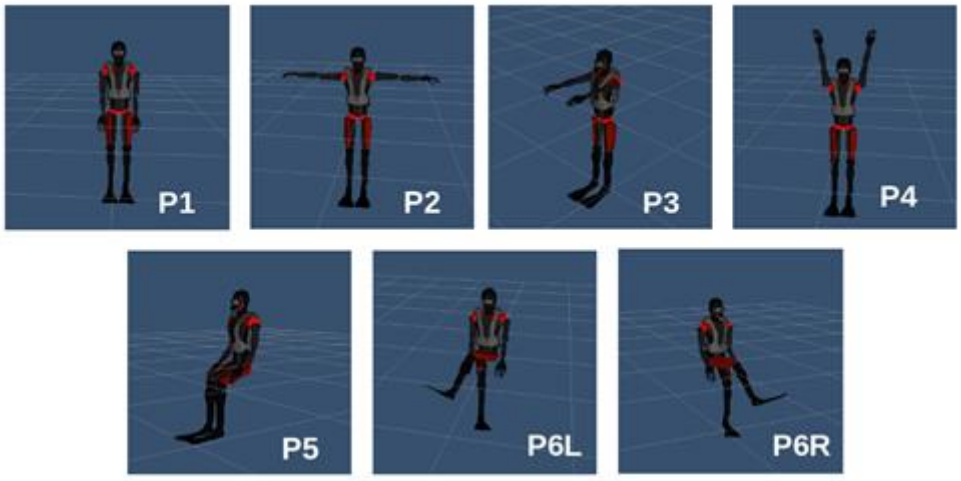

Figure 3.4. *Poses used in initial classification tests with Dynamic time warping*

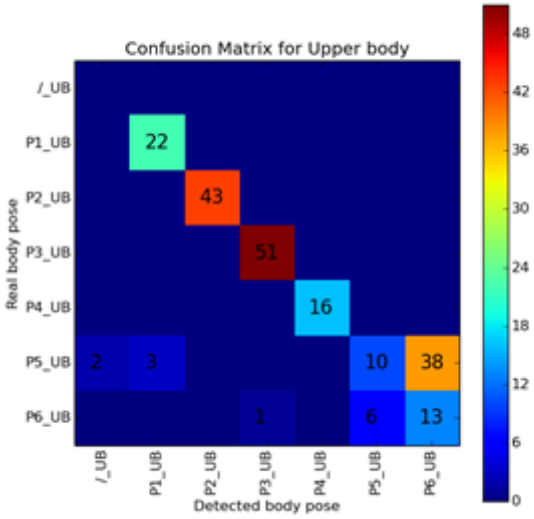

Figure 3.5. *Confusion matrix for Dynamic time warping tests*

The second approach that is currently being tested are artificial neural networks. Inputs to the network are calculated joint orientations of the diver, for both current and past frames. This allows time awareness and recognition of dynamic activity. Neural network models are currently being assessed to find a good structure for our task. A simple and shallow (1 hidden layer) feed-forward neural network has been tested so far, and has shown good results on a small dataset with static poses. Tests with dynamic activities are currently being performed.

3.2 Stereo-based estimation – Local sensing

The previously discussed DiverNet offers good accuracy when there is a correct calibration of the inertial sensors. It provides information about the main body parts of the diver, which helps to generate a complete human model. However, this information is passed through acoustics from the diver suit's modem to the underwater vehicle (BUDDY) which causes a delay in the information and makes processing available only every 5 seconds approximately. For the application, it is not indispensable to have a full pose estimation of the diver's limbs at all times, but it is important to at least have the diver's heading i.e. the direction to where the diver is swimming to. In this way, the BUDDY can position itself always in front of the diver in case it has to guide him to a certain location or if it needs to record the diver's gestures to perform a certain task.

To obtain the diver's heading between each DiverNet's measurement, we compute principal component analysis (PCA)[2] in the point clouds generated from the stereo images. In dense point clouds, we can visually find a correlation between the PCA eigenvectors and the pose of the diver. For example, when the diver is almost facing the camera, the smallest or previous to smallest eigenvector is perpendicular to the diver's chest area indicating the diver's heading (or the parallel opposite). Likewise, when the diver has turned 90° from the camera, the greatest eigenvector goes across the diver's body; so its projection to the Y-Z plane indicates the diver's orientation. Figure 3.6 shows some examples where the PCA eigenvectors can be visually correlated to the diver's heading.

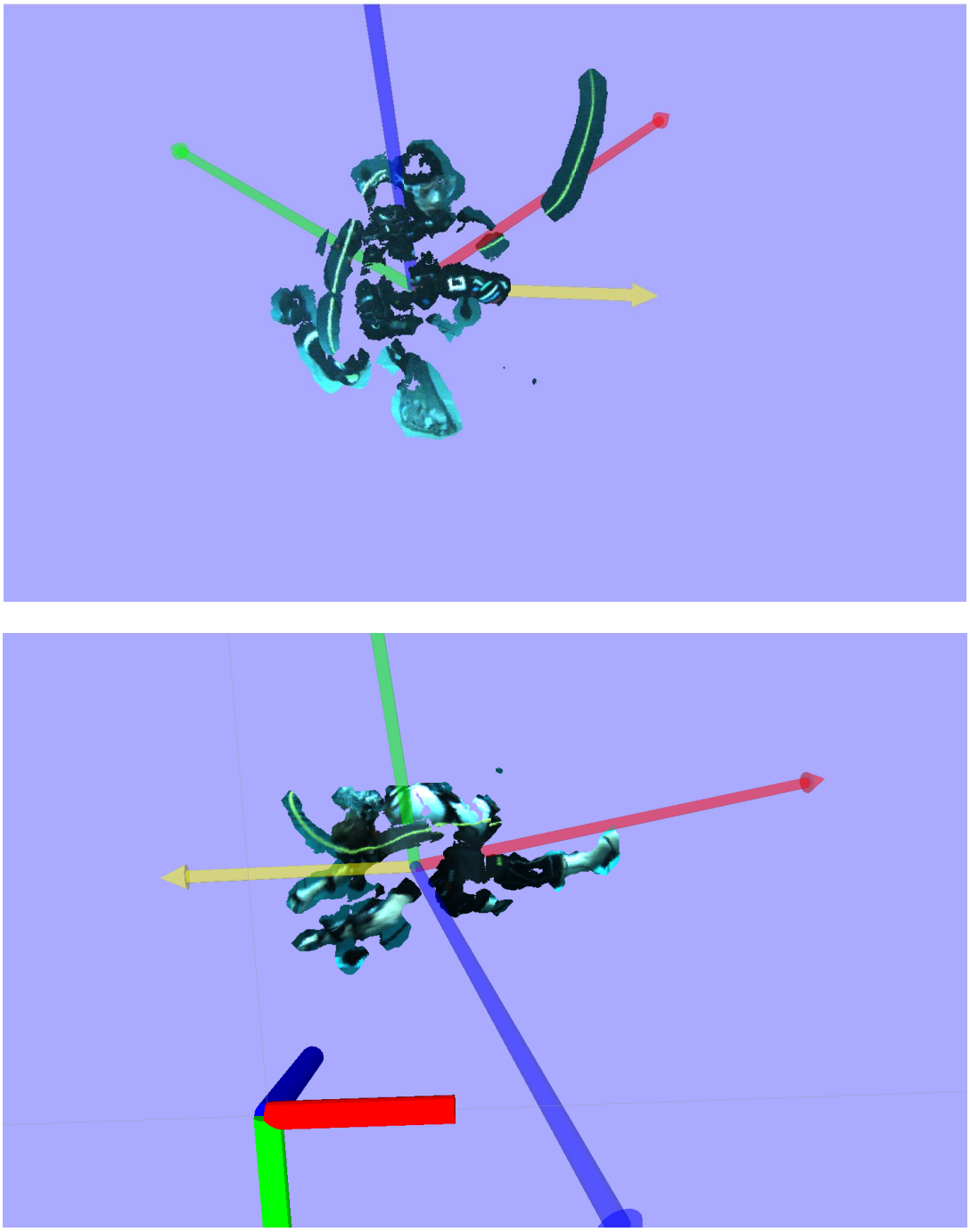

Figure 3.6. *Diver point clouds with PCA eigenvectors superimposed, color coded as Red, Green, Blue from the one with the greatest eigenvalue to the lowest. Yellow arrow indicates the diver heading according to our method. (Top) The green eigenvector is parallel - inverse - to the diver's heading. (Bottom) The red eigenvector is almost parallel - inverse - to the diver's heading. It's projection in the Y-Z plane, formed by the blue and red axes shown in the image, would be more accurate.*

It would be possible to establish a relation between PCA eigenvectors and every possible diver's pose; however, point clouds from stereo images rely on feature matching and the rather textureless underwater environment (diver suit an uniform background) causes the point clouds to be sparse or with holes, possibly differing substantially from frame to frame. Hence, the relation between PCA eigenvectors and diver's heading is not one-to-one anymore, but there's still some statistical

correlation. For this reason, a Random Forest classifier [3] was trained with collected data from 20 divers (see section 3.2.2 below); with enough data, the classifier can find these non-obvious relations and assign degrees of confidence (probabilities).

3.2.1 Pointcloud processing

In order to have reliable information about the diver's pose; the point cloud should not include many points outside the diver's body. For this reason, a series of image processing techniques are used in the generated disparity map. First a median filter is applied to eliminate random noise caused by feature mis-matching. Then, erosion and dilation processes are done sequentially to eliminate holes within the point cloud and small sets of isolated points; usually these sets are not nearly as big as the set that encompasses the diver's body. The resultant image (the rightmost in Figure 3.7) is used as a mask to only use the points on those coordinates as point cloud generators; in this way, PCA will provide more information about the diver and not the surrounding environment. Otherwise, a lot of points from the background, bubbles, reflections, etc, would have formed part of the point cloud. And all of this unnecessary information would have made the PCA computation less accurate. This is shown in the disparity map below.

Figure 3.7. *(Left) Right image from the stereo-rig. (MIddle) Disparity map generated by feature matching. (Right) Resultant mask after erosion and dilation cycles in the disparity map to eliminate noise and holes.*

3.2.2 Dataset Creation

To create the training dataset, we asked 20 participants to rotate in front of the stereo camera 360° in vertical and horizontal position (Padova, Italy; February, 2016). Since the classifier cannot output continuous values, all possible orientations were quantized in 8 bins of 22.5° as shown in the Figure 3.9.

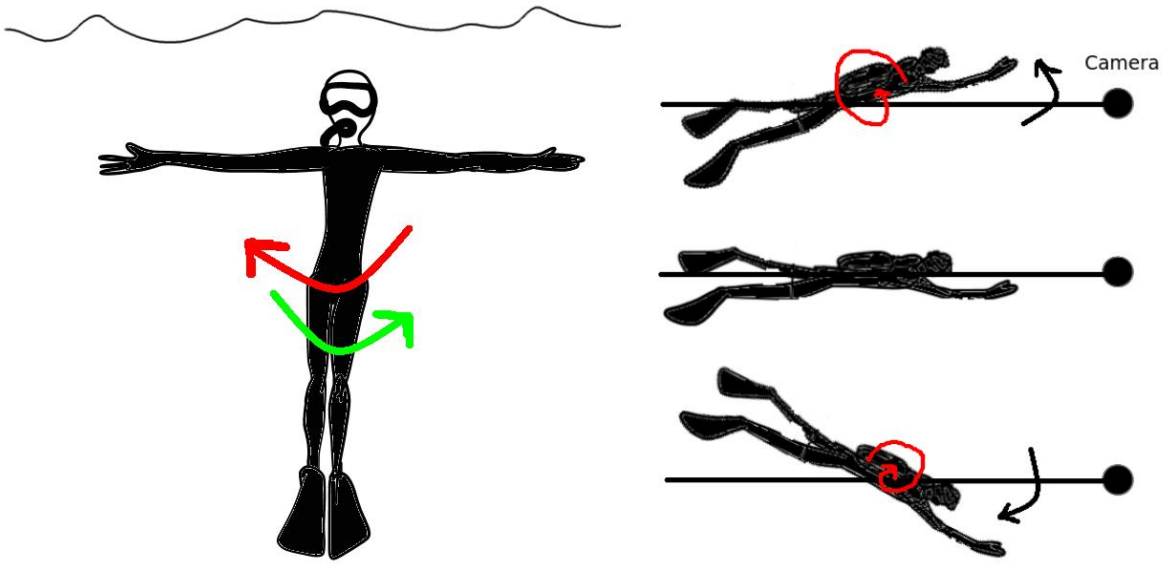

Figure 3.8. *Diagram showing how the the diver should position itself vertically (left) and horizontally (right) in front of the camera and make 360° turns in a clockwise and anticlockwise manner.*

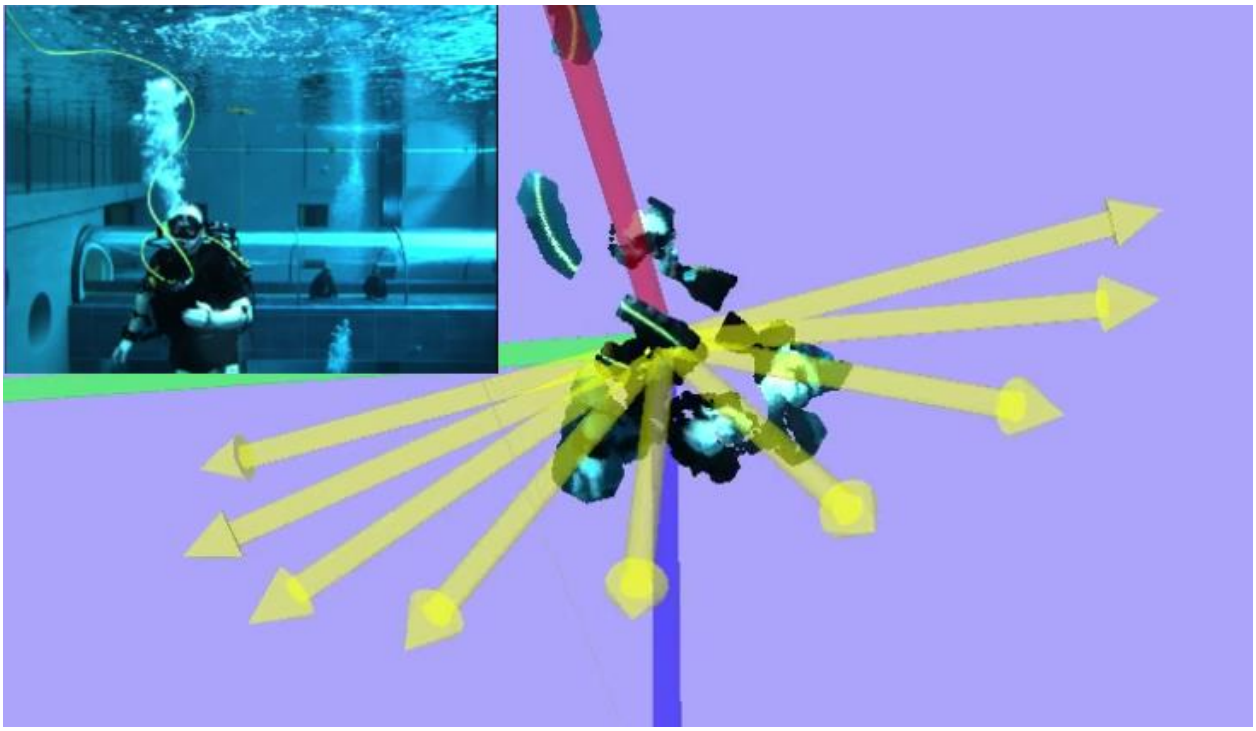

Figure 3.9. *Point cloud of the diver shown in the top left corner. Yellow arrows divide the Y-Z plane into 8 angular bins from -90° to 90° (zeros degrees corresponds to the diver facing the stereo camera).*

We are only interested in headings in the range of -90° and 90° because when BUDDY is looking at the diver from behind, it knows it has to quickly move in front of the diver; once in front, finer movements are necessary to position itself exactly in front of the diver (face to face). Based upon these constraints, the data was manually labelled by watching the diver's rotate while passing through visual markers, this is shown in Figure 3.9 as yellow arrows. Then we build a 15 element feature vector from the labelled data as such:

Eigen - value[0]	Eigen - vector[0]	Eigen - value[1]	Eigen - vector[1]	Eigen - value[2]	Eigen - vector[2]	Maximum variations (distance) along the eigenvectors

The first 12 elements of the feature are the ordered eigenvectors (3 dimensions) and their respective eigenvalues (1 dimension); and the last 3 elements are the maximum variations or absolute distances between pointcloud points along each eigenvector. As a result we have a dataset with 1270 samples, which were divided in sets of 889, 254 and 127 (70%,20%,10%) to create the training, validation and test set respectively and train through cross validation.

3.2.3 Results

On the test set the Extremely Randomized Forest classifier [4] achieves *71.6% accuracy*. However, as it can be seen Fig. 3.10, the majority of the predicted headings lie within 1 or 2 bins of the true heading. So, we can state that even when the classifier outputs an incorrect value, there is a 90% probability (less than 2 bins error) that it will output a heading close enough to the true value for it to be used by the control filter mentioned in Section 3.1.1. In Figure 3.10 we can also see that there are some errors of 6-7 bins difference; this happens only when the diver is turned 90° from the camera and the classifier's prediction is 90° heading but in the opposite direction. A hypothesis for this error is that the point clouds have similar distributions when the diver is facing in these directions; nonetheless, this can be solved in future work by including temporal data (e.g. kalman filter) because the diver cannot change direction so abruptly within the camera frame rate.

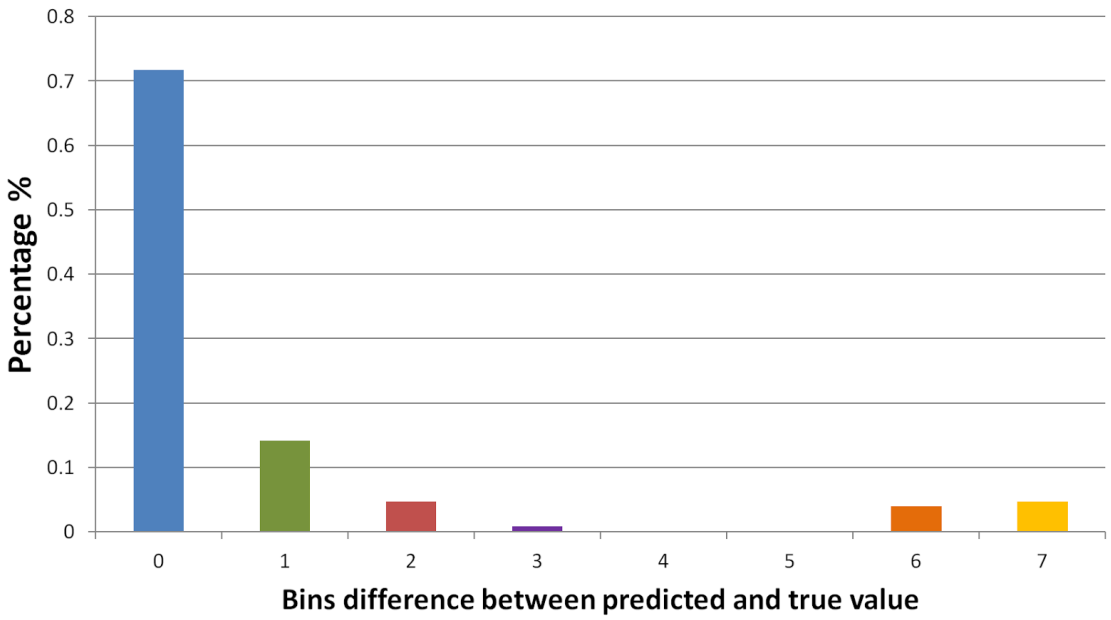

Figure 3.11. *Graph showing the percentage of times the classifier predicted a heading with a difference of N angular bins. N=0, represents the number of times the classifier output the true heading of the diver.*

During the CADDY Software integration Week in Zagreb, Croatia on May 18th, 2016; the algorithm was tested on never before seen data of one diver wearing the DiverNet, so that we could compare the output given by the inertial sensors and the classifier. The ground-truth we used was the measurement of the inertial sensor in the diver's chest. *The accuracy achieved was 83.4%*.

Also the limits of the sensor coverage (stereo camera) were tested, since enough points in the point cloud are needed to yield good results. Precise results could be obtained in the range of 1 to 3 meters, for this reason this method is categorized as local sensing. It is important to have in mind that this data will be used as an input to a filter, such as the one described in Section 3.1.1, to change the position and speed of the AUV smoothly. The control filter does not require values as precise as the ones given by the DiverNet at all times in order to guide the AUV in the correct way, as long as the output value is a close reference to the true value. Thus, for the BUDDY guide application that we are addressing, the trained classifier is enough; however, work to make the classifier more accurate and to make the heading quantization smaller will be done in the following months. This work will be further developed into a scientific publication as well.

As it is shown in Figure 3.11, besides the diver's heading, the bearing angle (angle between the camera's direct line of view and the diver's center of mass) and the range (distance between the camera and the diver's center of mass - magnitude of the blue arrow in Figure 3.11) are given as extra information for the control filter. A screenshot of the ROS message containing this information is given in Figure 3.12; the next step is to compute the variance of all these values from the labels probabilities the Random Forests inherently give.

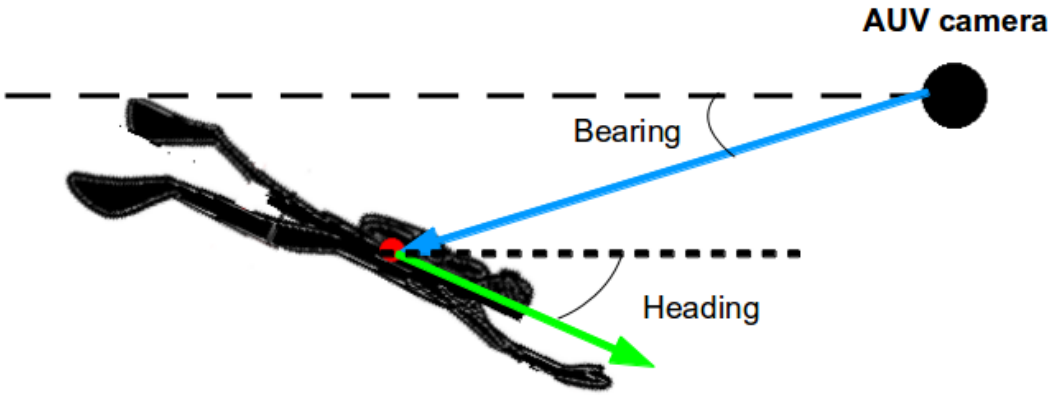

Figure 3.11. *Diagram of the diver's pose measurements done from the generated pointcloud, includes: bearing, heading and range (magnitude of blue arrow).*

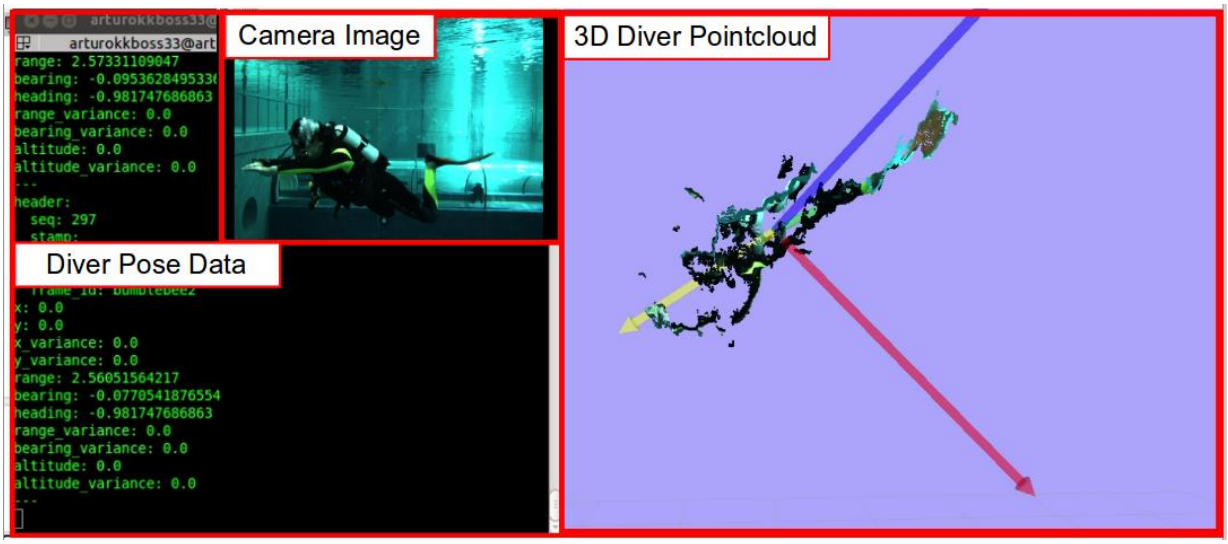

Figure 3.12. *Screenshot of ROS and Rviz showing the camera images, the diver's pointcloud (top view) and the terminal displaying all important measurements about the diver's pose.*

4 References

[1] Madgwick, Sebastian OH, Andrew JL Harrison, and Ravi Vaidyanathan. "Estimation of IMU and MARG orientation using a gradient descent algorithm." *Rehabilitation Robotics (ICORR), 2011 IEEE International Conference on*. IEEE, 2011.

[2] Jolliffe, I. T. "Principal Component Analysis". Springer-Verlag. 2002. doi:10.1007/b98835. ISBN 978-0-387-95442-4.

[3] A. Bosch, A. Zisserman and X. Munoz, "Image Classification using Random Forests and Ferns," *2007 IEEE 11th International Conference on Computer Vision*, Rio de Janeiro, 2007, pp. 1-8.

[4] P. Geurts, D. Ernst, and L. Wehenkel. Extremely randomized trees. Machine Learning, 36(1):3-42, 2006.