Estimating diver orientation from video using body markers

Ivor Rendulić, Aleksandar Bibulić, Nikola Mišković
University of Zagreb
Faculty of Electrical Engineering and Computing
Zagreb, Croatia
Email: ivor.rendulic@fer.hr
Email: aleksandar.bibulic@fer.hr
Email: nikola.miskovic@fer.hr

***Abstract*—Diving is a dangerous activity during which diver's safety is significantly compromised, both due to unpredictable environment and dependence on technical systems for life support. Introducing the concept of a cognitive autonomous underwater vehicle (AUV), that is capable of guiding, observing and assisting the diver at all times, requires precise positioning of the AUV relative to the diver. While diver orientation can be measured locally on the diver and transmitted to the underwater vehicle using acoustic communication channels, the problems of limited bandwidth and communication delay can influence the quality of AUV positioning. Another approach, which is considered in this paper, is to use remote sensing techniques, i.e. mono or stereo camera on board the AUV, to determine orientation of the diver based on markers placed on the divers shoulders. As an initial step, laboratory experiments were conducted where orientation of two physically coupled spherical markers is determined by using mono and stereo camera, and compared to the ground truth orientation obtained from an inertial measurement unit. The obtained results prove the concept in laboratory conditions, with the limitations imposed on the distance of markers relative to the camera.**

I. Introduction

Diving is an extremely dangerous activity not only because of the hazardous and unpredictable underwater environment in which divers operate, but also because of the fact that divers lives depend on technical equipment such as breathing regulators. One of the main objectives of CADDY project is to introduce an autonomous underwater vehicle that will act as cognitive robotic diving buddy, thus significantly reducing risks inherent to diving activities (see Fig. 1). This autonomous vehicle will act at the same time as a buddy "guide", leading the way for the diver; buddy "observer", looking out for the diver; and buddy "slave", helping the diver execute specific tasks. In order to achieve these functionalities, the vehicle has to perceive the diver (determine diver's position and orientation) in order to be able to position itself around the diver while at the same time keeping the diver within its field of view.

This work is supported by the European Commission under the FP7-ICT project "CADDY - Cognitive Autonomous Diving Buddy" Grant Agreement No. 611373.

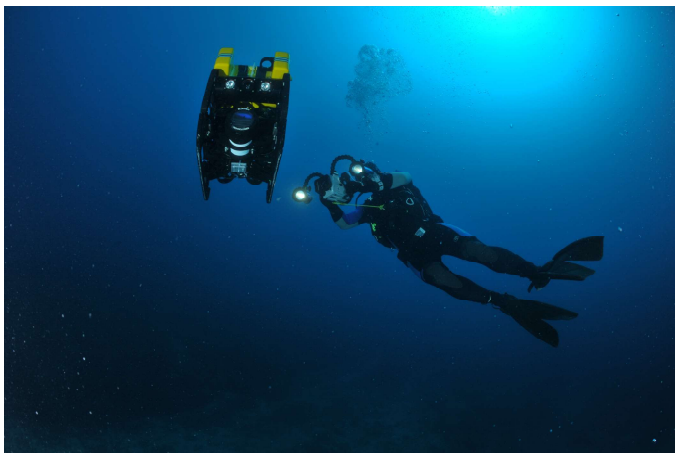


Fig. 1. A diver and a remotely operated underwater vehicle working together.

Underwater environment imposes many constraints on technical systems. Estimating orientation of a person, which can easily be done on dry land with inertial measurement sensors, becomes a non-trivial task. Sensor has to be placed in a waterproof casing and the data transfer at distances longer than a few meters has to be done either by acoustic communication or by wired link. Given the very high price of an acoustic communication system and many constraints caused by having a wire connected to a diver, it is reasonable to explore other remote sensing options.
In this paper we address the problem of estimating orientation and distance of a diver for the purpose of positioning an autonomous underwater vehicle (AUV) which serves as a cognitive diving buddy. As an alternative to having a sensor mounted on the diver and needing to transmit the data to our vehicle we are using video cameras to record the diver. Our vehicle is equipped with a stereo and a mono camera, both of which were used for the task.
To cope with often poor visibility issues, caused by water turbidity or low lighting, we are using highly visible body markers to mark predefined body parts of a diver. Altho-

ugh diver detection and distance estimation can be done in various conditions even without the markers, estimating the orientation is much more tricky if a distinct marker (e.g. mask or air tank) is not visible in the image. As a starting point in diver orientation estimation, this paper focuses on laboratory experiments on dry land. Even though this setup significantly simplifies the situation which is expected in the underwater environment, this step is crucial for determining the proof of concept. Separate approaches by both mono and stereo camera are taken to calculate the orientation and distance based on the detected spatial positions of the markers.
The paper is organized as follows. Sections II and IIi describe the principles of using mono and stereo camera for determining positions of visual markers, while Section IV describes how orientation is determined from the obtained data. Section V focuses on the experimental setup and reports results from laboratory experiments, while the paper is concluded with Section VI.

II. Mono Camera

The first approach that was used in estimating the orientation was by tracking the spatial position of two distinct visual markers with an ordinary mono camera.
It is clear that mono cameras have a major limitation in tasks that include estimating distances from objects. It is impossible to give a good estimate of depth from a single image and without any prior knowledge. In [1] authors use supervised learning to train a model for depth estimation in a single still image. Other approaches include estimating depth from video of a moving camera based on optical flow [2], [3].
For the problem of determining diver orientation, we are using a much simpler solution of estimating depth from the perceived size of a familiar object. Given a reference size of an object in pixels at a known distance, we are estimating the change of distance from the changes in size.
If an object is recorded at two different distances z_1 and $z_2 = sz_1$, the ratio of object areas in pixels in the two recordings P_{1pix}/P_{2pix} is equal to the ratio of areas A_2 and A_1 captured by the camera.

$$\begin{aligned} P_{1pix}/P_{2pix} &= \frac{P/A_1}{P/A_2} = \frac{A_2}{A_1} = \frac{a_2 b_2}{a_1 b_1} \\ &= \frac{k(2z_2 \operatorname{tg}(\alpha/2))^2}{k(2z_1 \operatorname{tg}(\alpha/2))^2} \\ &= \frac{z_2^2}{z_1^2} = \left(\frac{sz_1}{z_1}\right)^2 = s^2 \end{aligned}$$

In the formula above, P_{1pix} and P_{2pix} are the sizes of the recorded object in pixels when the object is recorded at distances z_1 and z_2 from the camera. A_1 and A_2 are areas of the field captured by the camera at distances z_1 and z_2. s is the ratio of distances z_2/z_1, α is the horizontal field of view angle and k is the height-to-width aspect ratio of the image $k = b/a$. All values are depicted on Fig. 2. Given that the ratio of object area sizes is inversely proportional to the square of the ratio of distances, square roots of object area sizes or any linear measure of objects (such as height, width or radius) are linearly inversely proportional to the ratio of distances. This enables us to estimate the unknown distance to the object by measuring it's size and knowing size at a reference distance.

$$z_{curr} = z_{ref}\sqrt{\frac{P_{ref,pix}}{P_{curr,pix}}}$$

Also, several measurements, such as area, width and height, can be combined to give a better estimate.
For the lateral movement from the camera two approaches are considered. Firstly, pixel size at a given distance can also be estimated from a pixel size at a reference distance and estimated depth of an object. However, it is reasonably to assume that the markers mounted on a human body (e.g. on the shoulders) maintain constant mutual distance l. This enables the calculation of relative positions of the markers, given the previously calculated depths z_1 and z_2.

$$\begin{aligned} t &= \sqrt{\frac{b_{pix}}{b}} \\ t &= \sqrt{\frac{(x_{2,pix} - x_{1,pix})^2 + (y_{2,pix} - y_{1,pix})^2}{l^2 - (z_2 - z_1)^2}} \\ x_2 &= (x_{2,pix} - x_{1,pix})/t \\ y_2 &= (y_{2,pix} - y_{1,pix})/t \end{aligned}$$

In the equations above, t is the ratio of distance measured in pixels and actual distance. It is calculated as a ratios of distance b, marked in 3. The numerator, b_{pix}, is the distance in pixels, and the denominator, actual distance b, is calculated from the previously calculated depth difference between the two markers. Based on the ratio t marker locations can be transformed from pixels to millimeters.

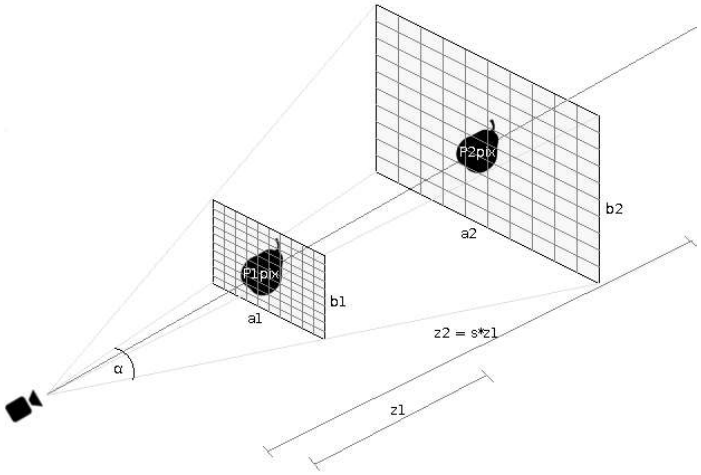


Fig. 2. Visualization of object recording at different distances.

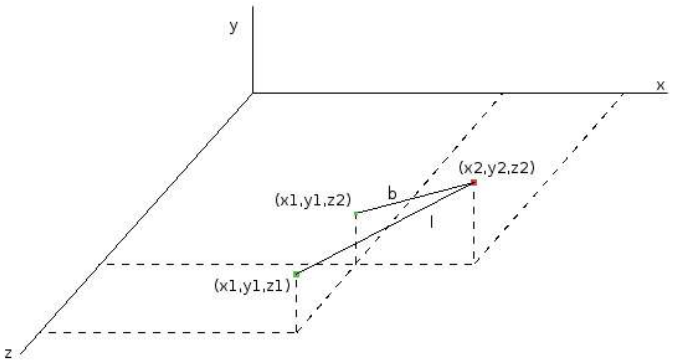


Fig. 3. Visualization of object recording at different distances.

III. Stereo Camera

In the second approach, orientation was estimated by using stereo camera. Stereo vision system is used to determine depth from two or more images taken at the same time from horizontally shifted cameras.
The calculation of depth includes the following steps of processing pairs of stereo images:

1) Calibration of a stereo camera system
2) Stereo image rectification
3) Search for correspondence points in stereo pair
4) Calculating distance from disparity

The process of calibration is to determine the estimation of the intrinsic parameters for each camera and estimation of the translation and rotation of the second camera relative to the first one. This is accomplished using a calibration plane with an asymmetric checkerboard. Multiple pairs of images of a calibration pattern is needed to calibrate the system. Detailed explanation of calibration process are omitted but can be found in in [4].
Estimated parameters are used to rectify a stereo pair of images. Rectification is a transformation process used to project two or more images onto a common image plane, which reduces the 2D stereo correspondence problem to a 1D problem (see e.g. [5]).
Now corresponding point in left image can be found in the same row in the right image. Points chosen are centers of the markers. Disparity is calculated as distance between corresponding pixels in the left and right image as

$$d = |x_{1pix} - x_{2pix}|$$

where x_{1pix} is x coordinate of the corresponding pixel in the left image and x_{2pix} is x coordinate of the corresponding pixel in the right image. Now we can tell that the geometry of the stereo camera system is known, meaning that x, y and z coordinates of the markers can be calculated as

$$z_p = T * \frac{f}{d}$$
$$x_p = x_{lpix} * \frac{f}{d}$$
$$y_p = y_{lpix} * \frac{f}{d}$$

where T is the distance between two cameras, f is the focal length, x_{lpix} and y_{lpix} are the coordinates of the coordinate system from the left camera.

IV. Orientation From Spatial Coordinates

After obtaining the spatial coordinates of both markers by either of the methods described in previous sections, orientation of the diver (e.g. assuming the markers are mounted on his shoulders) can be calculated by simple trigonometry. For representation of the orientation Euler angles yaw and roll are used.

$$\psi = \operatorname{arc\,tg}(\frac{z_2 - z_1}{x_2 - x_1})$$
$$\phi = \operatorname{arc\,tg}(\frac{y_2 - y_1}{\sqrt{(z_2 - z_1)^2 + (x_2 - x_1)^2}})$$

Here ψ is the yaw angle around the z-axis and ϕ is the roll angle around the x-axis. The position of the markers on a diver is shown in Fig. 4, along with the rotation axes. Calculation of all three angles cannot be done from only two markers and would require a third one.

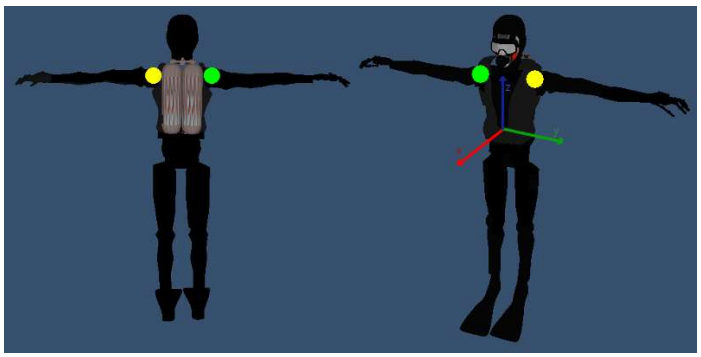


Fig. 4. Position of the markers on a diver and coordinate system of the diver for calculating orientation.

V. Experimental Setup and Results

To test the accuracy of the proposed systems an experimental setup consisting of two items (visual markers) mounted on opposite ends of a pole was used. An IMU sensor was used to provide reference orientation during the recording with a mono and a stereo camera. The setup is shown in Fig. 5.

Fig. 5. Experimental setup for visual orientation estimation.

A. IMU Reference Orientation

A Pololu MinIMU-9 was used to obtain reference orientation. It includes three 3-axis sensors, measuring acceleration, angular velocity and magnetic field. All sensors were sampled at 50Hz and time-stamped. The independently measured nine degrees of freedom enables the calculation of absolute orientation in space. Data from the three sensors is processed by a complementary filter described in [6] in order to get a precise estimate.

B. Visual Marker Detection

1) Mono Camera: Video acquisition was performed with an ordinary computer web camera in VGA resolution (640x480) for the mono orientation estimation experiment. Frames were captured at average of 15Hz and time-stamped to allow direct comparison to the IMU reference orientation.
Template- and color-based detection were used to obtain the positions (in pixels) of the x and y coordinates of the markers, and to give an estimate of the marker size in pixels.
Color-based detectors were used to simply distinguish pixels matching the marker colors in the image and find the areas that are most likely corresponding to the markers. This method allows for simple size estimation by counting the pixels that are matched to each marker.
For the template-based detection, OpenCV template matching library was used on a multiple scales of the marker image and the results were used to estimate the position and depth of the markers.

2) Stereo Camera: For the stereo video acquisition a monochromatic Point Grey Bumblebee XB3 was used. The camera has three sensors distanced at 12cm apart, allowing three separate stereo pairs to be formed for a single shot. Frames were taken at 5Hz and also time-stamped for comparison to the IMU.
The markers used in stereo vision test were spherical, so the detector used Hough Circle Transform [7].

Table I
Standard deviations of differences between orientations obtained by IMU and visual sensing.

	$< 1m$	$3m$
mono camera roll	$\sigma = 3.58°$	$\sigma = 10.54°$
mono camera yaw	$\sigma = 8.16°$	$\sigma = 24.14°$
stereo camera roll	$\sigma = 2.53°$	$\sigma = 7.67°$
stereo camera yaw	$\sigma = 4.01°$	$\sigma = 12.23°$

C. Results

In this section we present the results of the visual orientation estimation.

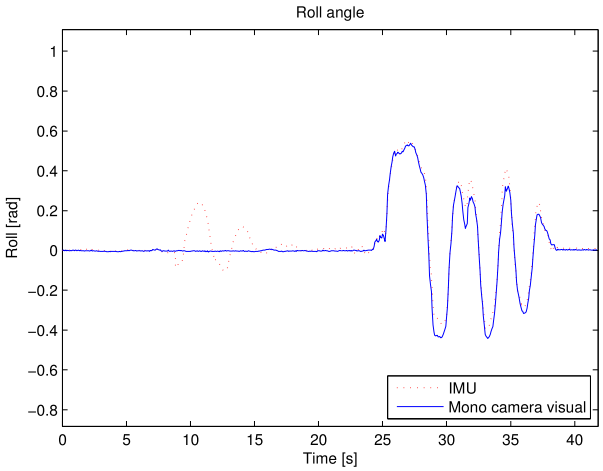
1) Mono Camera: Tests with the mono camera were conducted by tracking two almost-spherical markers whose diameter is approximately $10cm$. As expected, the depth estimation based on size change is heavily influenced by the distance to the object. The tests were conducted at two distances.
First, the setup was moved at distances between 0.5 and 1.0 meter from the camera. Results from those tests are visible in Fig. 6. The orientation obtained from the visual tracking is very accurate and closely follow the one obtained with the IMU. Then, the same markers were used at a larger distance from the camera. This impairs the ability of the marker detector to give a good estimate of the size, as well as limits the resolution of the detected marker size in pixels. The results shown in Fig. 7 show the orientation calculated when the setup was at slightly over 3.0 meters from the camera.
Standard deviation from the reference orientation obtained with IMU is shown in Table I.

2) Stereo Camera: Unlike the mono-based orientation estimation described above, the stereo based one does not suffer as much on longer distances from the accuracy of object size and position detection.
Tests were conducted with the same spherical markers as for mono camera. The distance between the markers and the camera was $1m$ in the first trial and $3m$ in the second one. Results of two trials are shown in Fig.8 and Fig.9.
Standard deviation from the reference IMU orientation is given in Table I.

VI. Conclusions and future work

This paper presented the results of laboratory experiments that were designed to determine two angles of orientation (yaw and roll) of physically coupled spherical markers by using mono and stereo camera. These markers will at a later stage be mounted on the divers shoulders, and will be used to determine diver's orientation relative to the autonomous underwater vehicle equipped with the cameras.

The initial results have shown that, at least in controlled laboratory conditions, both mono and stereo camera can be used to determine orientation. Since calculation of yaw and roll angle requires distance of the object to the camera, the experiments were carried out at two distances.



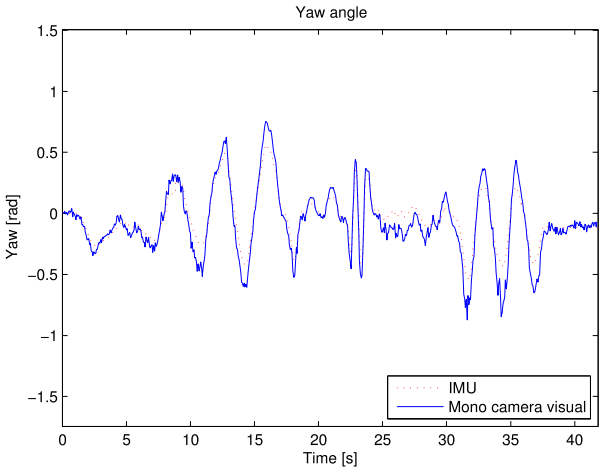
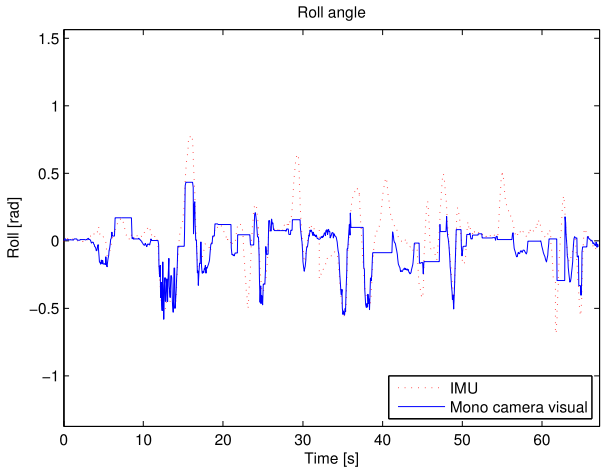


Fig. 6. First trial with mono camera, movement induces changes in both yaw and roll angles. Distance from the camera $< 1m$



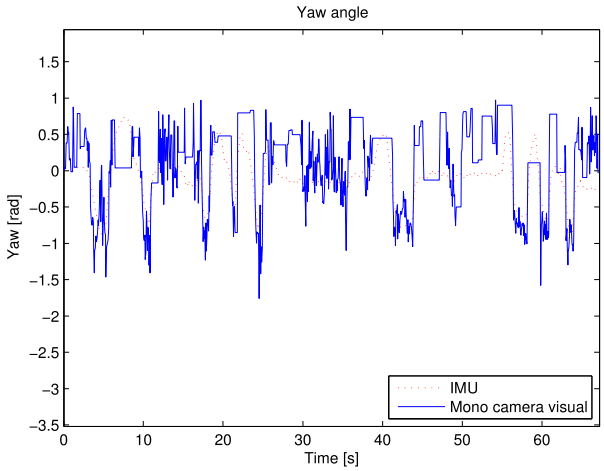


Fig. 7. Second trial with mono camera, movement induces changes in both yaw and roll angles. Distance from the camera $\approx 3m$

Ground–truth was set with measurements from an inertial measurement unit.

We have shown that results obtained by mono camera are very good at small distances, but are heavily deteriorated with increasing distance. On the other hand, measurements from the stereo–camera are less influenced by distance, mostly due to more precise distance measurements, keeping the standard deviation of measurements below $10°$.

Future work will include using the experimental setup underwater, in laboratory conditions. After that, the markers will be mounted on the diver and full field experiments will be conducted, in order to test the robustness of image processing in real conditions, with low visibility.

References

[1] A. Saxena, S. H. Chung, and A. Y. Ng, "3-d depth reconstruction from a single still image," *International journal of computer vision*, vol. 76, no. 1, pp. 53–69, 2008.

[2] B. Shahraray and M. K. Brown, "Robust depth estimation from optical flow," in *Computer Vision., Second International Conference on.* IEEE, 1988, pp. 641–650.

[3] W. Kruger, W. Enkelmann, and S. Rossle, "Real-time estimation and tracking of optical flow vectors for obstacle detection," in *Intelligent Vehicles' 95 Symposium., Proceedings of the.* IEEE, 1995, pp. 304–309.

[4] P. Hillman, "White paper: Camera calibration and stereo vision," *Lochrin Terrace, Edinburgh EH3 9QL, Tech. Rep*, 2005.

[5] C. Loop and Z. Zhang, "Computing rectifying homographies for stereo vision," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, vol. 1. IEEE, 1999.

[6] S. O. Madgwick, A. J. Harrison, and R. Vaidyanathan, "Estimation of imu and marg orientation using a gradient descent

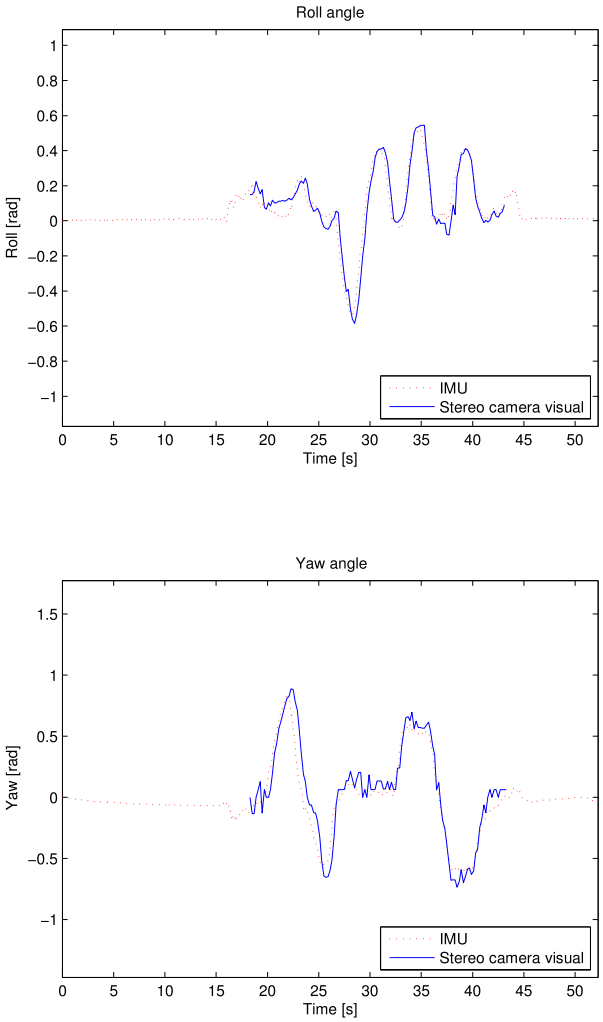


Fig. 8. First trial with stereo camera, movement induces changes in both yaw and roll angles. Distance from the camera $\approx 1m$

algorithm," in *Rehabilitation Robotics (ICORR), 2011 IEEE International Conference on.* IEEE, 2011, pp. 1–7.

[7] M. Smereka and I. Dulęba, "Circular object detection using a modified hough transform," *International Journal of Applied Mathematics and Computer Science*, vol. 18, no. 1, pp. 85–91, 2008.

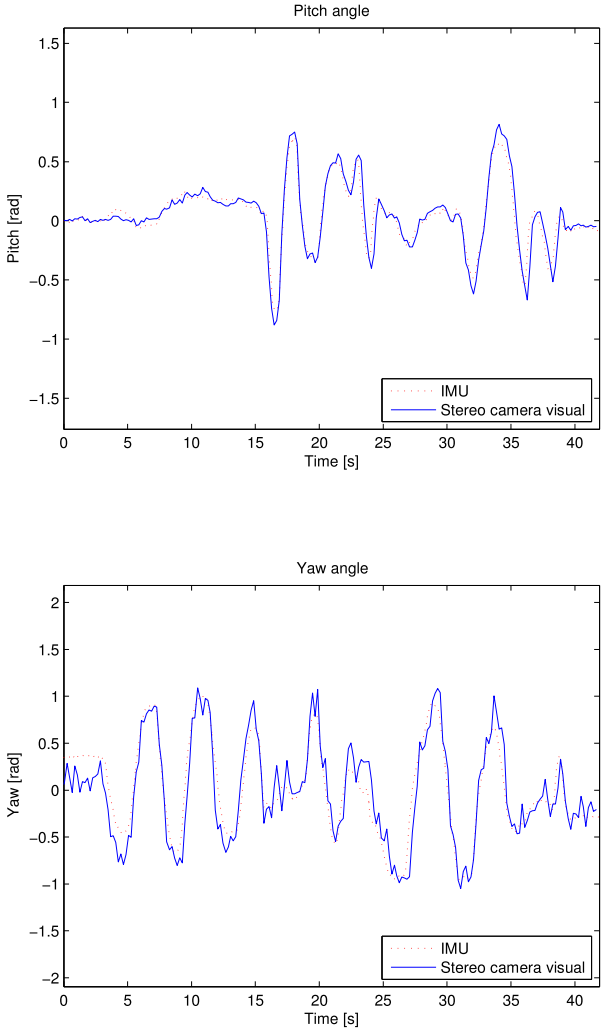


Fig. 9. Second trial with mono camera, movement induces changes in both yaw and roll angles. Distance from the camera $\approx 3m$