

Robust Outliers Detection and Classification for USBL Underwater Positioning Systems*

Marco Morgado¹, Paulo Oliveira^{1,2}, and Carlos Silvestre^{1,3}

¹ Institute for Systems and Robotics, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1, 1049-001, Lisbon, Portugal
{marcomorgado, pjcro, cjs}@isr.ist.utl.pt

² Department of Mechanical Engineering, Instituto Superior Técnico

³ Faculty of Science and Technology, University of Macau, Macau

Abstract. This paper presents a data classification algorithm able to detect corrupted measurements as outliers, with application to underwater ultra-short baseline (USBL) acoustic positioning systems. The devised framework is based on causal median filters that are readily implementable, and a set of theoretical analysis tools that allows for the design of the filter parameters is also presented. The design takes into account very specific implementation details of USBL acoustic positioning systems and also inherent non-ideal characteristics that include long period data outages. The outlier classifier is evaluated both in simulation and with experimental data from a prototype USBL acoustic positioning system fully developed in-house.

1 Introduction

Measurement outliers are naturally present in the output of every sensor package available on the market nowadays. The correct identification of these spurious perturbations often stands out as one of the important steps to the correct usage and successful integration of the aforementioned sensor packages into larger systems that are the building blocks of any robotic platform. This paper addresses the design and experimental validation of an outlier detector and classifier for underwater positioning systems.

The fast deployment, less complex hardware of small and compact arrays of receivers and increasing performance of modern factory-calibrated USBL positioning devices makes it suitable for faster intervention missions Napolitano et al. (2005). In generic operating conditions, a conductivity, temperature and depth (CTD) profile is normally required to account for the underwater sound velocity variations. Inverted USBL Vickery (1998) configurations, besides paving the way to future fully autonomous systems without the need to have surface mission support vessels, allows for the sound velocity to be considered constant while operating in the same underwater layer as the transponders (for instance, bottom operation while interrogating bottom placed transponders). The inverted USBL configuration is illustrated in Fig. 1.

* This work was supported by project FCT [PEst-OE/EEI/LA0009/2013], by project FCT PTDC/EEA-CRO/111197/2009 - MAST/AM, and by the EU Project TRIDENT (Contract No. 248497).

Due to several undesired aspects of the underwater sound propagation channel such as acoustic reverberation, layered underwater sound speed profiles, and mostly due to multipath phenomena, these type of acoustic positioning systems are highly susceptible to measurement outliers which need to be correctly identified. Otherwise these position measurement outliers can have a severe impact on systems that use them, degrading their performance downstream, as illustrated in Fig. 2, and worst case leading control and navigation systems to instability. This paper addresses the design of an outlier detector and classifier for a USBL positioning system and validates this classifier using experimental data obtained at sea with a USBL prototype fully developed in-house Morgado et al. (2010).

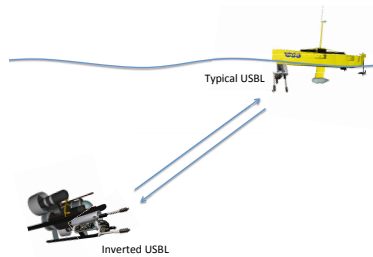


Fig. 1. Inverted USBL configuration as opposed to a typical installation floating on the sea surface - the prototype system developed in-house is also illustrated in the schematic attached to an underwater robotic in the inverted configuration and to the bow of an autonomous surface craft in the typical USBL configuration

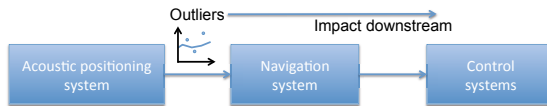


Fig. 2. Outliers from the acoustic positioning systems might have a severe degradation of the navigation system accuracy if not properly identified. In the worst case scenario it might even lead to instability of control systems that use this information downstream.

1.1 Paper Organization

The paper is organized as follows: Section 2.1 provides a review on the concepts of causal median filters that the outlier detector and classifier builds upon. Section 2.2 details the application of the median-based causal outlier detector and classifier to the USBL case. Some simulation results are analysed in Section 3 and Section 4 validates the usage of the devised classifier with real experimental data obtained at sea with the USBL prototype. Finally, Section 5 provides some concluding remarks and comments on future work to be developed within this subject.

2 Online Outlier Detection Algorithms

The detection and identification of possible outliers in the acoustic positioning measurements is of the utmost importance as already pointed out due to the fact that, if not correctly flagged, these spurious outliers might severely degrade the navigation systems performance that use this information, which information can also be critical to vital control systems on-board the underwater robotic vehicle. Albeit other more integrated solutions could be devised, that include designing navigation Kalman filters robust to outliers (see Ting et al. (2007a) and Gandhi and Mili (2010)), the idea of using instead a standalone outlier detector and classifier, that is coupled to the output of the acoustic positioning device, stems from the fact that not all navigation algorithms fit the framework of robust Kalman filters as presented in Ting et al. (2007a) and in Gandhi and Mili (2010). It is often desirable to have an outlier classifier detached from the dynamic filtering framework, thus allowing for several algorithms to be implemented independent of the outlier detection stage. Moreover, this setup allows for the USBL to provide position measurements with outliers correctly classified to a multitude of systems on-board.

The causal median on-line outlier classifier adopted in this work, is presented in this section and is based on the work presented in Menold P.H. and Allgower (1999). Section 2.1 provides an overview of the most important concepts of the causal median filter presented in Menold P.H. and Allgower (1999) and Section 2.2 explains the steps taken to adapt the causal median filter framework to the USBL outlier identification and classification problem. See Ting et al. (2007b) for a recent and alternative approach on the design of outlier detectors and classifiers using a Bayesian approach.

2.1 The Causal Median Filter

Most of the material presented in this section was carefully introduced in Menold P.H. and Allgower (1999) and it is introduced here to give the reader an overview of the theoretical basis for the design of the outlier classifier. Thus this section summarizes the most important concepts for the design of the outlier classifier. Consider the current observation x_k at time instant k and a data window W_k of fixed-width N

$$W_k = [x_{k-N+1} \cdots x_{k-1} x_k] \in \mathcal{R}^N.$$

If the values in W_k are sorted in descending or ascending order to obtain the sorted window R_k

$$R_k = \text{sort}(W_k)$$

the median x_k^\dagger is easily obtained as the mid-point of R_k as

$$x_k^\dagger = \begin{cases} R_k(\frac{N+1}{2}) & \text{if } N \text{ is odd,} \\ \frac{1}{2} (R_k(\frac{N}{2}) + R_k(\frac{N}{2} + 1)) & \text{if } N \text{ is even.} \end{cases}$$

The distance from the current data point x_k to the median value x_k^\dagger of the window W_k is given by

$$d_k = |x_k^\dagger - x_k|. \quad (1)$$

The data cleaning filter first identifies outliers by testing this distance d_k against a specified threshold $T_k \geq 0$ (which might depend on the data inside the window), and if the distance d_k exceeds the threshold T_k , then the current data point x_k is classified as an outlier. If the data point x_k is deemed an outlier, then it may be replaced by a prediction x_k^* to obtain a filtered sequence f_k given by

$$f_k = \begin{cases} x_k & \text{if } d_k \leq T_k, \\ x_k^* & \text{if } d_k > T_k. \end{cases}$$

or simply flagged to be an outlier so that systems that use this data downstream may now that it's not a reliable sample. The authors in Menold P.H. and Allgower (1999) mention several replacement strategies, which include, for instance, replacing the outliers by the current median value $x_k^* := x_k^\dagger$ or by the last valid value inside of the window W_k . If the outlier replacement actually takes place in the filtering framework, such setup is normally called a **data cleaning filter**. On the other hand, if the outliers are simply identified and marked, the setup is called an **outlier classifier**.

In the scope of this work we are not particularly interested in data cleaning filters since these tend to change the input data. We want to be able to provide raw acoustic USBL measurements to a myriad of systems and navigation filters on-board but with some sense of safety by flagging inappropriate data that might lead these navigation and control systems to instability. Navigation systems on-board the considered robotic platforms are typically based on dynamical systems that resemble the kinematics of rigid bodies and are able to provide open-loop numerical integration of other sensors such as accelerometers and rate gyros Morgado et al. (2008) when acoustic positioning systems data is not available or their measurements are suspected to be outliers. Moreover the effect of replacing outlier data points using the aforementioned strategies might introduce delays on the sequence and additional distortions on the noise characteristics of the signals that are difficult to model on the design of control algorithms and navigation filters. Thus, these replacement strategies should be used with appropriate care.

Threshold Selection. The threshold selection strategy adopted in this work and presented in Menold P.H. and Allgower (1999) is actually a combination of two strategies — the median absolute deviation (MAD) scale based threshold and a fixed lower bound for the threshold — and is given by

$$T_k = \max(cS_k, T_{min}) \quad (2)$$

where T_{min} is the lower bound for the threshold, S_k is an estimate for the MAD, and for some constant $c \in \mathcal{R}^+$, chosen independent of the data in the window W_k . The MAD scale estimate is defined as the median absolute deviation of the data points in the window W_k from the median x_k^\dagger , and is simply given by the median of the distances between all the data points in the window W_k and the median x_k^\dagger

$$D_k = [d_{k-N+1} \cdots d_{k-1} d_k] \in \mathcal{R}^N \quad (3)$$

where d_{k-i} with $i = \{0, 1, \dots, N-1\}$ is defined similarly to (1)

$$d_{k-i} = |x_k^\dagger - x_{k-i}|, \quad \forall i = \{0, 1, \dots, N-1\}.$$

Thus, the MAD scale estimate S_k is given by the median of D_k from (3). This unnormalized MAD scale estimate is often normalized Menold P.H. and Allgower (1999) to $\tilde{S}_k = S_k/0.6745 \approx 1.4826S_k$ to make it an unbiased estimate of the standard deviation for Gaussian data Huber and Ronchetti (1981). The choice of the scale parameter c in (2) will be addressed in Section 2.1.

The idea behind the dual strategy combination lies on the practical limitation with the MAD scale estimate being $S_k = 0$ for sequences that have, in a window of width N , at least $(N - 1)/2 + 1$ values (if N is odd, or $N/2 + 1$ if N is even), identical to the current data point x_k . If a lower bound T_{min} was not adopted, the threshold would be $T_k = 0$ in such cases regardless of parameter c . Thus T_{min} should be chosen taking into account the measurement noise level of the input signal and other parameters such as quantization and sensor resolution. Using this threshold selection rule, changes in the input sequence up to T_{min} are invariant, and as such it should not be chosen too large.

MAD Scale Parameter c . The results presented in this section provide a theoretical background on the design choice of the parameter c in (2). Most importantly they also provide lower bounds for c for certain type of sequences to be invariant under the MAD data cleaning filter. The following theorem establishes a lower bound for c under monotonic sequences that satisfy a growth rate restriction.

Theorem 1 ((Menold P.H. and Allgower, 1999, Theorem 5.1)). *Any monotonic sequence $\{x_k\}$ satisfying the growth rate restriction*

$$|x_{i+2} - x_{i+1}| \leq m|x_{i+1} - x_i|,$$

for some $m \in [0, 1]$ and $\forall i \in \mathcal{N}$, is invariant under the data cleaning filter of width $N = 4H + 1$ provided $c \geq 1 + m^H$.

Proof: See the proof of Theorem 5.1 in Menold P.H. and Allgower (1999).

An illustrative example of a monotonic decreasing sequence and the corresponding lower bound for c under three different window sizes is presented in Fig. 3. From this example it comes that as a rule of thumb, the parameter c should be larger than 2 for this type of sequences, that is $c \geq 2$.

The next set of results provide a basis to the lower bounding of the parameter c for two other distinct types of sequences defined in the following.

Definition 1 ((Menold P.H. and Allgower, 1999, Definition 5.1)). *A sequence of Type I satisfies the following conditions*

$$\begin{aligned} x_{k-2H} &= x_k^\dagger, & 0 < c_1 \leq c_2 < \infty, \\ c_1(2H - i) &\leq x_{k-i} - x_k^\dagger \leq c_2(2H - i) & \forall i \leq 2H, \\ c_1(2H - i) &\geq x_{k-i} - x_k^\dagger \geq c_2(2H - i) & \forall i > 2H, \end{aligned}$$

for all k and where x_k^\dagger is the median of the window of width $N = 4H + 1$.

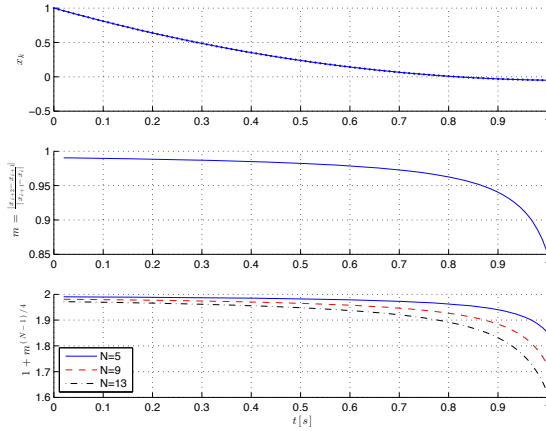


Fig. 3. Growth rate analysis of monotonic sequences and choice of the constant value $c \geq 1 + m^H$ with $H = (N-1)/4$

Definition 2 ((Menold P.H. and Allgower, 1999, Definition 5.2)). *A sequence of Type II satisfies the following conditions*

$$\begin{aligned}
 x_{k-2H} &= x_k^\dagger, \quad 0 > c_1 \geq c_2 > \infty, \\
 c_1(2H - i) &\geq x_{k-i} - x_k^\dagger \geq c_2(2H - i) \quad \forall i \leq 2H, \\
 c_1(2H - i) &\leq x_{k-i} - x_k^\dagger \leq c_2(2H - i) \quad \forall i > 2H,
 \end{aligned}$$

for all k and where x_k^\dagger is the median of the window of width $N = 4H + 1$.

The following theorem provides a lower bound for the parameter c under this type of sector bounded sequences defined in **Definitions 1** and **2**.

Theorem 2 ((Menold P.H. and Allgower, 1999, Theorem 5.2)). *Any sequence $\{x_k\}$ of type I or II is invariant under the MAD-based data cleaning filter of width $N = 4H + 1$ with $c \geq 2c_2/c_1$.*

Proof: See the proof of Theorem 5.2 in Menold P.H. and Allgower (1999).

Remark 1. The sequences of **Type II** in **Definition 2** are the decreasing analogous to the increasing **Type I** sequences in **Definition 1**.

In practice, the choice of the parameter c can be accomplished with the aid of the results presented in this section. For this purpose one could analyse sub-sequences of the nominal sequence and apply both Theorems 1 and 2 to compute a set of lower bounds for c , and then choose the largest lower bound that satisfies the invariance for the full sequence. On the other hand, it is recognizable that this procedure might be cumbersome, and probably the simplest way to choose a reasonable value for c is to examine and try out some values on training sets of the contaminated sequences. It is important to emphasize that c should also not be set too large, otherwise the outlier identification function will cease to be effective.

Table 1. USBL Data classification levels and flags

| Level | Flag | Description |
|-------|----------------|--|
| 0 | invalid | unrealisable solutions due to physical constraints of the USBL array: exceed the maximal allowed time delay between any two receivers on-board |
| 1 | valid | pass the physical limitations validation test but are yet unknown regarding its good or outlier |
| 2 | outlier | valid solution but clearly flagged as an outlier that violates the distance to the median of the window of valid samples |
| 3 | good | indicates that at least $2/3$ of the samples on the detection window were classified as valid |

Window Size N . The window size is also a very important parameter in the design of the outlier classifier, and it should be chosen to avoid observations from high dynamic range systems to be incorrectly considered outliers. Both the median x_k^\dagger and the MAD scale estimate S_k become less connected to local variations as N becomes too large and the analysis of a new measurement x_k less effective. On the other hand, N should also not be set too small in order to accommodate a reasonable amount of defective data patches. Data patches may occur for instance when a sensor saturates its output or errors in the measurements happen. For a window size of $N = 4H + 1$, both the median x_k^\dagger and the MAD scale estimate S_k are completely set to the value of a patch of $2H + 1$ samples. For instance, for $N = 9$, 5 patched outliers would undermine the effectiveness of the classifier.

2.2 Adaptation to the USBL System

The USBL positioning system provides measurements of the position of a transponder with respect to the reference frame of the robotic vehicle, that is, a $\mathbf{p} \in \mathcal{R}^3$. In order to adapt the outlier detection scheme to the USBL system, the algorithm outlined in Section 2.1 has to be extended to the three-dimensional case. The extension is fairly simple in which the window size is also extended to $W_k \in \mathcal{R}^{3 \times N}$ and the evaluation is performed separately for each of the three Cartesian coordinates.

The next improvement to be incorporated is the introduction of a classifier flag instead of performing outliers replacement, deriving what was named an **outlier classifier** in Section 2.1. A four level classification scheme is adopted that allows to introduce robustness to the classification process and the usage in downstream systems that require the classified data. The four levels can be summarized in Table 1.

Another feature that is needed due to the fact that underwater acoustics are highly susceptible to jamming and periods without actual measurements, is a time-elapsing window reset that removes observations from the window if their time tags are older than R seconds from the current system time. Finally, the global classification flag for each of the N -triplet values is assessed as follows: if any of its three values violates the threshold distance rule then the entire triplet is set as an **outlier**. The final

Table 2. Acoustic outlier classification algorithm

Algorithm *ClassifyData*(p, t)
 (* detect outliers and classify USBL positioning data *)
Input: p - current position measurement
Input: t - current measurement time
Output: class - classification level of current measurement
 (* persistent W_p positions window size $3 \times N$ init. to $\mathbf{0}$ *)
 (* persistent t_v last valid measured time tag init. to $\mathbf{0}$ *)

1. **if** $t - t_v > R$
2. **then** remove from W_p elements older than t_v
3. **else** insert the new data point in the window
4. $W_v \leftarrow$ select only valid elements from W_p
5. $m_k \leftarrow$ compute the row-wise median of W_v
6. $d_k \leftarrow$ compute the distances $|W_v - m_k|$
 (* Compute the MAD scale estimate *)
7. $S_k \leftarrow$ compute the row-wise median of d_k
 (* Normalize the MAD scale estimate *)
8. $\tilde{S}_k \leftarrow 1.4826 S_k$
 (* Threshold selection *)
9. $T \leftarrow \max(c\tilde{S}_k, T_{min})$
 (* Test the data point against the threshold *)
10. **if** $|p_k - m_k| > T$
11. **then** class = outlier
12. remove data point from the window W_p
13. **else** class = valid
14. update last valid time tag to current $t_v \leftarrow t$
15. **if** number of valid data points in $W_p > 2/3N$
16. **then** class = good

algorithm is outlined in Algorithm 2. The adaptation steps of the algorithm can be briefly summarized as follows:

- Extension to three dimensions
- Introduction of a classifier flag instead of performing outliers replacement
- Creation of a four-level classification scheme: **invalid**, **outlier**, **valid**, and **good**
- Introduction of an elapsed-time R window reset.

3 Simulation Analysis

The outlier classifier presented previously was first evaluated in simulation to assess its feasibility and performance. The nominal sequence to be tested was derived from the output of a second order spring-mass-damper system on the presence of small input step changes. Additive white Gaussian Noise was added to the output with a standard deviation of $\sigma = 0.03$. Ten percent of the values on the sequence were disturbed with outliers in random positions, with amplitudes in the interval $\pm[0.25, 0.6]$.

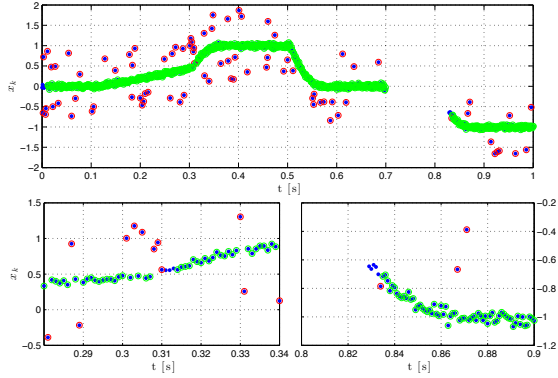


Fig. 4. Classification analysis in simulation results, with $N = 9$ and $c = 5$ — the parameter c was adjusted on this training set so that there are not false positive classifications as **good**

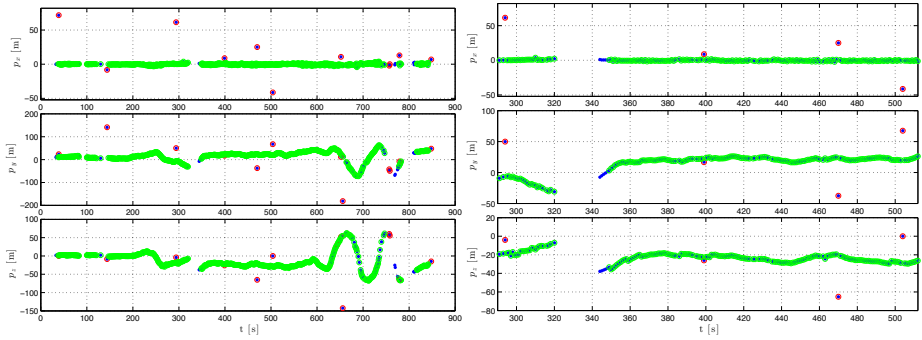


Fig. 5. Experimental results for the USBL positioning system during tests in Roses, Spain in October 2011. On the left: Overview of the total of approx. 900 seconds of data. On the right: Zoom from 290 to 512 seconds of operation.

The lower-bound for the threshold T_{min} was adjusted to accommodate disturbances on nominal, non-dynamically changing sequences up to 3σ of the additive white Gaussian noise perturbation, whereas the MAD scale estimate multiplier was initially set to $c = 6$ with a window size of $N = 9$. This first approach led to the conclusion that some values were being classified as false positive good values. It can be easily argued that is highly preferable to have false identifications of outliers on good values, rather than having outliers being classified as good values. Thus the value for the MAD scale estimate multiplier was adjusted to $c = 5$ and the classifier rerun on the same data. The remastered results are presented in Fig. 4, where it can be seen that there are no more false positives of good values while maintaining the performance on the remainder of the sequence. Lower-bounds for this c value can be found with the aid of Theorems 1 and 2 on sub-sequences of this training set, nonetheless it is always a good practice to

adjust this value bearing in mind the overall performance of the classifier on the entire training set.

4 Validation with Real Data

The devised outlier classifier was implemented and applied to a real USBL positioning system, fully developed in-house Morgado et al. (2010) and its outlier detection capabilities and performance are evaluated in this section. The parameters of the classifier were adjusted to: $c = 6$, a window-size of $N = 9$ samples, time reset constant of $R = 20$ seconds, and a threshold lower-bound of $T_{min} = 6$ meters. The outlier detection capability of the system in real world operation scenarios is evidenced in Fig. 5.

5 Conclusions

This paper presented an outlier detection and classifier algorithm with application to underwater acoustic positioning systems. The devised framework is based on causal median filters and a set of theoretical analysis tools that allows for the design of the filter parameters was also presented. Specific details that arise from the implementation of such an algorithm in real-world operation conditions were taken into account and a set of new features, such as a multi-level classification scheme and a time-based moving window reset, was added to cope with periods of acoustic data outage, which are quite typical in underwater scenarios. Interestingly enough, given the necessary window-size and computations, the outlier classifier is easily implementable in low-cost and low-power consumption digital signal processor (DSP) hardware. The outlier classifier was finally evaluated both in simulation and with experimental data from a prototype USBL acoustic positioning system fully developed in-house. Interesting ideas on future directions of research in this subject might include the validation of an adaptive algorithm for the choice of certain parameters in the filter.

References

- Gandhi, M.A., Mili, L.: Robust Kalman filter based on a generalized maximum-likelihood-type estimator. *IEEE Trans. Signal Processing* 58(5), 2509–2520 (2010)
- Huber, P.J., Ronchetti, E.: *Robust statistics*, vol. 1. Wiley Online Library (1981)
- Kinsey, J.C., Whitcomb, L.L.: Preliminary field experience with the DVLNAV integrated navigation system for oceanographic submersibles. *Control Engineering Practice* 12(12), 1541–1548 (2004) (invited paper)
- Kinsey, J.C., Eustice, R.M., Whitcomb, L.L.: A survey of underwater vehicle navigation: Recent advances and new challenges. In: *Proceedings of the 7th Conference on Manoeuvring and Control of Marine Craft (MCMC 2006)*, Lisbon, Portugal (2006)
- Larsen, M.B.: Synthetic long baseline navigation of underwater vehicles. In: *Proc. MTS/IEEE OCEANS 2004*, Providence, RI, USA, vol. 3, pp. 2043–2050 (September 2000)
- Lurton, X., Millard, N.W.: The feasibility of a very-long baseline acoustic positioning system for AUVs. In: *Proceedings of the MTS/IEEE OCEANS 2004 Conference*, Brest, France, vol. 3, pp. 403–408 (September 1994)

- Menold, R.K., Pearson, P.H., Allgower, F.: Online outlier detection and removal. In: Proceedings of the 7th International Conference on Control and Automation MED 1999, Haifa, Israel, pp. 1110–1134 (1999)
- Miller, P.A., Farrell, J.A., Zhao, Y., Djapic, V.: Autonomous underwater vehicle navigation. *IEEE Journal of Oceanic Engineering* 35(3), 663–678 (2010) ISSN 0364-9059, doi:10.1109/JOE.2010.2052691
- Milne, P.H.: *Underwater Acoustic Positioning Systems*. Gulf Pub. Co. (1983)
- Morgado, M., Oliveira, P., Silvestre, C., Vasconcelos, J.F.: Improving Aiding techniques for USBL Tightly-Coupled Inertial Navigation System. In: Proceedings of the IFAC World Congress 2008, Seoul, South Korea, IFAC (July 2008)
- Morgado, M., Oliveira, P., Silvestre, C.: Design and experimental evaluation of an integrated USBL/INS system for AUVs. In: Proceedings of the 2010 IEEE International Conference on Robotics and Automation (ICRA), Anchorage, AK, USA, pp. 4264–4269. IEEE (May 2010), doi:10.1109/ROBOT.2010.5509597
- Napolitano, F., Cretollier, F., Pelletier, H.: GAPS, combined USBL + INS + GPS tracking system for fast deployable and high accuracy multiple target positioning. In: Proceedings of the OCEANS 2005, Brest, France (June 2005)
- Pascoal, A., Oliveira, P., Silvestre, C., et al.: Robotic Ocean Vehicles for Marine Science Applications: the European ASIMOV Project. In: Proceedings of the OCEANS 2000, Rhode Island, USA (September 2000)
- Ting, J.A., Theodorou, E., Schaal, S.: A Kalman filter for robust outlier detection. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2007, pp. 1514–1519. IEEE (2007a)
- Ting, J.-A., D'Souza, A., Schaal, S.: Automatic outlier detection: A bayesian approach. In: Proceedings of the 2007 IEEE International Conference on Robotics and Automation, ICRA 2007, pp. 2489–2494 (April 2007b)
- Vaganay, J., Bellingham, J.G., Leonard, J.J.: Comparison of fix computation and filtering for autonomous acoustic navigation. *International Journal of Systems Science* 29(10), 1111–1122 (1998)
- Vickery, K.: Acoustic positioning systems. New concepts. In: Proc. 1998 Workshop Autonomous Underwater Vehicles, AUV 1998, Cambridge, MA, USA (August 1998)