

Pose Estimation for Underwater Vehicles using Light Beacons ^{*}

Nuno Gracias ^{*} Josep Bosch ^{*} Mohammad Ehsanul Karim ^{**}

^{} Computer Vision and Robotics Institute, Centre d'Investigació en Robòtica
Submarina, Universitat de Girona, 17003 Girona, Spain
(ngracias@silver.udg.edu; j.bosch@udg.edu).
^{**} (e-mail: ehsan.mce@gmail.com)*

Abstract: This paper presents an approach for estimating the relative location and orientation between two or more underwater vehicles operating in tight formation. One of the vehicles is equipped with a camera of wide field of view. The other vehicle(s) are equipped with active light markers to enable the use of computer vision for pose estimation. The pose estimation addresses two scenarios, which are both important from the operational point of view. The first pertains to the availability of an acoustic communication channel which allows for exchanging attitude data and acoustic ranging, and use it in the pose estimation procedure. The second corresponds to the exclusive use of a set of 4 or more optical beacons with no acoustic information exchange, which is a capability that has not been yet proposed nor demonstrated in underwater vehicles. The contributions can be summarize as (1) a novel method of estimating the pose of an autonomous underwater vehicle using light beacons and other sensors when available, (2) an automated marker configuration analysis approach. Performance of the pose estimation approach is evaluated using synthetic and real data.

Keywords: Robot navigation, Robot vision, Underwater vehicles, Cooperative navigation, Relative localization, Active markers, Formation Control

1. INTRODUCTION

Present day oceanographic research relies heavily on the use of remotely operated vehicles (ROV) and autonomous underwater vehicles (AUV), specially in deep-water operation. AUVs are becoming standard tools in applications as varied as environmental surveying, geology, archaeology, cable inspection, and several others relating to industry and the military. However, the existing technology is still immature for close-range surveying of rugged terrains, such as caves, narrow passages and overhangs, due to limitations on the terrain sensing and on the navigation accuracy.

The deployment of multiple AUVs in close formation has the potential to significantly expand the coverage swath in mapping missions that require close proximity to the seafloor, such optical or electromagnetic surveying. In areas of high topography, rigid arrays of sensors cannot be used safely, whereas AUV formations can provide the required degree of terrain compliance. However AUV formations require the ability to acquire precise estimates of the vehicle's relative position, specially when the vehicles are less than 10 meters apart. The existing techniques for relative position estimation rely solely on acoustic ranging, and are not accurate nor provide fast enough updated to ensure safety under such close range operation.

Under adequate visibility conditions, optical cameras can be very effective for computing precise position estimates, including full inter-vehicle poses. The effects of absorption and scattering often preclude the use of standard feature detectors as a solution to the problem of vision based formation sensing. To improve the chances of detecting point features and to identify individual vehicles, this paper proposes to endow the AUVs with active light markers, blinking with distinctive patterns to facilitate their recognition.

This paper focuses on the aspects of pose estimation and light beacon placement. The pose estimation addresses two scenarios, which are important from the operational point of view. The first pertains to the availability of an acoustic communication channel which allows for exchanging attitude sensor data and acoustic ranging, and use it in the pose estimation procedure. The second corresponds to the exclusive use of a set of 4 or more optical beacons with no acoustic information exchange, which is a capability that has not been yet proposed nor demonstrated in underwater vehicles. This concept is illustrated in Figure 1.

A related aspect, which is not addressed in this paper, is the detection, identification and tracking of the individual beacons in the image. This component is currently under development and will be addressed in a forthcoming paper.

The present work has been developed within the framework of the MORPH EU-FP7 project (2012-2015) described in Kalwa et al. (2012). This project proposes a novel concept of an underwater robotic system that emerges out of using different mobile robot-modules with distinct and complementary resources. These mobile robots navigate at a very close range as a group,

^{*} This research was supported by the EU FP7-Project MORPH (FP7-ICT-2011-7-288704), the Spanish National Project OMNIUS (CTM2013-46718-R), the Generalitat de Catalunya through the ACCIO/TecnoSpring program (TECSPR14-1-0050), "la Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya" and the University of Girona under a grant for the formation of researchers.

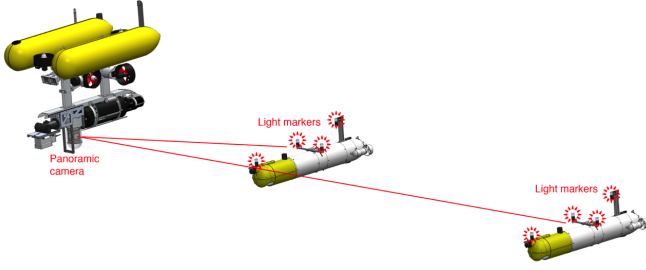


Fig. 1. Concept of the pose estimation using light beacons and a wide field-of-view camera.

and have the ability to adapt the formation to changes in the terrain. The most relevant concept with respect to this paper is that an underwater vehicle equipped with a multibeam sonar profiler, advances at the forefront of the formation, flying at a “safe” altitude from the sea-floor, while two other vehicles fly behind, very close to the bottom, acquiring images. As it can be deduced, an accurate knowledge of the poses of all robots during the missions is fundamental. The relative localization between vehicles is done through acoustic ranging in the case of medium to long distances. However for the case of short distances, where acoustics can not provide updates with enough precision and frequency to ensure safety, the vision-based method proposed in this paper comes as a natural solution. In order to have a sensor with the widest possible field of view, an omnidirectional underwater camera is used (Bosch et al., 2015). The camera was integrated in the Girona500 AUV (Ribas et al., 2012) which takes the role of the leader vehicle.

2. SELECTED RELATED WORK

The use of easily identifiable light sources for pose estimation has gained momentum in recent years in applications involving micro aerial robots. Recent examples are the work of Censi et al. (2013) and Faessler et al. (2014) where favorable visibility conditions allow the use of fast cameras and infrared LEDs to provide very fast pose updates. However, in underwater applications, where the detection and identification of the light sources is far more challenging, such capability has not been demonstrated yet.

In the context of underwater docking, Krupinski et al. (2009) designed an object pose estimation algorithm by fusing vision, a Doppler Velocity Logger (DVL), and a Fiber Optics Gyroscope (FOG) in an Extended Kalman Filter (EKF), using also passive and active optical markers. The passive marker based system relied on extracting SIFT (Scale-invariant feature transform) point features on the docking system. Conversely, the active marker system used LEDs on the docking station blinking at a fixed frequency. Although efficient in autonomous docking, the proposed approach for detection and recognition is not suitable in the context of formation sensing for multiple AUVs. The main reason is the use the SIFT descriptors for point features, which lead to poor performance under turbidity (Garcia and Gracias (2011)) and requires adequate illumination of the vehicles, which may not be feasible in deep underwater settings.

Lu et al. (2000) designed a fast and globally convergent pose estimation algorithm titled Orthogonal Iteration (OI) or LHM by reformulating the definition of pose estimation, as that of minimizing the *object-space* collinearity error. The algorithm is

flexible because the rotation matrix R can be initiated manually. Lepetit et al. (2009) designed a faster non-iterative solution to the problem where the n 3D points are expressed as a weighted sum of four virtual control points, thereby reducing the problem to estimating the coordinates of these control points. Moreover, experiments suggest that the algorithm performs similar to LHM or OI. However it is faster and more stable. Li et al. (2012) recently proposed a non-iterative solution to the problem by solving a seventh order polynomial system where the reference points are divided into 3-point subsets which results in a series of forth order polynomials. The algorithm is efficient, stable, faster and can deal with planar case and non-planar case.

3. BEACON-BASED POSE ESTIMATION

The following notation is used in this paper.

- n = Number of observed light beacons
- ρ = Acoustic range measurement
- ϕ = Roll angle
- θ = Pitch angle
- ψ = Yaw angle
- s = Follower AUV reference frame
- g = Earth or geo-reference frame
- l = Leader AUV reference frame
- c = Camera reference frame
- R_a^b = Rotation matrix of frame a with respect to b
- \mathbf{t}_a^b = vector of 3D translation of frame a with respect to b
- K = Intrinsic camera parameter matrix
- $\Theta_a^b = [\mathbf{t}_a^b \ \phi \ \theta \ \psi]$ vector of 6DOF pose parameters

We consider two AUVs equipped with Fiber Optic Gyroscopes (FOG) which may communicate using acoustic modems. The leader vehicle is equipped with a panoramic camera. The camera in the leader vehicle acquires an image of the follower AUV and establishes n correspondences $[\mathbf{X}_1^s, \mathbf{X}_2^s, \dots, \mathbf{X}_n^s] \mapsto [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ on the vehicle. Concurrently, the leader measures the acoustic range ρ_s^l , acquires its own orientation information $\phi_l^g, \theta_l^g, \psi_l^g$ from the on-board FOG, and also receives acoustic message containing the follower AUV’s orientation information $\phi_s^g, \theta_s^g, \psi_s^g$. Their respective rotation matrices are R_l^g and R_s^g .

3.1 Initialization

The pose estimation method is based on the minimization of a cost function comprising error terms from light beacon reprojection, plus range and attitude measurements when available. Depending on the number of beacons observed, the minimization is initialized using one of the following two methods.

Multi-sensor Initialization This initialization uses the attitude sensors of both vehicles and the inter-vehicle range measurement ρ_s^l . The 3D location \mathbf{X}^c of a single light beacon in the camera frame is approximated by

$$\mathbf{X}^c = \rho_s^l K^{-1} \mathbf{x} \quad (1)$$

from where an estimate of \mathbf{t}_s^c can be obtained as

$$\mathbf{t}_{sini}^c = \mathbf{X}^c - R_s^c \mathbf{X}^s \quad (2)$$

The matrix R_s^c is obtained by simple rotation composition as

$$R_s^c = (R_l^g \cdot R_c^l)^T \cdot R_s^g \quad (3)$$

where R_l^g and R_s^g are rotations matrices obtained directly from the attitude sensors of the vehicles, and R_c^l is a known rotation

matrix relating the camera frame and the leader AUV frame. The initial pose of the follower AUV is,

$$\Theta_{sini}^c = [\mathbf{t}_{sini}^c \ \phi_s^c \ \theta_s^c \ \psi_s^c]; \quad (4)$$

This initialization is used when $n < 4$. When more than one light beacon is available, then a single beacon is selected.

Optical-only Initialization LHM can provide a pose estimate when $n \geq 3$. However, to avoid the issues arising from ambiguous solutions under noisy data, it is only used to provide an initialization when $n \geq 4$ and no other sensor data is available.

3.2 The Non-linear Cost Function

The non-linear cost function comprises the following three terms.

Light beacon image reprojection error Given a pose parameter vector Θ_s^c , the image projection of a light beacon can be predicted using a perspective pinhole camera model

$$\mathbf{x}_{pred} = K[R_s^c \ | \ \mathbf{t}_s^c] \mathbf{X}^s \quad (5)$$

where \mathbf{X}^s is the 3D location of the beacon (described in the follower's reference frame), \mathbf{x}_{pred} is the predicted image projection, and R_s^c and \mathbf{t}_s^c are obtained from Θ_s^c . The uncertainty on the observed points \mathbf{x}_{obs} is modelled as additive zero-mean Gaussian noise with covariance Σ_{im} . Using the Cholesky decomposition $\Sigma_{im} = LL^T$, the weighted error term is:

$$e_{im} = L^{-1}(\mathbf{x}_{pred} - \mathbf{x}_{obs}) \quad (6)$$

Range error Assuming a standard deviation σ_ρ for the range sensor, the weighted range error is

$$\rho_{pred} = \sqrt{\mathbf{t}_{x_s}^c{}^2 + \mathbf{t}_{y_s}^c{}^2 + \mathbf{t}_{z_s}^c{}^2} \quad (7)$$

$$e_\rho = \frac{\rho_{pred} - \rho_s^l}{\sigma_\rho}$$

Attitude error The weighted orientation error is computed similarly as e_{im} , where $\Sigma_{fog} = LL^T$ is covariance matrix of the attitude sensor.

$$e_{fog} = L^{-1}([\phi_s^c \ \theta_s^c \ \psi_s^c]^T - [\phi_s^c \ \theta_s^c \ \psi_s^c]^T) \quad (8)$$

The cost function is the sum of the squared weighted residues:

$$cost = e_{im}^T e_{im} + e_\rho^2 + e_{fog}^T e_{fog} \quad (9)$$

and is minimized using the Levenberg-Marquardt algorithm (Moré (1978)).

4. AUTOMATED MARKER CONFIGURATION ANALYSIS

The correct placement of the active markers on the follower AUV plays an important role in the estimation of its pose. A given marker configuration can be adequate for a certain range of poses, but might lead to large inaccuracies for some other range. This section details a conceptually simple method which, given a certain marker configuration, analyzes the placement of the markers and studies their impact on the accuracy and effectiveness in optical formation sensing. It is also used to automatically optimize the configuration of markers in a vehicle through an iterative process.

A selected marker configuration is analyzed considering different AUV poses. A simple Monte Carlo Simulation (MCS),

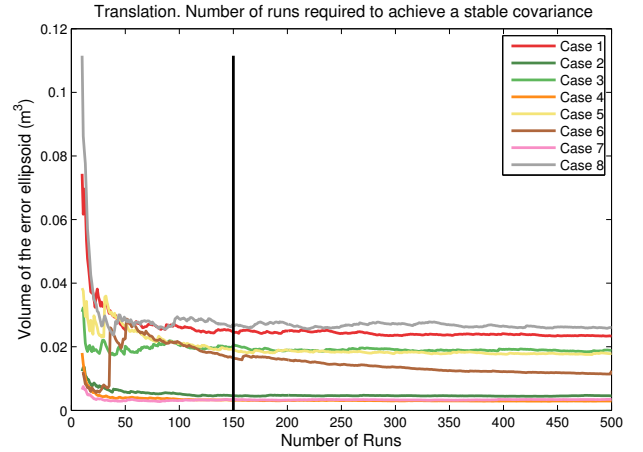


Fig. 2. Volumes of the error ellipsoids for a set of 8 poses, as a function of the number of MCS iterations.

is used to obtain an uncertainty estimate for each pose. The procedure is the following:

- (1) These different poses are chosen randomly taking into account the typical distances and orientations of the AUV during a MORPH mission. The AUV orientation will typically have its $[Roll, Pitch, Yaw]$ angles varying inside $[\pm 10^\circ, \pm 10^\circ, \pm 90^\circ]$, and the position $[X, Y, Z]$ varying inside $[\pm 5, \pm 5, (6 - 11)]m$.
- (2) For each pose the MCS is performed. A new instance of image noise was generated for each MCS iteration, considering the same image noise model as defined by Σ_{im} and only with visual information.
- (3) The covariance matrices of the pose parameters, and their associated error ellipsoid volumes, are computed using 150 runs. The volume of the largest ellipsoid was used to rank the marker configuration.

Figure 2 shows the stability of the pose estimation algorithm for the selected marker configuration, considering 8 different poses. This graph is useful to experimentally determine the number of MCS runs that are required to have a meaningful statistic for the error covariance matrix. From the graph it can be seen that more than 150 runs does not bring additional benefit. The reason for choosing the volume of the position error ellipsoid to evaluate the quality of the beacon configuration comes from the fact that errors in orientations are not as critical as errors in position, from the formation safety standpoint.

An simple algorithm was implemented to find the best configuration of beacons. We assume that a CAD model of the the follower AUV is available, which defines the positions where the light beacons can be installed on the surface of the vehicle. The idea is as follows:

- (1) The user defines the number of markers k to be placed on the object, and n configurations to be analysed.
- (2) A set of n configurations of k markers are randomly selected from the surface of the AUV.
- (3) A score is given to each configuration, according to the analysis method detailed previously. The configuration with the highest score is selected.

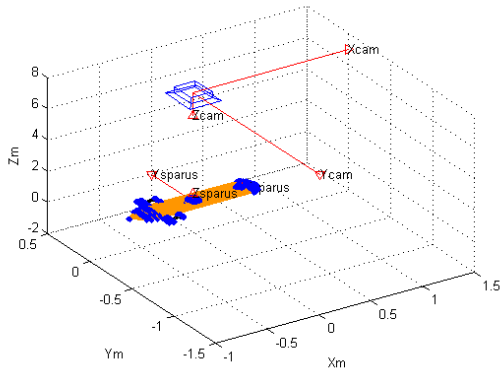


Fig. 3. Setup for the marker placement algorithm. Blue points represent the possible location of the markers.

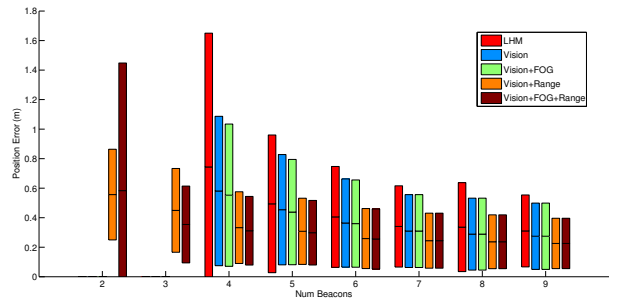
5. POSE ESTIMATION RESULTS

A comparison of the sensor fusion and vision-based object pose estimation algorithms was performed using a Monte Carlo simulation. The method of analysis was adopted from Li et al. (2012). Given a virtual perspective camera with image size 1616×1232 pixels, focal length of 860 pixels and zero skew, a set of 3D points were randomly generated in the camera reference frame. The 3D points were uniformly distributed in a 3D volume defined by the limits: $[1, 0] \times [-0.25, 0.25] \times [-0.25, 0.25]$ (in meters). This volume represents approximately the space where the beacons could be located in the Sparus II AUV. In every iteration the relative pose of the virtual AUV was randomly generated in a location visible from the virtual camera taking into account the most typical poses during real missions. Three independent simulations were performed and some of their primary statistics - mean, and standard deviation are highlighted. The statistics were based on 500 test runs which were randomly generated based on the simulation type. The following simulations were conducted:

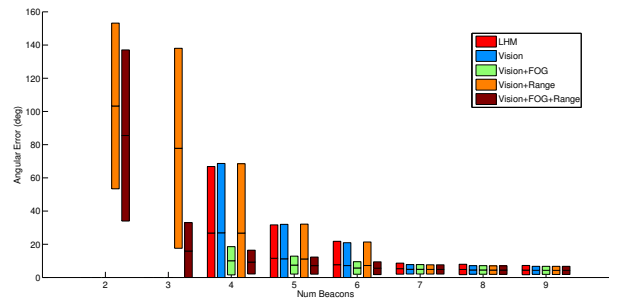
- Varying the number of observed beacons, with image noise of 2 pixels and vehicle distance roughly 9 meters (Figure 4).
- Varying the image noise level, considering 4 observed beacons and vehicle distance roughly 9 meters (Figure 5).
- Varying the vehicle's distance to the camera coordinate frame, considering 4 observed beacons and image noise of 2 pixels standard deviation (Figure 6).

In these simulations the standard deviation of the range sensor noise is $\sigma_p = 30cm$ whereas the covariance of the FOG (attitude sensor) noise is assumed to be $\Sigma_{fog} = \text{diag}(25, 25, 25) \text{ deg}^2$. The simulations indicate that the set Vision+FOG+Range provides the most accurate pose estimates. This is due to the translation and orientation being constrained by the range data and the orientation respectively. This becomes more evident from the results of Vision+FOG and Vision+Range. It can be seen in figure 4 that the position error of Vision+Range is similar to Vision+FOG+Range, whereas the orientation is significantly noisier. Similarly, the orientation error of Vision+FOG is comparable to that of Vision+FOG+Range. Figure 4 also indicates that both the position and orientation errors reduce with the increase of the number of beacons observations.

Figure 5 shows that the uncertainty of the pose parameters grows approximately linearly with the level of noise on the



(a) Mean and standard deviation of position error (in m)



(b) Mean and standard deviation of angular error (in deg)

Fig. 4. Performance of the pose estimation with varying number of observed beacons considering $\sigma_{im} = 2$ pixel (image noise), $\sigma_{fog} = 5^\circ$, $\sigma_p = 0.5m$ (range noise) at a distance of approximately 9m with 500 iterations.

beacon image projections. An exception is the case of position estimate using acoustic ranging, where the estimates are relatively unaffected. Figure 6 shows, as would be expected, that the position uncertainty grows with the distance of the AUV and that the orientation estimate are highly unreliable except when using the attitude sensors. Regarding computation effort, the pose estimation requires less than 40ms when using 4 beacons in Vision+FOG+Range.

The algorithm has been tested with preliminary data from sea trials of the MORPH project in the Azores archipelago. The Girona500 AUV was at the front of a formation that included a Sparus II AUV and the Seacat AUV, both equipped with light markers. During the mission, the estimated distance between their modems along with the navigation of all the vehicles and images from the omnidirectional camera captured from the Girona500 were recorded for further analysis. Figure 7(a) shows a real undistorted image from the omnidirectional camera used to test the algorithm, while figure 7(b) shows the reconstructed pose of the follower vehicle computed only taking account the 4 markers position.

Table 1 shows a comparison of the relative orientation and translation of the follower AUV with respect to the leader AUV. The data available from the vehicles were their heading and the distance between their modems. Even though they are valuable in order to compare them with the estimated results, they cannot be trusted as ground truth, due to the inaccuracy of the FOG used and the fact the ranges between the vehicles were received at a very low rate of a sample per second. The results between the different methods analyzed are coherent between them, as well as the graphical reconstruction of the pose.

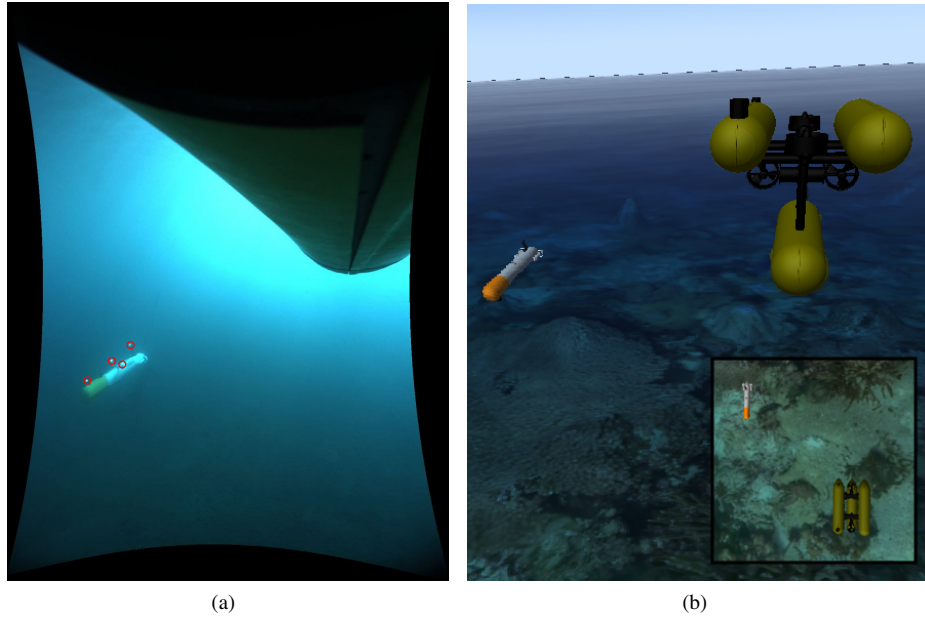


Fig. 7. Pose estimation example from sea trial data. Input image (with four beacons marked by the red circles) (a), and a 3D view of the AUV formation, created from the estimated relative pose (b).

Table 1. Comparison between the orientation and translation between the AUV frames computed with the different methods proposed for one sample image (Fig. 7).

Parameter	Rotation	Translation	Range
Vehicles sensors	[-0.0003,-0.1735, -0.166]	Not available	5.26
3 Lights + Range	[0.0907, -0.2179, -0.0891]	[-2.5289, 3.8772, 2.3316]	5.18
4 Lights (LHM)	[-0.0199, -0.1598, -0.0594]	[-2.2943, 3.5756, 2.2374]	4.81
4 Lights (Refinement)	[0.0424, -0.1919, -0.0746]	[-2.4121, 3.7300, 2.2862]	5.0
4 Lights + Range	[0.0841, -0.2143, -0.0794]	[-2.5205, 3.8672, 2.3291]	5.17

6. RESULTS OF THE AUTOMATED MARKER PLACEMENT

The automated marker placement approach was tested with a run of 500 different marker configuration of 4 beacons. The algorithm ranked each one of these configurations according to the largest error ellipsoid, computed for 8 different poses in a MCS of 500 runs. The histogram of the results for the 500 configurations is shown in figure 8. The best configuration is illustrated in figure 9. The final volume of the error ellipsoid is $0.0231 m^3$. Simulation is performed considering an image noise covariance matrix of $\Sigma_{im} = diag(3, 3) pixel^2$. The case below is for the best configuration found and its worst pose case studied. The theoretical (ground truth) pose (in meters and radians) of the follower AUV is,

$$[-1.552 \ 3.4329 \ 9.7106 \ -1.4760 \ 0.8848 \ 1.563]$$

The mean pose of the MCS (in meters and radians) of the follower AUV is,

$$[-1.5574 \ 3.4698 \ 9.8192 \ -1.6715 \ 0.4556 \ 1.15683]$$

The small volume of the ellipsoid error illustrates the reliability and adequacy of the marker configuration.

7. CONCLUDING REMARKS AND FUTURE WORK

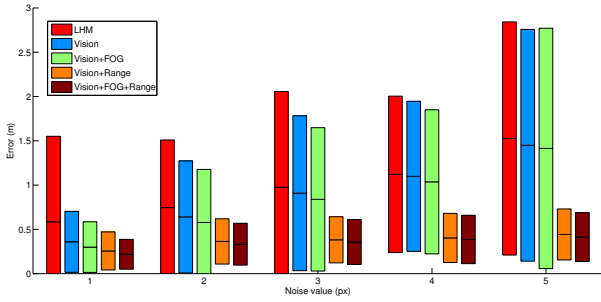
This paper presented a vision-aided pose estimation approach for AUVs, based on multi-sensor fusion comprising vision, attitude and range sensors. An important aspect of the approach is the possibility of performing the pose estimation with just light markers and no acoustic communication.

Given a possible marker configuration, either defined manually or obtained from an the CAD model of an AUV, the paper also presents a conceptually simple method to analyze the placement of the active light markers on the vehicle, and study their impact on the accuracy and effectiveness in optical formation sensing.

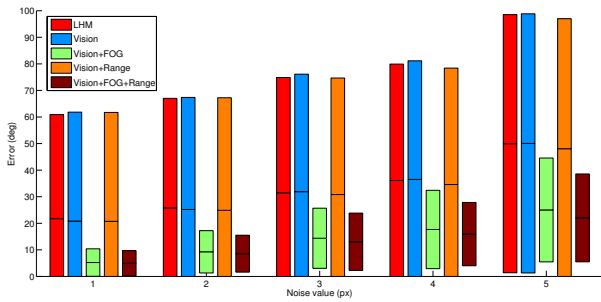
Current work is addressing the automatic detection and identification of the light markers through the use of signature temporal patterns. Such patterns are designed to allow the tracking and identification under low frame rates, and possibly large inter-frame motions.

REFERENCES

- Bosch, J., Gracias, N., Ridao, P., and Ribas, D. (2015). Omnidirectional underwater camera design and calibration. *Sensors*, 15(3), 6033–6065. doi:10.3390/s150306033. URL <http://www.mdpi.com/1424-8220/15/3/6033>.
- Censi, A., Strubel, J., Brandli, C., Delbruck, T., and Scaramuzza, D. (2013). Low-latency localization by active led markers tracking using a dynamic vision sensor. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 891–898. Tokyo, Japan.
- Faessler, M., Mueggler, E., Schwabe, K., and Scaramuzza, D. (2014). A monocular pose estimation system based on infrared leds. In *IEEE International Conference on Robotics and Automation (ICRA)*.
- Garcia, R. and Gracias, N. (2011). Detection of interest points in turbid underwater images. In *OCEANS, 2011 IEEE-Spain*. IEEE.

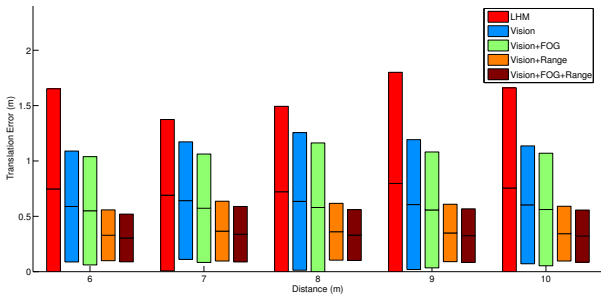


(a) Mean and standard deviation of position error (in m)

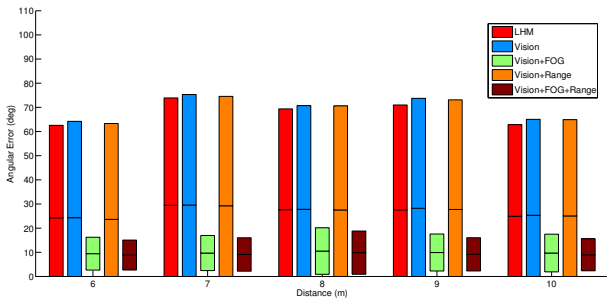


(b) Mean and standard deviation of angular error (in deg)

Fig. 5. Performance of pose estimation with varying image noise (*in pixels*), considering 4 observed beacons, $\sigma_{fog} = 5^\circ$, $\sigma_p = 0.5m$ (range noise) at a distance of approximately 9m with 500 iterations.



(a) Mean and standard deviation of position error (in m)



(b) Mean and standard deviation of angular error (in degrees)

Fig. 6. Performance of sensor fusion algorithms by varying the AUV's distance from the camera (*in meters*) considering 4 observed beacons, $\sigma_{im} = 2$ pixel (image noise), $\sigma_{fog} = 5^\circ$, $\sigma_p = 0.5m$ (range noise) with 500 iterations.

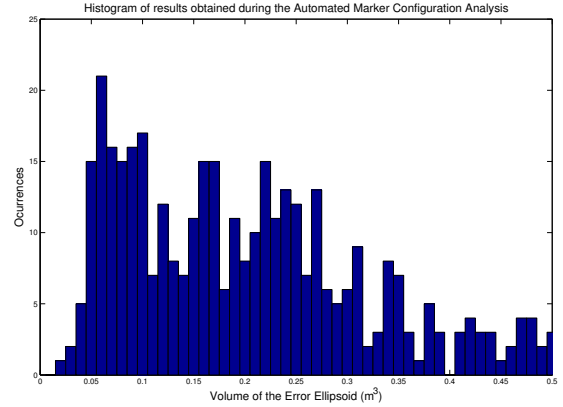


Fig. 8. Histogram of the scores for all the configurations analysed.

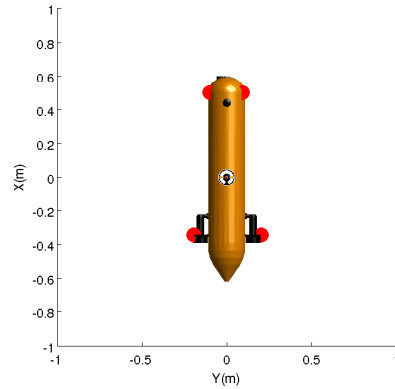


Fig. 9. Location of the best four-marker configuration found automatically (red blobs).

Kalwa, J., Pascoal, A., Ridao, P., Birk, A., Eichhorn, M., and Brignone, L. (2012). The european R&D-project MORPH: Marine robotic systems of self-organizing, logically linked physical nodes. In *IFAC Workshop on Navigation, Guidance and Control of Underwater Vehicles (NGCUV)*.

Krupinski, S., Maurelli, F., Mallios, A., Sotiropoulos, P., and Palmer, T. (2009). Towards AUV docking on sub-sea structures. In *OCEANS 2009 - EUROPE*.

Lepetit, V., Moreno-Noguer, F., and Fua, P. (2009). Epnp: An accurate o(n) solution to the pnp problem. *Int. J. Comput. Vision*, 81(2), 155–166.

Li, S., Xu, C., and Xie, M. (2012). A robust o(n) solution to the perspective-n-point problem. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(7), 1444–1450.

Lu, C.P., Hager, G., and Mjolsness, E. (2000). Fast and globally convergent pose estimation from video images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(6), 610–622.

Moré, J.J. (1978). The Levenberg-Marquardt algorithm: implementation and theory. In *Numerical analysis*, 105–116. Springer.

Ribas, D., Palomeras, N., Ridao, P., Carreras, M., and Mallios, A. (2012). Girona 500 AUV: From Survey to Intervention. *IEEE/ASME Transactions on Mechatronics*, 17(1), 46–53.