Mapping Swarms of Resource-Limited Sensor Motes: Solely Using Distance Measurements and Non-Unique Identifiers

Erik H.A. Duisterwinkel, Gijs Dubbelman, Libertario Demi, Elena Talnishnikh, Jan W.M. Bergmans and Heinrich J. Wörtche

***Abstract*— This work is on 3-D localization of sensor motes in massive swarms based solely on 1-D relative distance-measurements between neighbouring motes. We target applications in remote and difficult-to-access environments such as the exploration and mapping of the interior of oil reservoirs where hundreds or thousands of motes are used. These applications bring forward the need to use highly miniaturized sensor motes of less than 1 centimeter, thereby significantly limiting measurement and processing capabilities. These constraints, in combination with additional limitations posed by the environments, impede the communication of unique hardware identifiers, as well as communication with external, fixed beacons.**

We propose solving this challenging localization task by a novel RANSAC algorithm that can cope with noisy 1-D relative distance measurements and non-unique communication identifiers. It uses local geometric consistency, to resolve the ambiguity caused by non-unique communication identifiers and outlier measurements, and thereby is able to robustly assign unique hardware identifiers to be used for global non-linear graph optimization.

Extensive simulations show that this novel localization method is able to fully reconstruct the positions of the motes, in cases when the number of communication identifiers is only 2% of the number of motes. When the number of communication identifiers is lower, the algorithm exhibits graceful degradation.

I. Introduction

The exploration of remote, deep underground or difficult-to-access environments has been subject of study for decades. Current challenges include the mapping and exploration of the interior of narrow subterranean cavities like (oil) reservoirs and piping systems, which have high economic value and societal importance [1], and the investigation of the dynamics in industrial (multi-fluid) mixing tanks. The use of many (e.g. thousands), small and cheap mass-produced micro-sensor systems – from now on referred to as *sensor motes* which are operated in a *swarm* – is a promising way to approach these challenges. The specific sensor motes for mapping have many more promising applications, but are currently not available yet. The development for the larger sized sensor motes (5 cm diameter) for e.g. the mapping and exploration of a mixing tank is under way [2]. Many hardware challenges still need to be addressed before large swarms of miniaturized sensor motes (centimeter-sized or

This work was co-financed by the European Union (European Fund for Regional Development), the Dutch Ministry of Economic affairs (Peaks in the Delta), the Province of Drenthe and the Municipality of Assen.
Erik H.A. Duisterwinkel, Elena Talnishnikh, and Heinrich J. Wörtche are with INCAS3, [ErikDuisterwinkel, ElenaTalnishnikh, HeinrichWoertche]@incas3.eu. Gijs Dubbelman, Libertario Demi and Jan W. M. Bergmans are with the Signal Processing Group of Eindhoven University of Technology, [G.Dubbelman, L.Demi, J.W.M.Bergmans]@tue.nl

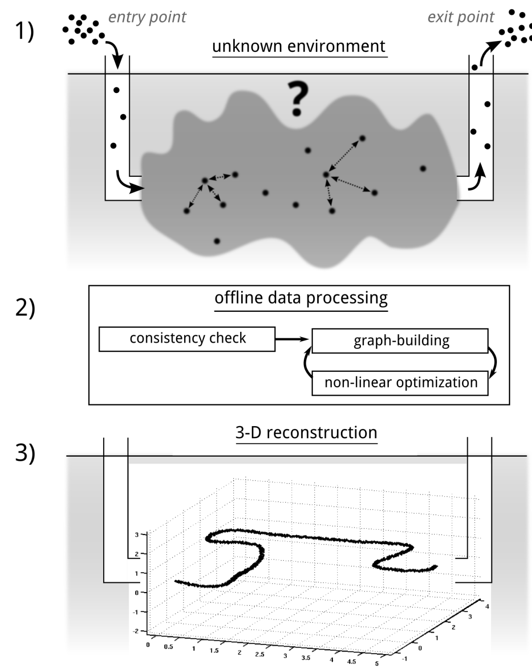

Fig. 1: Swarms of sensor motes are inserted in the environment of interest; they traverse the environment using its internal dynamics (e.g. flow); and are extracted in order to retrieve the data. Once the motes are distributed in the environment, distance measurements between neighbouring sensor motes are taken. Offline, this data can be processed to determine the motes positions relative to each other.

smaller) can be developed and deployed effectively. Trade-offs must be made in the hardware design with respect to mote capabilities and mote size. Improved mote capabilities, e.g. larger sensing radius or better signal-to-noise ratios, inevitably make cost-effective miniaturization of motes more challenging. Therefore, in our study, we set the challenge to perform reliable mote localization using severely limited mote capabilities. Although, these limitations might not be relevant to current motes, we believe they are relevant for yet-to-be-developed highly miniaturized motes of the near future [3]. We show that advances in off-line localization algorithms, can, to a large extent, compensate for severely limited mote capabilities. This is important as it implies that using large swarms of highly miniaturized motes in the future, is a realistic scenario. At the same time, our research provides lower bounds on mote capabilities that can guide the hardware design of yet-to-be-developed motes.

Each type of environment that is to be explored brings forward a specific set of constraints; in the application cases we are considering, the environments prevent the use of large, complex or existing sensing and localization systems, due to various issues, discussed in Sec. I-A. Exploration can in that case be performed by injecting swarms of sensor motes into the flooded environment, as e.g. illustrated in Fig. 1. The motes traverse the environment due to the internal dynamics (e.g. flow) and disperse over the volume of interest. The motes are extracted from the environment and the measured data that is stored in their internal memory can be analysed offline. The shape, or *map*, of the sensor swarm can be used to infer the geometric structure and size of the environment. The individual mote positions relative to each other – from which we can determine such a map – can be estimated from inter-more distance measurements [4], [5]. As illustrated in Fig. 2, estimating the positions of the motes can be seen as a specific kind of graph problem, $\mathcal{G}=(\mathcal{V},\mathcal{E})$, where the vertices $\mathcal{V}$ are the mote positions and the edges $\mathcal{E}$ the distances between them. The challenge is to obtain a robust initial estimate of the mote positions in order for non-linear graph optimization algorithms to succeed [6]. In contrast to earlier work [7] where a similar problem has been considered, in our work [8], an initial estimate is obtained using a novel robust Random Sampling Consensus (RANSAC) algorithm [9], [10], applied to general lateration techniques.

In this work, we explore the limits of swarm-based 3-D localization of extremely resource limited sensor motes with the specific set of constraints posed by the application and the environment. Our work differs from previous works [7], [9], [10], [12], [13], [14], [15], [16], [17] by,

1) using solely distance measurements (also called *range-only*), i.e. no direction information or additional sensor information, like bearing, odometry or inertia;
2) the distance measurements are performed using non-unique identifiers;
3) no additional data is exchanged, measurements are stored in memory for offline analysis;
4) not depending on external communication to e.g. fixed beacons; and,
5) sparse connectivity in a large swarm.

Although each separate constraint has been considered earlier; this work is, to the best of our knowledge, the first to show feasibility of reconstructing sensor swarms when considering all mentioned constraints and under realistic conditions. The main difference with our previous work in [8], is using non-unique identifiers (constraint 2) which has far-reaching consequences on the localization [11]. In [12], fully anonymous measurements are considered for robot localization, but cannot be compared to our problem, as the localization algorithm relies on the significantly larger sensing, online processing and communication capabilities of the robots. A similar complete problem as ours is considered in [18], [19] where a different solution is proposed to resolve the ambiguities. In both papers, only Gaussian noise is considered and the effect of ambiguities and noise on the final localization is not shown.

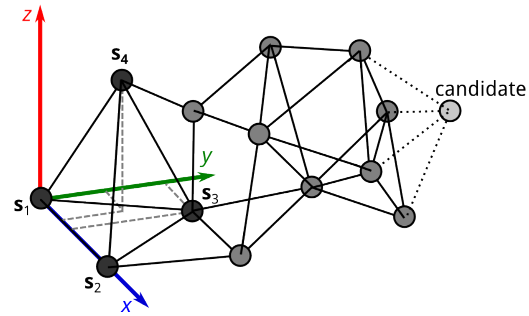

Fig. 2: Localizing motes within a swarm as a graph problem, $\mathcal{G}=(\mathcal{V},\mathcal{E})$, with the mote positions as vertices $\mathcal{V}$ and the distances between motes as edges $\mathcal{E}$. The initial four mote positions $\{\mathbf{s}_1,\mathbf{s}_2,\mathbf{s}_3,\mathbf{s}_4\}$ are chosen to define the coordinate system to solve the global reflection and rotation ambiguity. Additional motes can be added to the graph using general lateration techniques. For clarity of the figure, not all (required) edges are drawn.

This paper describes in Sec. II the exact problem that is considered. In Sec. III we elaborate on the localization algorithm. Sec. IV describes the simulation model we use to test the novel mapping method and localization algorithm in pipe-like environments. In Sec. V, we numerically evaluate the conditions/criteria for the localization algorithm to reliably estimate the mote positions in the swarm, while reducing the number of available communication identifiers. Our conclusions are provided in Sec. VI.

A. Constraints on sensing devices

In our application cases, we consider environments that are enclosed and/or deep underground and might have a narrow passage. Examples of these are the sand-free channels in heavy oil containing sandstone formations which are formed during oil extraction (Cold Heavy Oil Production using Sand, CHOPS) [1]. These sand-free channels, also called *wormholes*, are typically around one to several centimeters in diameter; extend over a few hundred meters; and are located 200-400 meter underground. Placing beacons is infeasible since the structure deep underground spans large distances and is not known beforehand. Communication using radio is not feasible due to the extreme high salinity. Acoustic communication however, is possible within the wormholes for communication between neighbouring sensor motes.

Passage through a wormhole is only possible if the physical dimensions of the motes are smaller than that of the local structure. For this specific application case it limits the sensor size to less than 1 centimeter in diameter, causing severe limitations on the instrumentation which can be taken aboard of the mote. At these scales, the major limitation is the energy storage, which limits the communication range, data rate and packet sizes, and the amount of processing that can be performed on-board. For these reasons we are looking into 1-D relative distance-only measurements which can be performed using ultrasound time of arrival or time of flight. This can for example be achieved using highly miniaturizable ultrasound technology, e.g. [20].

Identifiers can be encoded into the ranging pulse using e.g. BPSK or CDMA. But as ultrasound transducers are generally narrow-banded, data encoded in signal pulses causes signal pulses to be long [21], [20]. As these environments are small and reflective and the speed of sound is five orders of magnitude lower than that of radio waves, signal overlap due to emission of signals by neighbouring motes and their multi-paths is expected to be significant [21]. Elaborate communication techniques to deal with this, require more complex processing [22], which might not be feasible on the miniaturized motes. In order to reduce multi-path in these environments, we attempt to limit the length of the ranging pulse, and therefore, have to limit the quantity of identification information encoded in it. The motes unique hardware identifier (UID) can therefore not be encoded in the ranging-pulse, but only highly abbreviated and non-unique communication identifiers (CID) can be communicated. Consequently, received ranging-pulses from neighbouring motes can only be identified up to a large ambiguity.

II. Problem description

Following the operational procedure shown in Fig. 1, once a 'steady-state' distribution of the motes in the volume of interest is achieved, the motes perform simultaneously distance measurements to neighbouring motes. This can be achieved using e.g. a preset time after insertion. The ranging measurements are performed using ultrasound time-of-flight (TOF) or time-of-arrival (TOA). Omnidirectional ultrasound emission and reception is considered (as opposed to directional), because there is no a priori or online knowledge of the positions of neighbouring mote. Each mote reaches all neighbouring motes that are in line-of-sight, and within a specific sensing radius r_s.

We assume a homogeneous speed of sound, and that the movement of the motes during the distance measurements can be ignored, i.e. flow velocity of fluid is small compared to speed of sound in fluid . For now, the motes are also assumed to be synchronized such that obtaining the distances using TOA or TOF is straightforward. However, achieving this synchronization under the stringent communication constraints in the real application is not straightforward, but believed to be possible [3], [4]; it is a topic for future research.

A. Reconstructing mote positions

A general method of reconstructing the positions of the sensor motes based solely on distance measurements relies on general lateration methods. The graph $\mathcal{G}$ can be grown as illustrated in Fig. 2 and detailed in [8]. Every candidate mote, c, with known distances to four non-coplanar motes (of which the positions $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4$ are known) can be added to the graph. The position of mote c will be the intersection of the spheres with radii $\bar{d}_{1,c}, \bar{d}_{2,c}, \bar{d}_{3,c}, \bar{d}_{4,c}$ and centers at $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4$. Here $\bar{d}_{n,c}$ denotes the true distance from mote $n = \{1, 2, 3, 4\}$ to mote c. The initial four mote positions are chosen such that they define the coordinate system, therewith fixing the general reflection and rotation ambiguity: $\mathbf{s}_1 \in \{0, 0, 0\}$, $\mathbf{s}_2 \in \{\mathbb{R}, 0, 0\}$, $\mathbf{s}_3 \in \{\mathbb{R}, \mathbb{R}^+, 0\}$ and $\mathbf{s}_4 \in \{\mathbb{R}, \mathbb{R}, \mathbb{R}^+\}$.

The ranging measurements are performed in twofold, mote i measuring distance to mote j, $d_{\bar{i},\bar{j}}$, and vice versa, $d_{\bar{j},\bar{i}}$. In the data processing, this allows for the possibility to perform a forward-backward consistency check in determining the distance between the motes.

B. Identity and distance ambiguity

The motes are assigned at random one of the n_f available CIDs; all CIDs are distributed uniformly among the motes. The motes only emit this non-unique CID in the ranging pulse, rather than a unique UID. Consequently, the measured data per mote i will consist of a set of measured distances $\mathrm{d}_{\bar{i},\mathrm{CID}(j)}$ for all neighbouring motes j within its sensing radius r_s. These distance measurements are not uniquely associated with neighbouring motes j, but rather only to motes with $\mathrm{CID}(j)$. Therefore, the forward-backward consistency check, cannot be used to *uniquely define* a hypothesised distance between mote pairs, as in [8], but leaves ambiguities.

In order to build the graph, the non-unique distance measurement, $\mathrm{d}_{\bar{i},\mathrm{CID}(j)}$, should be associated with the uniquely identified motes $\bar{i}$ and j. This can be attempted by first considering all possible arrangements that are consistent with current measurements; similar as in [12].

With perfect distance measurements, the challenge to associate measurement $d_{\bar{i},\mathrm{CID}(j)}$ with mote j and $d_{\bar{j},\mathrm{CID}(i)}$ with mote i is straightforward. Searching for all possible arrangements that are consistent with the observation, i.e., $d_{\bar{i},\mathrm{CID}(j)} = d_{\bar{j},\mathrm{CID}(i)}$ will give most of the times an unambiguous correct result, since it is unlikely that another set of motes with similar CID pair have exactly the same distance.

However, distance measurements are subject to noise and result in an erroneous distance determination. The searching condition $\mathcal{F}$ for associating motes with measurements should be extended to

$$\mathcal{F}(\mathrm{d}_{a,b}, \mathrm{d}_{c,d}) = \begin{cases} \frac{\mathrm{d}_{a,b}+\mathrm{d}_{c,d}}{2} & \text{when } \mathrm{d}_{a,b} \simeq \mathrm{d}_{c,d} \\ \varnothing & \text{otherwise} \end{cases} \tag{1}$$

where $\mathrm{d}_{a,b} \simeq \mathrm{d}_{c,d}$ means $\mathrm{d}_{a,b}$ and $\mathrm{d}_{c,d}$ are *similar* to each other. We can define this similarity by the condition

$$\mathrm{d}_{a,b} \simeq \mathrm{d}_{c,d} \quad \text{when} \quad \begin{cases} |\mathrm{d}_{a,b} - \mathrm{d}_{c,d}| \leq \epsilon_r & \text{and} \\ \mathrm{CID}(a) = \mathrm{CID}(d) & \text{and} \\ \mathrm{CID}(b) = \mathrm{CID}(c) \end{cases} \tag{2}$$

where ϵ_r is set as threshold value for maximum allowed difference in distance measurement for which two measurements are considered inliers. This threshold should be related to the noise. It can e.g. be obtained using trial and error if the error model is unknown.

The obtained hypothesised distance between mote $\bar{i}$ and mote j is not unambiguous. As also illustrated in Fig. 3, the hypothesised distances and associations $\mathrm{h}_{i,j}$ for measurement $\mathrm{d}_{\bar{i},\mathrm{CID}(j)}$ can consist of three types of contributions:

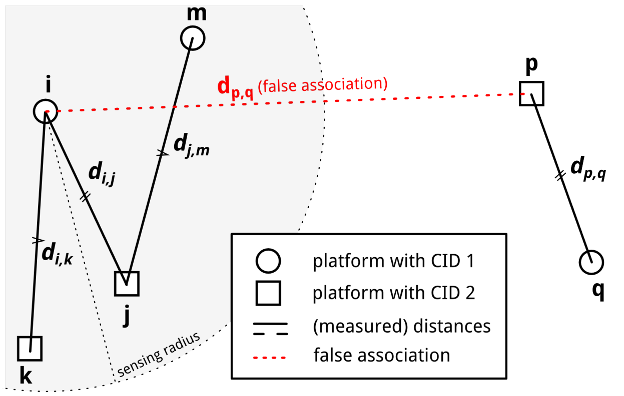

Fig. 3: Platforms with $\text{UID} = \{i, j, k, m, p, q\}$ using only two distinct CIDs (indicated by ◯ and □) cause ambiguities in associating UID with measured distances. Identity ambiguity arises due to *similar* distance between pairs with similar CID within the swarm (Eq. 2). Distance ambiguities arise due to a plurality of motes with similar CIDs within a sensing radius.

$$\mathrm{h}_{i,j} = \begin{cases} \mathcal{F}(\mathrm{d}_{\bar{i},\bar{j}}, \mathrm{d}_{\bar{j},\bar{i}}) & (a) \\ \mathcal{F}(\mathrm{d}_{\bar{i},j}, \mathrm{d}_{p,q}) & (b) \\ \mathcal{F}(\mathrm{d}_{\bar{i},k}, \mathrm{d}_{j,m}) & (c) \end{cases} \quad (3)$$

a) the measurement between the real motes $\bar{i}$ and $\bar{j}$;
b) *identity ambiguity*: an association to a mote p which is believed to be at similar distance as mote $\bar{j}$ but not necessarily within sensing range of $\bar{i}$ (see Fig. 3);
c) *distance ambiguity*: an association with the correct mote $\bar{j}$ but with a distance belonging to the measured distance to another mote k within its sensing radius (see Fig. 3).

The identity ambiguities are based on the statistical likelihood that somewhere in the swarm, a mote pair with similar CIDs and distance is present. Assuming uniform distribution of the motes, the average amount of identity ambiguities in $h_{i,j}$ scales with $\mathbb{A}_I \propto \epsilon_r n_s N / n_f^2$, in which n_s is the average amount of neighbouring motes within the sensing radius r_s and N the total amount of motes in the swarm. Equally, the distance ambiguities are based on the statistical likelihood that neighbouring motes give rise to confusion. The average amount of distance ambiguities per $\mathrm{d}_{\bar{i},\text{CID}(j)}$ then scales with $\mathbb{A}_D \propto \epsilon_r n_s^2 / n_f^2$.

Both type of ambiguities, $\mathbb{A}_I$ and $\mathbb{A}_D$, increase with an increasing amount of neighbouring platforms n_s, leading to a more challenging task associating measurements with the correct motes. However, our RANSAC algorithm, which is described in the next section, exploits the fact that measurements of true neighbouring motes can be used to resolve ambiguities by providing geometric consistency checks. An increase in ambiguities can be compensated for by a larger amount of consistent measurements from neighbouring motes.

III. Swarm localization algorithm

From the ambiguous set of hypothesised distances and associations $h_{i,j}$ in Equation 3, the graph $\mathcal{G}$ should be robustly built. We use a similar approach as in [8] but made significant modifications to prune false entries in $h_{i,j}$.

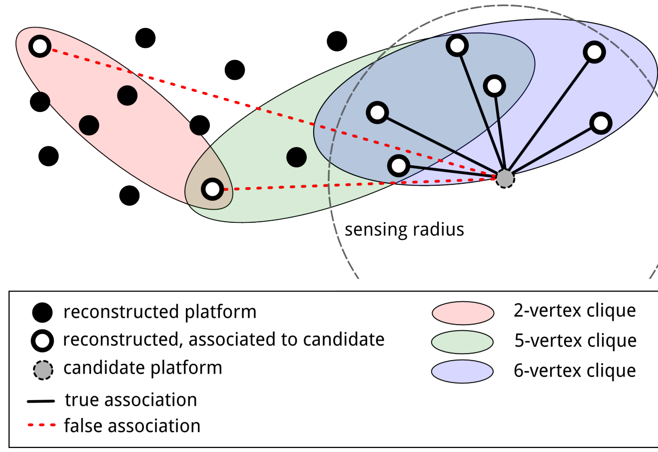

Fig. 4: RANSAC graph growing algorithm uses an inlier-outlier voting system to filter out outlier distance measurements and proposes a position for candidate motes to add to the graph. Using cliques of motes in the graph that are within twice the sensing radius helps in reducing the chance motes are positioned wrongly.

The main steps in our localization method are described next in Algo. 1 and Sec. III-A through Sec. III-D.

A. Initial seed selection

Four neighbouring motes are selected as initial seed to start the graph growing. Their positions, $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4$, define the coordinate system as described in Sec. II-A. The motes are selected based on their connectivity and the stability of their geometric configuration which is obtained by general lateration techniques. Since $h_{i.j}$ contains ambiguities, the initial seed selection is conditional until the graph growing has successfully added several motes to the graph $\mathcal{G}$. If this is not possible, a new initial seed is selected and the process is repeated.

B. RANSAC graph growing

Additional motes can be added to the graph when the motes have at least four connections to already reconstructed motes. This set of candidate motes is denoted as $\mathcal{C}$. Figure 4 illustrates a simplified 2-D situation in which a candidate mote $c \in \mathcal{C}$ has associations to its true neighbouring motes in $\mathcal{V}$, but also false associations to motes somewhere else in the graph due to identity ambiguities.

A RANSAC approach can be used to attempt to correctly position the candidate mote based on true associations only. The inlier-outlier voting mechanism is similar as in [8] but is adjusted to handle ambiguities. This is described in Algo. 2.

When a candidate mote c is considered, all motes that are already reconstructed in the graph and that are associated with this c are selected, $A = \{a : a \in \mathcal{V},\ h_{a,c} \neq \varnothing\}$. A set of three of these motes is selected, $\{a_p \in A\}$, to *propose* – using general lateration – a position for c up to a reflection ambiguity. But since no guarantees can be given whether the entries in $h_{a_p,c}$ are true distances or true associations, the

Algorithm 1: Localization

Data: $d_{\tilde{i},\text{CID}(j)}$, $\text{CID}(\{1, ..., N\})$
Result: graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
Hypothesise distances and associations $\text{h}_{i,j}$ using $\mathcal{F}$ (Eq. 1&3);
Create stable seed from 4 motes (Sec. III-A);
while *unreconstructed motes left* **do**
  select all $c \in \mathcal{C}$ that observe ≥ 4 motes in $\mathcal{V}$;
  sort c according to most likelihood to succeed (Sec. III-C);
  perform RANSAC on c (Sec. III-B & Algo. 2);
  if *motes c with sufficient RANSAC-score* **then**
    add motes with highest RANSAC-score to $\mathcal{G}$;
  else if *mote added since last non-linear refinement* **then**
    run non-linear refinement (Sec. III-D);
  else abort algorithm ;
  if *m new motes have been added to $\mathcal{G}$* **then**
    run non-linear refinement (Sec. III-D);

Algorithm 2: RANSAC

Data: sorted c^*, $h_{i,j}$, $\mathcal{Q}$
Result: RANSAC-score and proposed positions of c^* in $\mathcal{G}$
while *no good enough RANSAC-score* **do**
  select (next) largest clique in $\mathcal{Q}$ with associated candidate b;
  select all possible sets of 3 motes $\{a : a \in \mathcal{Q}(b)\}$;
  for *all a* **do**
    for *all a_p = possible (3-)sets of distances in $h_{a,b}$* **do**
      hypothesise position for b, based on lateration;
      if *hypothesised position $\notin \mathbb{R}^3$* **then** continue ;
      select all voters $\{a_v : (a_v \cup a_p) = a, (a_v \cap a_p) = \varnothing\}$;
      voters support hypothesis when observation $h_{a_v,b}$ is consistent with hypothesised position;
      voters oppose otherwise;
      if *# of supporters > # of opposers* **then**
        RANSAC-score hypothesis is # of supporters;
        if *RANSAC-score is good enough* **then**
          break **while**-loop;

proposed positions do not need to be viable or close to the true position. Therefore, each of the other associated motes $\{a_v : (a_v \cup a_p) = A, (a_v \cap a_p) = \varnothing\}$ are used to *vote* for the proposed position. Consensus is reached when the majority of the voting motes, including the three proposing motes, have at least a 50% majority. The reflection ambiguity is chosen based on which of the two positions received more supporting votes. When consensus is reached, the amount of supporting votes is called the *RANSAC-score* for the specific proposal.

This step in the RANSAC algorithm is repeated with each time a different set of three proposers a_p, until all possibilities are exhausted or until a proposal received a specific threshold in RANSAC-score. In the latter case, it is then considered *good enough* for addition in the graph. When no 'good enough' condition is reached, the RANSAC procedure is repeated for a next candidate mote until all possibilities are exhausted or the 'good enough' condition is found. The candidate with the highest RANSAC-score is added to the graph in $\mathcal{V}$ with edges only to the supporting motes and corresponding $h_{a,c}$ entry in $\mathcal{E}$. The 'good enough' condition for the RANSAC is chosen heuristically.

Algorithm 3: Sorting of candidate motes

Data: c, $h_{i,j}$, $\mathcal{V}$
Result: sorted c^* and cliques $\mathcal{Q}(c)$
for *every b in c* **do**
  select all motes $a \in \mathcal{V}$ with existing $h_{a,b}$;
  create graph $\mathcal{G}^*(b) = (\mathcal{V}(a), \mathcal{E}^*)$ in which all vertices $V(a)$ are connected with an edge $\mathcal{E}^*$ if their positions are within $2r_s$ of each other;
  find cliques $\mathcal{Q}(b)$ in $\mathcal{G}^*(b)$;
sort c according to clique size of largest cliques in $\mathcal{Q}(c)$;

C. Sorting candidate motes

The amount of false associations from candidate motes to the motes in the already reconstructed graph grows on average linearly with the amount of motes in the graph as explained in Sec. II-B. As a consequence, candidate motes might get voted to a wrong position in the graph during the RANSAC process in cases where the true neighbouring motes are outnumbered.

To reduce the chance of this to happen, a sorting order of candidate motes is made before the RANSAC algorithm is performed on them. The sorting is based on the likelihood that candidate motes have enough true neighbouring motes already in the graph. As such, candidate motes that are more prone to be positioned based on false connections, are considered later. This gives the chance that more of its true neighbouring motes will be added to the graph first. The algorithm is stated in Algo. 3 and described next.

For each candidate mote $\{b : b \in \mathcal{C}\}$, a list is made of associated motes that have already been positioned in the graph $A = \{a : a \in \mathcal{V},\ h_{a,b} \neq \varnothing\}$. A new graph $\mathcal{G}^* = (\mathcal{V}(A), \mathcal{E}^*)$ is made with these motes as vertices. These vertices are connected with edges $\mathcal{E}^*$ when their positions are less than twice the sensing radius, $2r_s$, apart. In this new graph, maximal cliques are listed as $\mathcal{Q}(b)$. These cliques indicate the groups of motes that are within the sensing radius of a proposed candidate mote that can potentially be agreed on by all of these motes. The example in Fig. 4 shows these cliques in coloured oval areas. The amounts of motes in these cliques are registered and are used as sorting order in which RANSAC is performed. Platforms with larger sized cliques in $\mathcal{Q}(b)$ are more likely to have more associations with true neighbouring motes.

D. Robust non-linear refinement

The stepwise addition of new motes to the graph introduces build-up of errors. These errors in positions can prevent other motes from being added. In order to reduce this error build-up, a global non-linear optimizer algorithm is performed [6], [8], after every m newly added motes, or failure to add a new mote.

There is a clear trade-off between the size of m and the processing time required for the localization algorithm. In our work, m is chosen arbitrarily to 10. In future work, this can e.g. be adjusted dynamically based on uncertainty of previously added motes.

IV. Simulation setup

The swarm-based distance-only localization can be used in a variety of environments. In this paper we will be using long, pipe-like structures as an example. The diameter of the pipe is large enough such that motes can be positioned all around one another instead of in one line along the pipe axis.

Simulations are performed 4-7 times, each time with a different pipe-structure. The course path of the loop-less, non-overlapping pipe without branches is randomly chosen with incremental steps of 1 meter over a single {x,y,z}-axis. The pipe axis is smoothed out by drawing a spline through these points. The pipe diameter is chosen to be fixed at 8 cm and there are 100 motes in every meter of pipe-length.

The positions of the motes are assumed to be uniformly distributed within the pipe; r_s is chosen such that each mote has on average n_s neighbouring motes. Parameter n_s is one of the parameters that will be swept. The motes sensing radius r_s, in which ranging measurements are possible, is for all motes the same.

In reality, homogeneous distribution of mote positions is difficult to achieve, but allows for thorough study of all other parameters involved.

A. Measurement noise

The measured distance between motes will be affected by different imperfections in the system. The distance measurement errors are modelled similar to what we have described in [8], which means we add additive Gaussian noise to account for inaccuracies in the timing of TOF and multiplicative Gaussian noise to account for imperfect knowledge of the speed of the (ultrasound) signal through the medium. Let $\bar{\mathrm{d}}$ be the true distance between a set of motes, then the measured distance d is modelled by

$$\mathrm{d}_{\bar{\mathrm{i}},\bar{\mathrm{j}}} = \bar{\mathrm{d}}_{\bar{i},\bar{j}} + \bar{\mathrm{d}}_{\bar{i},\bar{j}}\mathcal{N}_{\mathrm{m}} + \mathcal{N}_{\mathrm{a}} + \mathcal{B}\mathcal{U} \tag{4}$$

with $\mathcal{N}_{\mathrm{m}}$ and $\mathcal{N}_{\mathrm{a}}$ being perturbations from the zero-mean Gaussian distributions $\mathcal{N}(0, \sigma_{\mathrm{m}}^2)$ and $\mathcal{N}(0, \sigma_{\mathrm{a}}^2)$ respectively.

Outlier noise is added to the measurement with the term $\mathcal{B}\mathcal{U}$ in which $\mathcal{U}$ is a perturbation drawn from a uniform distribution in the range $[-r_s, r_s]$ and $\mathcal{B} \in \{0, 1\}$. The chance that a measurement has outlier noise, i.e. $\mathcal{B} = 1$, is determined by a Bernoulli distribution such that on average a percentage ω is an outlier. Any measurement that lies outside of the sensing range $(0, r_s]$ is discarded from further processing.

Identification noise, that accounts for the possible errors in determining the senders CID, is included by assigning a CID that is drawn randomly from the available set of CIDs to a measurement. The chance that a measurement has identification noise is determined by the ratio ϕ.

Initially, in Sec. V, we consider no outlier and identification noise, i.e. $\omega = 0\%$ and $\phi = 0\%$, and we consider three Gaussian noise levels: $\sigma_{\mathrm{m}} = \{0.30,\ 1.0,\ 3.0\} \cdot 10^{-3}$ and $\sigma_{\mathrm{a}} = \{0.36,\ 1.3,\ 3.6\} \cdot 10^{-3}\ m$. This yields an average standard deviation of the distance measurement noise of $\nu \approx \{0.4\%,\ 1.33\%,\ 4\%\}$ expressed in percentage of the average sensing radius when $n_s = 20$.

In Sec. V-A we investigate the robustness against additional outlier and identification noise with noise levels ω and ϕ ranging between 0% and 20%.

B. Performance parameters

To analyse the performance of the localization algorithm, we investigate the parameter *recall*, $\mathbb{X}$, which is the percentage of motes reconstructed by the localization algorithm. The recall parameter can only be interpreted correctly when also considering the *accuracy*. The accuracy is described by both the relative error and the absolute error of the mote positions. The relative error gives the error of the distance between two neighbouring motes, expressed as $\hat{d}/\bar{d}$, where $\hat{d}$ is the reconstructed distance between motes and $\bar{d}$ the true distance between them. The absolute error gives the error between the reconstructed position and the true position of the motes: $|\hat{\mathbf{s}}-\bar{\mathbf{s}}|$. Since the reconstruction is performed based on relative position only, i.e. relative to the initial four motes that determine the coordinate system, the absolute error is only useful when the reconstructed swarm is aligned with its ground truth. In order to solve this ambiguity, a linear fit between the 3-D positions of the initial four motes and their true positions is performed.

These experiments are designed to evaluate conditions/criteria of the localization algorithm for reliable reconstruction of the sensor swarm with non-unique identifiers. Besides the pipe structure and the motes positions, the parameters which play a role are the amount of motes in use (N), the amount of CID (n_f), the average amount of neighbouring motes within the sensing radius (n_s) and the distance measurement noise levels (ν, ω, ϕ).

A parameter set yields successful reconstructions when: a) the amount of added motes, based mainly on false associations, is below 5% of $\mathbb{X}$; and, b) the local pipe structure (its diameter and its course axial direction) is reconstructed correctly. Build-up of small directional changes over the course of the pipe axis are allowed for a reconstruction to be defined successful. In this work, this is assessed rather subjectively. As addition on b), successful reconstruction is also achieved when there is at one point in the reconstructed swarm a large deviation in the course path of the pipe axis relative to its true direction, but the positions of the motes after the this point are reconstructed correctly relative to the neighbouring motes. An example of such a *line-break* is shown in Fig. 7.a.

V. Numerical simulations

The *recall* $\mathbb{X}$ results of the experiments without outlier and identification noise are illustrated in Figure 5. Indicated in green are the parameter sets which yield successful reconstructions, grouped in: $\mathbb{X} = 100\%$, $\mathbb{X} \geq 80\%$ and $\mathbb{X} \geq 20\%$. Unsuccessful reconstructions are indicated in white, grouped in: $\mathbb{X} \geq 50\%$ (cross) and $\mathbb{X} < 50\%$ (plain). The green hatched areas indicate that either a maximum of one iteration has a significant pipe axis line break or several iterations have a minor pipe axis line break.

It can be seen that at each N and noise level ν, there exists a minimum n_s and $n_f \ll N$ for which successful reconstruction can be achieved with 100% recall. Increasing n_f beyond this point does not significantly improve the recall or accuracy parameters.

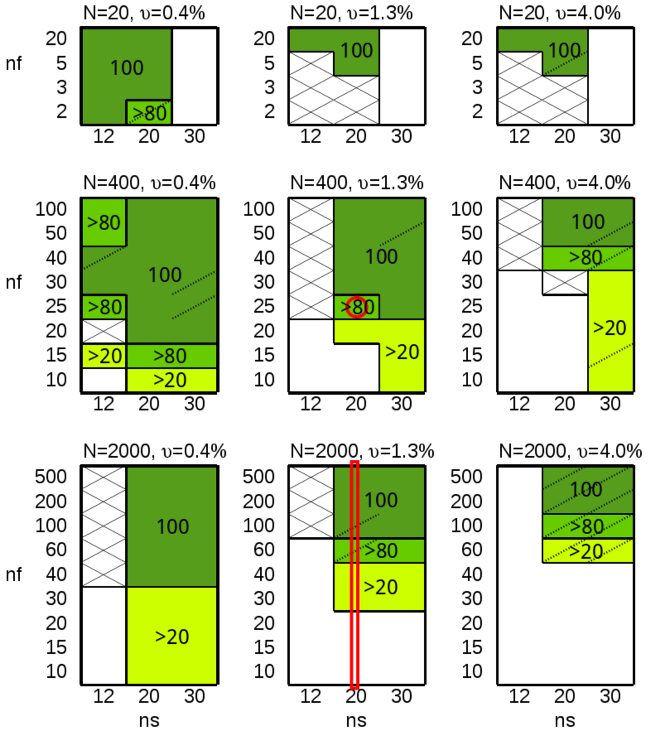

Fig. 5: Reconstruction of sensor swarm from simulations with different parameter sets N, ν, n_f, n_s. Recall $\mathbb{X}$ in five categories: sufficient reconstruction: $\mathbb{X} \geq \{100\%,\ 80\%,\ 20\%\}$ (green); unsuccessful reconstruction $\mathbb{X} \geq 50\%$ (white with cross) and $\mathbb{X} < 50\%$ (plain white). Red rectangle marks dataset used in Fig. 6, red cirle marks parameterset used in Sec. V-A.

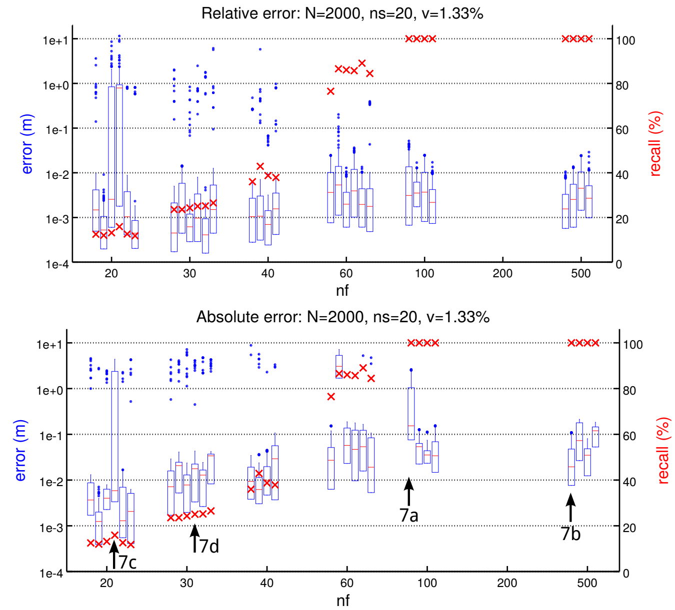

Fig. 6: Recall (red '×') plotted together with relative error (top) and absolute error (bottom) for all iterations with specified parameter sets (marked in Fig. 5 with red rectangle). Reconstructions of four selected runs (arrows) are shown in Fig. 7.

Figure 6, shows the relative and absolute accuracy of the reconstructions with the specific set of parameters highlighted in Figure 5 with a red line. The errors are shown in boxplots for all reconstructed sensor motes. Only the ten highest outliers are plotted ($>1.5\times$ the interquartile range); the lowest quartile and lower outliers are not plotted. As a reference, the recall is also plotted.

The amount of neighbouring motes n_s has a crucial role in whether a successful reconstruction can be achieved. In these experiments $n_s = \{12, 20, 30\}$ has been used and the recall results show that in most cases it is not sufficient to choose $n_s = 12$. In most cases, there is a significant chance that somewhere in the graph growing process, there are too little true neighbouring platforms already reconstructed for inlier votes in the RANSAC algorithm. A minimum of $n_s = 20$ is required for reliable reconstruction at larger N. This confirms our findings in [8]. Increasing n_s from 20 to 30, however, does not increase the recall significantly at these noise levels.

Figure 7 shows four reconstruction of different iterations with the parameter set from Figure 6. Figure 7.a and 7.b show the reconstruction of pipe structure with equal mote placement, both with a recall $\mathbb{X} = 100\%$. Figure 7.a has been made with $n_f = 100$ and shows a significant line break in the reconstructed pipe axis. Figure 7.b with $n_f = 500$ has significantly better reconstruction at this critical point in the swarm.

Figure 7.c and Figure 7.d show reconstructions in an equal pipe structure and mote placement, but with a different n_f. The lower amount of n_f increase the amount of ambiguities introduced in the dataset, and due to the coherent nature of these ambiguities, RANSAC can in this particular case not distinguish between correct candidate motes and false candidate motes to add to the graph. The reconstruction in Figure 7.c has an average total ambiguity factor (both distance and identity) of $\mathbb{A} = 480\%$ versus $\mathbb{A} = 210\%$ in Figure 7.d.

The reconstructions with parameter set $N = 400$, $\nu = 4\%$, $n_s = 30$, $n_f = 15$ has a dataset with the largest average ambiguity factor for which the reconstruction is still considered successful. With an average ambiguity factor of $\mathbb{A} = 1700\%$ it still has a reliable recall of $\mathbb{X} = 20\%$. The reconstruction of these iterations was terminated due to a self-imposed time constraint. All of the partial reconstructions (with sufficient accuracy) with $n_s = 30$ as well as the highly ambiguous datasets with $n_s = 20$ have been terminated due to this time-constraint. Some other high ambiguous datasets with $\mathbb{A} > 2000\%$ have been preventively terminated due to the large amount of total entries in $h_{i,j}$ (larger than 2×10^5).

A. Non-Gaussian outlier and identification noise

The results described above are based on measurement datasets without non-Gaussian outlier and identification noise. But as these types of noise are inevitable in realistic scenarios, we also study the localization algorithm on datasets including outlier and identification noise. The unmodified algorithm is performed with ω and ϕ ranging from 0% to 20%. This is performed on the parameter set $N = 400, n_s = 20, n_f = 25, \mu = 1.33\%$ (marked in Fig 5

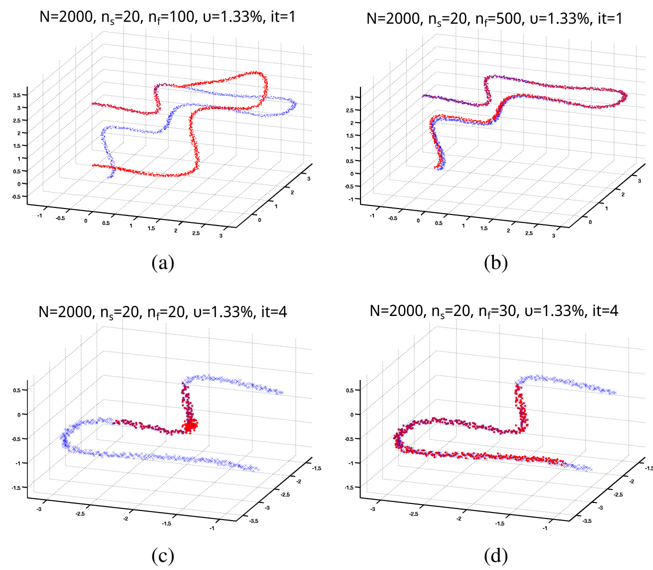

Fig. 7: (Parts of) reconstructions of four different runs (indicated with arrows in Fig. 6). Blue are true positions, red are reconstructed positions.

TABLE I: Recall $\mathbb{X}$ in percentage with outlier noise ω and identification noise ϕ of dataset $N=400, n_s=20, n_f=25, \mu=1.33\%$. The median of 10 iterations is taken; the lowest recall of the iterations is between brackets.

Outlier noise, ω	Identification noise, ϕ 0%	1%	5%	10%	20%
0%	99 (92)	100 (91)	99 (15)	96 (88)	65 (7)
1%	100 (89)	99 (91)	99 (87)	100 (72)	83 (1)
5%	100 (88)	100 (89)	98 (74)	100 (15)	27 (1)
10%	100 (75)	100 (87)	94 (12)	93 (9)	8 (1)
20%	83 (2)	70 (9)	64 (1)	21 (1)	1 (1)

with red circle). Recall results are shown in Table I. Outlier and identification noise of up to 10% still yields successful reconstruction with $\mathbb{X} \simeq 100\%$ for the majority of the iterations.

VI. Conclusion

The localization algorithm proposed in this paper was found to provide robust reconstructions of the motes relative positions in the novel swarm-based mapping approach with the use of non-unique communication identifiers. It uses a novel RANSAC method to obtain local geometric consistency of neighbouring motes for reconstruction. A realistic time-of-flight noise model has been used, which includes imperfect knowledge of the speed and timing of ranging pulses, as well as outlier noise in range measurements and communication identifiers.

Our work shows that depending on the noise levels and the geometric structure of the environment, there is a minimum amount of communication identifiers required for successful reconstruction. This can in our case be as low as 2% of the amount of motes in the swarm. Under the specific constraints that we are investigating, increasing the amount of identifiers above this minimum, or even using unique identification, does not significantly improve localization. This relaxation aids the feasibility of sensor swarm mapping using resource-limited sensor motes.

This work currently only considers a static case in which motes have enough energy to perform a single distance measurement. When considering cases in which the motes are able to perform subsequent multiple measurement, additional analysis can be performed: e.g. ambiguities can be filtered out better and the motes distribution throughout the environment can be less homogeneous; this is subject of future research.

References

[1] E. Talnishnikh *et al.*, "Micro Motes: A Highly Penetrating Probe for Inaccessible Environments", in *Intelligent Environmental Sensing*, edt. H. Leung, Springer International Publishing, 2015, pp 33-49.

[2] "Xploring WiseMotes", INCAS^3.

[3] EU Horizon 2020 FET-Open project: PHOENIX. www.phoenix-project.eu

[4] E.H.A. Duisterwinkel, "Asymmetric Multi-Way Ranging for Resource-Limited Nodes", in *8th EAI International Conference on Ad Hoc Networks*, Sept. 2016, Ottawa, Canada

[5] H. Wymeersch *et al.*, "Cooperative Localization in Wireless Networks", in *Proc. IEEE*, 97(2):427450, Feb. 2009.

[6] R. Kuemmerle *et al.*, "g2o: A General Framework for Graph Optimization", In *IEEE Int. Conf. Robotics and Automation*, 2011.

[7] S. Schlupkothen *et al.*, "A Novel Low-Complexity Numerical Localization Method for Dynamic Wireless Sensor Networks", *IEEE Trans. Signal Processing*, Aug. 2015

[8] G. Dubbelman *et al.*, "Robust Sensor Cloud Localization from Range Measurements", in *IEEE Int. Conf. Intelligent Robots and Systems*, Chicago, Illinois, USA, Sept. 2014.

[9] J. Djugash *et al.*, "Range-only SLAM for Robots Operating Cooperatively with Sensor Networks", in *IEEE Int. Conf. Robotics and Automation*, p.2078-2084, Orlando, Florida, USA, 2006.

[10] E. Menegatti *et al.*, "Range-only SLAM with a Mobile Robot and a Wireless Sensor Network", in *IEEE Int. Conf. Robotics and Automation*, p.8-14, Kobe, Japan, May 2009.

[11] E. Duisterwinkel *et al.*, "Environment mapping and localization with an uncontrolled swarm of ultrasound sensor motes", in *Proc. of Meetings on Acoustics*, 20(1), 2014.

[12] A. Franchi *et al.*, "Mutual Localization in Multi-Robot Systems using Anonymous Relative Measurements", *Int. Journal of Robotics Research*, vol. 32, no. 11 (2013): p.1302-1322.

[13] G. Han *et al.*, "Localization algorithms of Wireless Sensor Networks: a survey", *Telecommunication Systems*, 52(4):2419-2436, August 2011.

[14] M. Erol-kantarci and H. T. Mouftah, "Localization Techniques for Underwater Acoustic Sensor Networks", *IEEE Commun. Mag.*, p.152-158, Dec. 2010.

[15] J. A. Rothermich *et al.*, "Distributed Localization and Mapping with a Robotic Swarm", *Swarm Robotics*, Springer Berlin Heidelberg, 2005.

[16] J. McLurkin and J. Smith, "Distributed Algorithms for Dispersion in Indoor Environments Using a Swarm of Autonomous Mobile Robots", *Distributed Autonomous Robotic Systems 6*, Springer Japan, p.399-408, 2007.

[17] D. Hahnel, "Mapping and Localization with RFID Technology", In *IEEE Int. Conf. Robotics and Automation* (ICRA), New Orleans, LA (USA), 2004

[18] S. Schlupkothen and G. Ascheid, "Localization of wireless sensor networks with concurrently used identification sequences", in *IEEE Ad Hoc Networking Workshop* (MED-HOC-NET), 2015

[19] S. Schlupkothen *et al.*, "A dynamic programming algorithm for resolving transmit-ambiguities in the localization of WSN", in *Mediterranean Ad Hoc Networking Workshop* (Med-Hoc-Net), Vilanova i la Geltru, Spain, June 2016, pp. 1-8

[20] Z.D. Deng *et al.*, "An injectable acoustic transmitter for juvenile salmon", *Scientific Reports 5*, 2015.

[21] I.F. Akyildiz *et al.*, "Underwater acoustic sensor networks: research challenges", *Ad Hoc Networks 3*, 2005, p.257279.

[22] D. Dardari *et al.*, "Ranging With Ultrawide Bandwidth Signals in Multipath Environments", *Proc. IEEE*, vol. 97, 2009.