Localization for Large-Scale Underwater Sensor Networks *

Zhong Zhou[1], Jun-Hong Cui[1], and Shengli Zhou[2]

[1] Computer Science& Engineering Dept, University of Connecticut, Storrs, CT, USA ,06269
[2] Electrical & Computer Engineering Dept, University of Connecticut, Storrs, CT, USA, 06269
{zhong.zhou,jcui,shengli}@engr.uconn.edu

Abstract. In this paper, we study the localization problem in large-scale underwater sensor networks. The adverse aqueous environments, the node mobility, and the large network scale all pose new challenges, and most current localization schemes are not applicable. We propose a hierarchical approach which divides the whole localization process into two sub-processes: anchor node localization and ordinary node localization. Many existing techniques can be used in the former. For the ordinary node localization process, we propose a distributed localization scheme which novelly integrates a 3-dimensional Euclidean distance estimation method with a recursive location estimation method. Simulation results show that our proposed solution can achieve high localization coverage with relatively small localization error and low communication overhead in large-scale 3-dimensional underwater sensor networks.

1 Introduction

Recently, there has been a rapidly growing interest in monitoring aqueous environments for scientific exploration, commercial exploitation and coastline protection. The ideal vehicle for this type of extensive monitoring is a distributed underwater system with networked wireless sensors, referred to as Underwater Wireless Sensor Network (UWSN) [1, 9]. For most UWSNs, localization service is an indispensable part. For example, in the long-term non-time-critical aquatic monitoring service [9, 13], localization is a must-do task to get useful location-aware data. Location information is also needed for geo-routing which is proved to be more efficient than pure flooding in UWSNs [20]. In this paper, we investigate the localization issue for large-scale UWSNs.

Localization has been widely explored for terrestrial wireless sensor networks, with many localization schemes being proposed so far. Generally speaking, these schemes can be classified into two categories: range-based schemes and range-free schemes. The former covers the protocols that use absolute point-to-point distance (i.e., range) estimates or angle estimates to calculate locations [12, 14, 6, 5, 18, 15], while the latter makes no assumptions about the availability or validity of such range information [7, 17, 16, 11, 19]. Although range-based protocols can provide more accurate position estimates, they need additional hardware for distance measures, which will increase the network cost. On the other hand, range-free schemes do not need additional hardware

support, but can only provide coarse position estimates. In this paper, we are more interested in accurate localization, which is requested by a range of applications, such as estuary monitoring and pollutant tracking [9]. Moreover, in UWSNs, acoustic channels are naturally employed, and range measurements using acoustic signals are much more accurate than using radio [9, 20]. Thus, range-based schemes are potentially good choice for UWSNs. Due to the unique characteristics (such as low communication bandwidth, node mobility, and 3-dimensional node deployment) of UWSNs [1, 9], however, the applicability of the existing range-based schemes is yet to be investigated.

There are also several schemes proposed for the localization service in underwater acoustic networks [4, 3, 21, 10]. These solutions are mainly designed for small-scale networks (usually with tens of nodes or even less). For large-scale UWSNs, hundreds or thousands of sensor nodes are deployed in a wide underwater area. Directly applying these localization schemes proposed for small scale underwater networks in large-scale networks is often inefficient and costly.

In this paper, for the first time, we explore the localization problem in large-scale UWSNs. We propose a hierarchical approach, dividing the whole localization process into two sub-processes: anchor node localization and ordinary node localization. Many existing approaches can be used in anchor node localization. For ordinary node localization, we propose a novel distributed method based on a 3-dimensional Euclidean distance estimation method and a recursive location estimation method. Simulation results show that our localization scheme can achieve high localization coverage with accurate location estimation and low communication overhead in large-scale 3-dimensional underwater sensor networks.

The rest of this paper is organized as follows. In Section 2, we describe our localization scheme. Simulation results are then presented in Section 3. And finally we draw conclusions in Section 4.

2 Localization for Large-Scale UWSNs

2.1 Overview

We consider a typical UWSN environment as shown in Fig. 1. There are three types of nodes in the network: surface buoys, anchor nodes, and ordinary nodes. Surface buoys are nodes that drift on the water surface. These buoys are often equipped with common GPS and can get their absolute locations from GPS or by other means. Anchor nodes are those who can directly contact the surface buoys to get their absolute positions. These nodes can also communicate with ordinary nodes and assist them to do localization. Ordinary nodes are those who can not directly talk to the surface buoys because of cost or some other constraints but can communicate with the anchor nodes to estimate their own positions. To handle the large scale of UWSNs, we propose a hierarchical localization approach. In this approach, the whole localization process is divided into two sub-processes: anchor node localization and ordinary node localization. At the beginning, only the surface buoys know their locations through common GPS or by other means. Four or more buoys are needed in our system. These buoys work as the "satellites" for the whole network, and anchor nodes can be localized by
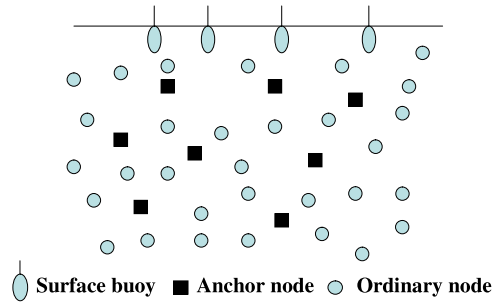
Fig. 1. A typical large-scale underwater sensor network setting

these surface buoys. Using surface buoys to locate underwater objects has been extensively investigated and many existing systems, such as [4] and [3], can be employed in the anchor node localization process. In this paper, we will not contribute to this part. Instead, we mainly tackle the problem of ordinary node localization, for which we propose a distributed range-based scheme, novelly integrating a 3-dimensional Euclidean distance estimation method and a recursive location estimation method. We describe this scheme in the following section.

2.2 Ordinary Node Localization

In 3-dimensional UWSNs, for a range-based localization scheme, ordinary nodes have to estimate their distances to more than 4 anchor nodes and calculate their locations by triangulation methods, which are commonly used in GPS systems. In a large-scale UWSN, however, not all ordinary nodes can directly measure their distances to 4 or more anchor nodes, thus some multi-hop distance estimation methods have to be developed.

In [18], the authors proposed and compared three multi-hop distance estimation methods: DV-Hop, DV-Distance and Euclidean. Even for two dimensional terrestrial sensor networks, the performance of DV-Hop and DV-Distance degrades dramatically in anisotropic topologies, while the Euclidean method can achieve much more accurate results and behave more consistently in both anisotropic and isotropic networks than other methods [18]. In a UWSN, since the sensor nodes are constantly moving due to many environment factors, the network topology may change unpredictably with time and space. Thus, the Euclidean method is expected to be more suitable for UWSNs than other approaches.

In our scheme, we employ a hybrid approach based on a 3-dimensional Euclidean distance estimation method and a recursive location estimation method to get the ordinary node positions. When combined with the recursive method, the inherent problems of the Euclidean method such as high communication cost and low localization coverage can be greatly alleviated. Next, we first discuss these two methods, examining why they can be seamlessly integrated together. Then we describe the ordinary node localization process in detail.

3-Dimensional Euclidean Distance Estimation In [18], a Euclidean distance propagation method is proposed for two dimensional sensor networks. Here, we extend it into 3-dimensional networks.

We use an example to illustrate the method. Referring to Fig. 2, if an ordinary node E wants to estimate its distance to anchor node A, it needs to know at least three (one-hop) neighbors (e.g., B, C, and D) which have distance estimates to A. Note that nodes A, B, C and D should not be co-plane and any three nodes out of A, B, C, D and E should not be co-line. Moreover, E needs to know its two-hop distance estimates, that is, E should have the length information of EB, BA, EC, CA, ED, DA, DB, DC, and BC. The 3-dimensional Euclidean distance estimation works as follows: First, node E uses edge BA, CA, BC to construct the basic localization plane. Since the lengths of edges DB, DA and DC are already known (to E), the position of D can be easily estimated. There exist at most two possible positions for D. Because E knows the lengths of edges ED, EB and EC, corresponding to the two possible positions of D, there will be at most four possible solutions for E's position. The choice among these four possibilities is made locally by voting when E has more immediate neighbors with estimates to A. If it cannot be decided, the distance estimate to A is not available until E gets more information from its neighbors.
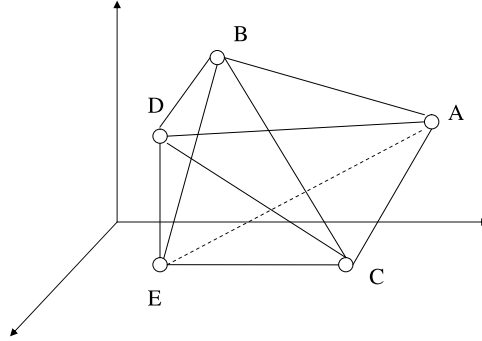
**Fig. 2.** 3-dimensional Euclidean estimation

Recursive Location Estimation In [2], the authors propose an iterative framework to extend the position estimation from a few reference nodes throughout the whole network. System coverage increases recursively as nodes with newly estimated positions join the reference node set, which is initialized to include anchor nodes.

This recursive location estimation method is illustrated in Fig. 3. In the figure, node 1 can get its location information from four neighboring anchor nodes A, B, C and D. If the location estimation error is small enough, node 1 can be regarded as a new reference node for other nodes. Then, it will broadcast its own location information. When node 2 gets to know the locations of C, D, E and 1 as well as the distances to these nodes, it can calculate its own location. On the other hand, if the location estimation error is

large, node 1 cannot be treated as a reference node and will not broadcast its location. In our scheme, the following formula is used to estimate the *location error* δ:

$$\delta = \sum_i \left| (u - x_i)^2 + (v - y_i)^2 + (w - z_i)^2 - l_i^2 \right|, \tag{1}$$

where (u, v, w) are the estimated coordinates of the unknown node, (x_i, y_i, z_i) are the reference node i's location, l_i is the measured distance between the unknown node and node i.
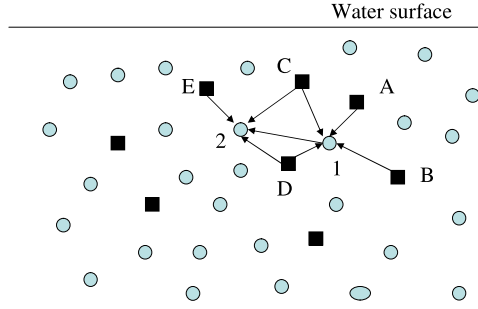
Fig. 3. Recursive location estimation

In order to alleviate the error propagation effect, every reference node in the system has a *confidence value* η. For the initial reference nodes (i.e., the anchor nodes), η is set to be the largest, while for a new reference node, η is associated with its location error. In our scheme, η is calculated as follows

$$\eta = \begin{cases} 1 & if \ \ node \ \ is \ \ the \ \ \ initial \ \ \ anchor \\ 1 - \dfrac{\delta}{\sum\limits_i (u - x_i)^2 + (v - y_i)^2 + (w - z_i)^2} & others \end{cases} \tag{2}$$

We can see that η is essentially a normalized δ. A critical value λ (referred to as "*confidence threshold*" later) is set. When $\eta > \lambda$, the unknown node can become a reference node. Otherwise, it will continue to be non-localized. When a node gets to know its distances to more than four nodes, it will choose four according to their η values and calculate its location.

Ordinary Node Localization Process In the ordinary node localization process, there are two types of nodes: reference nodes and non-localized nodes. In the *initialization phase*, all anchor nodes label themselves as reference nodes and set their confidence values to 1. All the ordinary nodes are non-localized nodes. With the advance of the localization process, more and more ordinary nodes are localized and become reference nodes. There are two types of messages: localization messages and beacon messages.

Localization messages are used for information exchange among non-localized nodes and reference nodes, while beacon messages are designed for distance estimates. During the localization process, each node (including reference nodes and non-localized nodes) periodically broadcasts a beacon message, containing its id. And all the neighboring nodes which receive this beacon message can estimate their distances to this node using techniques, such as TOA (time of arrival). We describe the actions of the two types of nodes as follows.

Reference Nodes: Each reference node periodically broadcasts a localization message which contains its coordinates, node id, and confidence value.

Non-localized Nodes: Each non-localized node maintains a counter, n, of localized messages it broadcasts. We set a threshold N (referred to as "*localization message threshold*") to limit the maximum number of localization messages each node can send. In other words, N is used to control the localization overhead. Besides, each non-localized node also keeps a counter, m, of the reference nodes to which it knows the distances. Once the localization process starts, each non-localized node keeps checking m. There are two cases:

(1) $m < 4$. This non-localized node broadcast a localization message which contains all its received reference nodes' locations and its estimated distances to these nodes. Its measured distances to all one-hop neighbors are also included in this localization message. Besides, this node uses the 3-dimensional Euclidean distance estimation approach to estimate its distances to more non-neighboring reference nodes. After this step, the set of its known reference nodes is updated. Correspondingly, m is updated and the node returns to the m-checking procedure.

(2) $m \geq 4$. This non-localized node selects 4 reference nodes with the highest confidence values for location estimation. After it gets its location, it computes confidence value η. If η is larger than or equal to the confidence threshold λ, then it is localized and labels itself as a new reference node. Otherwise, if η is smaller than λ, the node will take the same actions as described in case (1).

The complete localization procedure of an ordinary node is illustrated in Fig. 4.

3   Performance Evaluation

In this section, we evaluate the performance of our proposed localization scheme through simulation.

3.1   Simulation Settings

In our simulation experiments, 500 sensor nodes are randomly distributed in a $100m \times 100m \times 100m$ region. We define node density as the expected number of nodes in a node's neighborhood, hence node density is equivalent to node degree. We control the node density by changing the communication range of each node while keeping the area of deployment the same. Range (i.e., distance) measurements between nodes are assumed to follow normal distributions, with real distances as mean values and standard deviations to be one percent of real distances. 5%, 10% and 20% anchor nodes are considered in our simulations. Besides our scheme, we also simulate a Euclidean
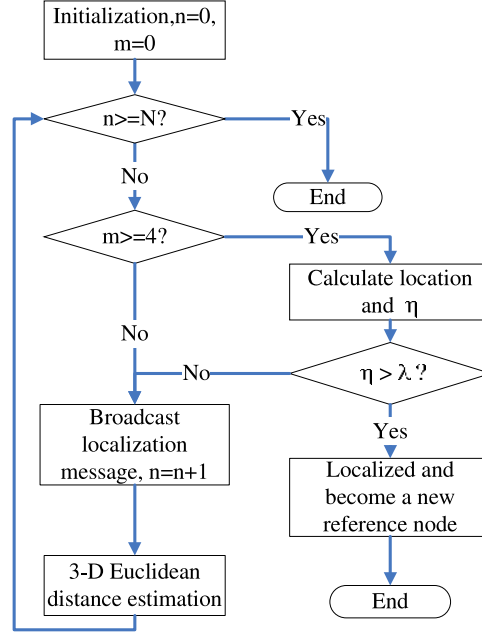
Fig. 4. Ordinary node localization process

scheme and a recursive scheme for comparison. The recursive scheme here is the same as in [2]. As for the Euclidean scheme, we use the three dimensional Euclidean distance estimation as the distance propagation method and then use the triangulation method to estimate an ordinary node's position if it gets to know four or more reference nodes. It works almost the same as the Euclidean scheme for two dimensional networks [18].

We consider three performance metrics: *localization coverage*, *localization error* and *average communication cost*. *Localization coverage* is defined as the ratio of the localizable nodes to the total nodes. *Localization error* is the average distance between the estimated positions and the real positions of all nodes. As in [18, 8], for our simulations, we normalize this absolute localization error to the node communication range R. *Average communication cost* is defined as the overall messages (including beacon messages and localization messages) exchanged in the network divided by the number of localized nodes.

3.2 Performance in Static Networks

In this set of simulations, nodes in the network are fixed. The confidence threshold λ is set to 0.98, and the localization message threshold N is set to 5. We change the node density (i.e., node degree) from 8 to 16 and compare our scheme with the Euclidean scheme and the recursive scheme. The results are plotted in Fig. 5, Fig. 6, and Fig. 7.
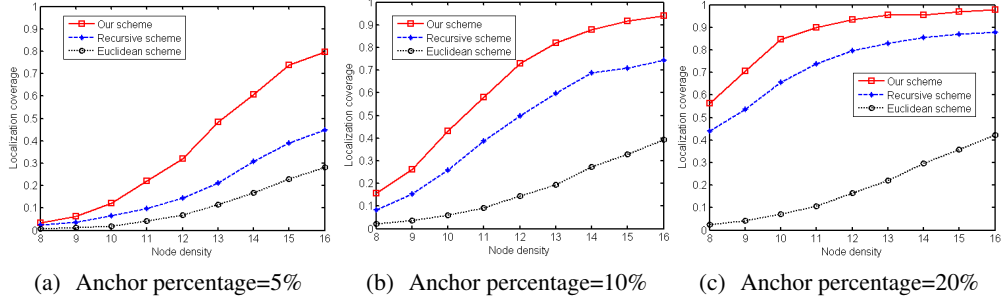
Localization Coverage Fig. 5 shows that our scheme outperforms both Euclidean scheme and recursive scheme in terms of localization coverage. This is reasonable since any node which can be located by either Euclidean scheme or recursive scheme can also be located by our scheme. The localization coverage of our scheme increases monotonically with the node density. But when the node density is relatively large, the coverage reaches a relatively high value and will not change much after that. For example, when the anchor percentage is 20%, the localization coverage reaches 94% at node density 12 and does not increase much with the node degree lifted. And we can also see that the more the anchors, the higher the localization coverage. For example, if the anchor percentage is 5%, the localization coverage can only reach 0.4 when the node density is 13, but if the anchor percentage is 10%, the localization coverage can reach 0.8 when the node density is 13. This suggests us that in sparse networks, we can increase the number of anchor nodes to achieve higher localization coverage.

Localization Error Fig. 6 plots the relationship between the localization error and the node density. We can observe that when the node density is relatively small, the localization error of our scheme is almost the same as that of the other two schemes. With the increase of the node density, the localization error of our scheme will increase and become a little larger than recursive scheme but much smaller than Euclidean scheme. This is because with the increase of the node density, the localization coverage of our scheme increases much faster than the other two schemes, as leads the growth of the localization error. But this growth is much slower rate than that of the localization coverage. As the node density continues to increase beyond some point, the localization error of our scheme will decrease slowly. This can be explained as follows. When the node density reaches a certain point, most sensor nodes can localize themselves. If we continue to increase the node density, ordinary nodes will get to know more anchor nodes and have more choices to calculate their locations. Thus, the localization error will decrease. But, as show in Fig. 6, this decrease is very limited. For example, when the anchor percentage is 5%, if we increase the node density from 13 to 16, the localization error only decreases from 0.3 to 0.27. Thus, in practice, we cannot expect to reduce the localization error by simply lifting the node density. Fig. 6 also shows us that the localization error will decrease observably with the anchor percentage. For example, at node density 13, when the anchor percentage is 5%, the localization error is 0.3. But when the anchor percentage is enlarged to 20%, it reduces to 0.05. Thus, more anchor nodes can translate into smaller localization errors.

Communication Cost Fig. 7 shows the average communication cost with the changing node density. In the recursive localization scheme, only nodes with known locations broadcast messages and other nodes keep silent. Therefore, the average communication cost of this scheme is very small. For our scheme, when the node density is small, it introduces larger communication cost than the recursive scheme. This is because in our scheme, when the network is sparse, although many nodes exchange beacon messages, they cannot finally localize themselves. In other words, these beacon messages are actually "wasted" in the localization process. But with the increase of the node density, this waste becomes smaller and smaller, and the average communication cost of our scheme

becomes closer and closer to the recursive scheme. From the figure, we can also observe that the average communication cost of our scheme decrease with the increase of anchor percentage. Compared with the Euclidean localization scheme, our scheme can always achieve much lower communication cost. This is due to that fact that the recursive component in our scheme help to find more reference nodes much faster than the Euclidean localization scheme.

(a) Anchor percentage=5%   (b) Anchor percentage=10%   (c) Anchor percentage=20%

Fig. 5. Localization coverage

(a) Anchor percentage=5%   (b) Anchor percentage=10%   (c) Anchor percentage=20%
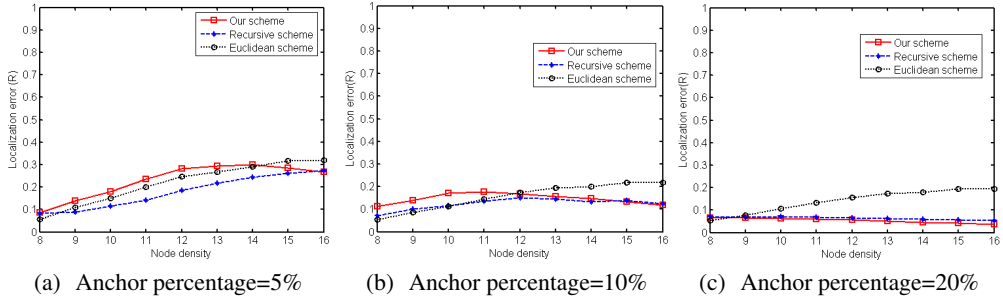
Fig. 6. Localization error

Discussions It is shown in [8] that range-based ad hoc localization schemes have high requirements on the node density of the networks. The paper also shows that in a two dimensional network, the node density needs to be at least 11 in order to localize 95% nodes with less than 5% localization error when 20% anchor nodes are present in the network. From Fig. 6(c), we can observe that when there are 20% anchors, our scheme can localize more than 95% nodes with less than 5% localization error if the node
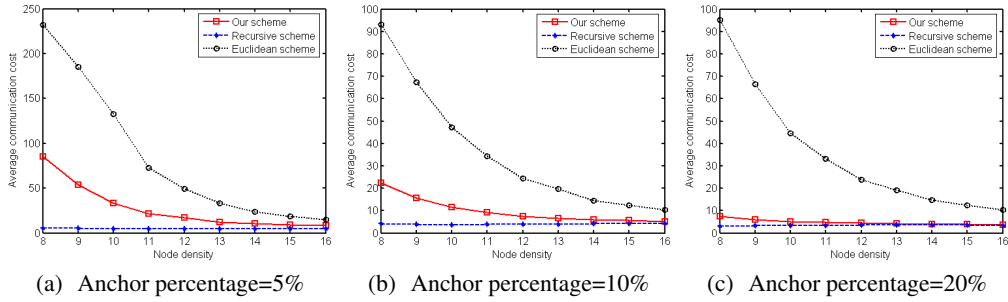
(a) Anchor percentage=5%   (b) Anchor percentage=10%   (c) Anchor percentage=20%

Fig. 7. Average communication cost

density is 12 in a 3-dimensional UWSN. Compared with the results in [8] for two dimensional networks, our scheme can achieve the same performance in 3-dimensional networks, with the connectivity requirement increased from 11 to 12. This indicates the good performance of our proposed scheme. On the other hand, this connectivity requirement of 12 may be still a little high for UWSNs with expensive sensor nodes or sparse deployment. One possible solution is to distinguish between the sensor's localization range and communication range. This means that we can increase the transmission power for the localization and beacon messages. In this way, the localization connectivity requirement can be satisfied while the contention among data will not increase much.

Besides the aforementioned results, we also study the impact of confidence threshold λ, the impact of the localization message threshold N, and the performance in mobile networks. In the following, we briefly summarize our findings for each aspect. Due to space limit, however, we do not include the detailed results in this paper. Interested readers can refer to our technical report [22].

Impact of Confidence Threshold: This study suggests us that by changing the confidence threshold, we can control the tradeoff between the localization error, the localization coverage and the average communication cost. For example, with the increase of the confidence threshold, the localization coverage and the localization error will decrease, while the average communication cost will increase. For UWSNs where location information is only used for geo-routing, high localization accuracy is not required [11], but a high localization coverage is desired. For this type of networks, the confidence threshold can be set to a relatively small value. While for UWSNs which require high precise location information, the confidence value should be set to a relatively large value. Some adaptive algorithms can be used to control this important parameter to provide performance guarantees.

Impact of Localization Message Threshold: This study tells us that for a network setting, there is a critical value of N. When N is smaller than this value, the localization coverage, the localization error and the average communication cost will increase rapidly. When N is larger than this value, the localization coverage and the localization

error will not change much and are relatively stable. But the communication cost will continue to increase. This indicates that beyond the critical value, increasing N will only increase the communication cost and will not bring any benefits. Thus, in practice we need to carefully choose N according to the network environments. In our previously presented simulations, we set N to 5, which is the critical value of N for the considered network setting.

Performance in Mobile Networks: We also conduct simulations to evaluate the performance of our scheme in mobile networks, and the results show that the localization coverage and average communication cost are not affected much by the node mobility, while the localization error increases noticeably with the node moving speed. This is mainly due to that fact that the average distance measurement error increases with the average moving speed, as naturally causes the increase of the final localization error.

4   Conclusions

In this paper, we presented a hierarchical localization approach for large-scale UWSNs. In this approach, the whole localization process consists of two sub-processes: anchor node localization and ordinary node localization. We focused on the ordinary node localization, for which we proposed a distributed scheme which novelly integrates a 3-dimensional Euclidean distance estimation method and a recursive localization method. Simulation results showed that our scheme can achieve high localization coverage with relatively small localization error and low communication cost. Besides, we also investigated the tradeoffs among the node density, the anchor percentage, the localization error, the localization coverage and the communication cost in our scheme. Different networks may have different requirements for these parameters. Via changing the confidence threshold parameter of our scheme, we can well control these tradeoffs.

References

1. I. F. Akyildiz, D. Pompili, and T. Melodia. Challenges for efficient communication in underwater acoustic sensor networks. *ACM SIGBED Review*, 1(1):3–8, Jul 2004.
2. J. Albowitz, A. Chen, and L. Zhang. Recursive position estimation in sensor networks. In *Proceedings of IEEE ICNP*, pages 35–41, Nov 2001.
3. T. C. Austin, R. P. Stokey, and K. M. Sharp. PARADIGM: a buoy-based system for auv navigation and tracking. In *Proceedings of MTS/IEEE Oceans*, 2000.
4. C. Bechaz and H. Thomas. GIB system: The underwater GPS solution. In *Proceedings of 5th Europe Conference on Underwater Acoustics*, May 2000.
5. P. Biswas and Y. Ye. Theory of semidefinite programming relaxation for sensor network localization. *To appear in matehmatical programming*.
6. P. Biswas and Y. Ye. Semidefinite programming for ad hoc wireless sensor network localization. In *Proceedings of IPSN*, pages 46–54, Apr 2004.
7. N. Bulusu, J. Heidemann, and D. Estrin. GPS-less low cost outdoor localization for very small devices. *IEEE Personal Communications Magazine*, pages 28–34, Oct 2000.
8. K. K. Chintalapudi, A. Dhariwal, R. Govindan, and G. Sukhatme. Ad-hoc localization using range and sectoring. In *Proceedings of IEEE Infocom*, pages 2662–2672, Mar 2004.

9. J.-H. Cui, J. Kong, M. Gerla, and S. Zhou. Challenges: building scalable mobile underwater wireless sensor networks for aquatic applications. *IEEE Network, Special Issue on Wireless Sensor Networking*, pages 12–18, May 2006.

10. J. E. Garcia. Ad hoc positioning for sensors in underwater acoustic networks. In *Proceedings of MTS/IEEE Oceans*, pages 2338–2340, 2004.

11. T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher. Range-free localization schemes for large scale sensor networks. In *Proceedings of 9th annual internatonal conference on mobile computing and networking*, pages 81–95, Sep 2003.

12. Kenneth and D.Frampton. Acoustic self-localization in a distributed sensor network. *IEEE Sensors Journals*, 6:166–172, Feb 2006.

13. J. Kong, J.-H. Cui, D. Wu, and M. Gerla. Building underwater ad-hoc networks and sensor networks for large scale real-time aquatic application. In *Proceedings of IEEE Military Communications Conference (MILCOM'05), Atlantic City, New Jersey, USA*, pages 1535–1541, Oct 2005.

14. A. Mahajian and M. Walworth. 3-D position sensing using the differences in the Time-of-Flights from a wave source to various receivers. *IEEE Transactions on Robotics and Automation*, 17:91–94, Feb 2001.

15. D. Moore, J. Leonard, D. Rus, and S. Teller. Robust distributed network localization with noisy range measurements. In *Proceedings of Sensys*, pages 50–61, Nov 2004.

16. R. Nagpal, H. Shrobe, and J. Bachrach. Organizing a global coordinate system from local inforamtion on an ad hoc sensor network. In *Proceedings of IPSN*, Apr 2003.

17. D. Nichulescu and B. Nath. DV based positioning in ad hoc networks. *Springer, Telecommunication Systems*, pages 267–280, Oct 2003.

18. D. Niculescu and B. Nathi. Ad hoc positioning system (APS). In *Proceedings of IEEE Globecom*, pages 2926–2931, Nov 2001.

19. H. Wu, C. Wang, and N.-F. Tzeng. Novel self-configurable positioning technique for multi-hop wireless networks. *IEEE/ACM Transaction on Networking*, pages 609–621, Jun 2005.

20. P. Xie, L. Lao, and J.-H. Cui. VBF: vector-based forwarding protocol for underwater sensor networks. In *Proceedings of IFIP Networking*, May 2006.

21. Y. Zhang and L. Cheng. A distributed protocol for multi-hop underwater robot positioning. In *Proceedings of IEEE International Conference on Robotics and Biomimetics*, pages 480–484, Aug 2004.

22. Z. Zhou, J.-H. Cui, and S. Zhou. Localization for large-scale underwater sensor networks. *UCONN CSE Technical Report: UbiNet-TR06-04, http://www.cse.uconn.edu/ jcui/publications.html*, Dec. 2006.