



RESEARCH ARTICLE

Reducing Uncertainty in Collective Perception Using Self-Organizing Hierarchy

Aryo Jamshidpey^{1,2}, Marco Dorigo¹, and Mary Katherine Heinrich^{1*}¹IRIDIA, Université Libre de Bruxelles, Brussels, Belgium. ²University of Ottawa, Ottawa, Canada.*Address correspondence to: mary.katherine.heinrich@ulb.be

In collective perception, agents sample spatial data and use the samples to agree on some estimate. In this paper, we identify the sources of statistical uncertainty that occur in collective perception and note that improving the accuracy of fully decentralized approaches, beyond a certain threshold, might be intractable. We propose self-organizing hierarchy as an approach to improve accuracy in collective perception by reducing or eliminating some of the sources of uncertainty. Using self-organizing hierarchy, aspects of centralization and decentralization can be combined: robots can understand their relative positions system-wide and fuse their information at one point, without requiring, e.g., a fully connected or static communication network. In this way, multi-sensor fusion techniques that were designed for fully centralized systems can be applied to a self-organized system for the first time, without losing the key practical benefits of decentralization. We implement simple proof-of-concept fusion in a self-organizing hierarchy approach and test it against three fully decentralized benchmark approaches. We test the perceptual accuracy of the approaches for absolute conditions that are uniform time-invariant, time-varying, and spatially nonuniform with high heterogeneity, as well as the scalability and fault tolerance of their accuracy. We show that, under our tested conditions, the self-organizing hierarchy approach is generally more accurate, more consistent, and faster than the other approaches and also that its accuracy is more scalable and comparably fault-tolerant. Under spatially nonuniform conditions, our results indicate that the four approaches are comparable in terms of similarity to the reference samples. In future work, extending these results to additional methods, such as collective probability distribution fitting, is likely to be much more straightforward in the self-organizing hierarchy approach than in the decentralized approaches.

Introduction

Increased autonomy in robot swarms is an open challenge. For example, in collective decision making, robot swarms cannot yet autonomously identify when a collective decision needs to be made and trigger the process [1]. Full collective autonomy would require task-general approaches to several constituent capabilities, including accurate collective perception and manageable collective actuation. However, existing approaches are often task-specific and features such as accuracy and manageability are challenging in swarm robotics—innovation is required if swarm autonomy is to increase markedly.

One approach to these challenges would be introducing some hierarchy into an otherwise fully decentralized system. Self-organizing hierarchy has been identified as a crucial research direction for the future of robot swarms [2,3]. However, if hierarchy is implemented in robot swarms, features such as scalability and fault tolerance must not be lost. Indeed, the motivation for studying self-organizing hierarchy is to combine aspects of centralized and decentralized control, ideally to get the benefits of both in one system. In this paper, we propose a self-organizing hierarchy approach to collective perception based on the existing concept of mergeable nervous systems (MNS) [4,5]. We empirically compared the proposed method to three fully decentralized

approaches as benchmarks, assessing whether accuracy is improved and scalability and fault tolerance are preserved.

In robot swarms, collective perception—i.e., the perception of an environment by a group of agents collaborating in a self-organized manner—can be viewed as a type of collective decision-making [6]. The swarm must both collect information and converge on a shared understanding of that information. Swarm robotics approaches to collective perception (e.g., [7,8]) are generally scalable and fault-tolerant because they do not rely on fully connected or static communication networks and do not have single points of failure, such as base stations or fixed leaders. These approaches also have strong potential for autonomy because they do not require access to external infrastructure or extensive prior knowledge. However, because fully decentralized approaches reach a collective decision via consensus, accuracy is challenging (compared to fully centralized approaches), and convergence times can be long [9].

Centralized approaches to perception generally make use of information fusion. Multi-sensor and multi-robot fusion problems are well understood, and existing methods are powerful [10–13]. However, these approaches typically know the positions and often also the poses of all robots or sensors in the system, using global positioning infrastructure, predefined or static positions, or other solutions that restrict scalability or

Citation: Jamshidpey A, Dorigo M, Heinrich MK. Reducing Uncertainty in Collective Perception Using Self-Organizing Hierarchy. *Intell. Comput.* 2023;2:Article 0044. <https://doi.org/10.34133/icomputing.0044>

Submitted 27 October 2022
Accepted 7 June 2023
Published 13 September 2023

Copyright © 2023 Aryo Jamshidpey et al. Exclusive licensee Zhejiang Lab. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY 4.0).

fault tolerance (compared to swarm robotics approaches). For most perception problems, fully decentralized information fusion has not yet been developed.

We propose self-organizing hierarchy as a way for robots to understand their relative positions system-wide and fuse their collective information at one point, without relying on restrictive features such as fixed positions, a fully connected or fixed communication topology, external infrastructure, or prior knowledge.

Related work

In this subsection, we give an overview of the existing fully decentralized approaches to collective perception with robots, in which perception is usually formulated as a best-of- n decision-making problem. We also give an overview of the existing multi-robot perception approaches that are not fully decentralized, which usually focus on the information fusion problem. We then discuss the topic of information fusion in self-organized systems, for which there are almost no existing approaches. Finally, we briefly discuss the perception of absolute versus relative conditions.

Collective and multi-robot perception approaches are summarized in the Table. They are organized according to whether perceptual information is fused at the data-level (e.g., images), feature-level, or decision-level, based on the multi-sensor fusion levels defined in [13]. They are also organized according to the type of perceptual decision being made: discrete best-of- n decisions [6], estimates of continuous values [6], or target tracking, e.g., of moving objects or stationary landmarks, including for mapping or simultaneous localization and mapping (SLAM).

Fully decentralized perception

All fully decentralized approaches to collective perception use a dynamic communication network that is not fully connected (i.e., robots cannot broadcast globally, and local connections are not static). The majority of these approaches are set up as best-of- n decision-making problems (see [6]), where a robot swarm compares multiple options according to a given criterion. In one common setup, robots sense colors in an environment and compare them according to the criterion of highest

representation [8,14–18]. In another setup, robots compare discrete zones according to a criterion of quality, which can be sensed from anywhere in the zone [7,19,20]. In similar setups, robots aggregate at a certain type of environmental feature, by either collectively deciding on an appropriate threshold to detect it [21] or sharing detected maximums so robots at a local optimum are triggered to explore further [22]. In [23], robots estimate the density of tiles scattered in the environment by sharing perceptual information, including sample counts and values for time-based decrementation. In [24], robots collaboratively track a moving target by sharing locally based beliefs about the target position with their neighbors, using incomplete knowledge of robot positions and orientations.

Not fully decentralized perception

All approaches that are not fully decentralized use fixed or known robot positions during information fusion. In most of these approaches, robots collaboratively track targets using known robot positions, e.g., by data-level fusion of sensor readings [26], feature-level fusion of perceptual information and associated uncertainties [28], or decision-level fusion of estimated positions [27], which can be supported by manual annotations merged at a base station [30]. In some tracking approaches, targets are tracked as part of mapping or SLAM, e.g., by decision-level fusion of estimated positions at a base station [29], and sometimes supported by prior information about the environment [31]. In SLAM, the positions of robots are not known beforehand, but the information needed to estimate all robot positions is available during the process of fusing information about the tracked targets (e.g., stationary landmarks). Besides target tracking, the literature also includes a few non-fully decentralized approaches for best-of- n decision making. In [32], robots establish a fixed communication network to train a system-wide artificial neural network that classifies a global light pattern in the environment. In [33], static robots with known relative positions send infrared sensor readings to a base station to identify objects from a predefined set. In [34], robots identify hand gestures from a predefined set and merge their opinions according to known robot positions. In this approach, the communication network is fully connected during information fusion (although the impact of packet loss is studied).

Fusion for robot swarms

Access to explicit or implicit positional information is required for the majority of existing multi-sensor fusion approaches (see [10–13]). These approaches are well-developed and would be useful for many applications if they could be implemented in robot systems. However, if these approaches were applied to robot systems using global broadcast or networks with fixed topologies, the systems' scalability and fault tolerance would be insufficient. Ideally, existing approaches that require positional information would be implemented in robot systems in a self-organized manner, but this combination is challenging. In the Table, there is one approach [27] that both fuses information based on known robot positions (i.e., starred in the Table) and can operate under dynamic topologies (i.e., blue and bold in the Table). However, this approach assumes full knowledge of absolute robot positions and orientations, the availability of which cannot always be guaranteed. In short, position-guided information fusion methods cannot currently be used for self-organized collective perception. In this paper, we propose self-organizing hierarchy based on the MNS concept as a general

Table. Existing approaches to collective perception and multi-robot perception, organized according to the level at which perceptual information is fused (see [13]) and the type of decision being made (best-of- n ; continuous, see [6]; or tracking). In the bold blue references, the topology of the communication network is dynamic and not fully connected. In starred (*) references, all relative robot positions in the system are either fixed or globally known and are used explicitly or implicitly during information fusion.

Decision type	Tracking	[25, *] [26, *]	[27, *] [28, *]	[29, *] [30, *] [31, *] [24]
	Continuous	[21]	[23] [22]	
	Best-of- n	[32, *] [33, *]		[8] [14] [17] [16] [18] [15] [7] [20] [19] [34, *]
	Data-level	Feature-level	Decision-level	
	Information fusion type			

framework in which existing multi-sensor fusion techniques could be implemented for collective perception, without using restrictive mechanisms (e.g., a fixed central coordinating entity or fixed communication topology) that impede desirable swarm robotics traits.

Perceiving absolutes

The fully decentralized approaches in the literature are applied to either the perception of a relative condition (e.g., whether red or blue is more represented) or the completion of a targeted action using sensed information (e.g., aggregation [22]). Results of these studies are not necessarily directly transferable to perception of an absolute condition. For instance, when perceiving a more represented color, if a swarm underestimates or overestimates absolute density, it might do so for each color somewhat consistently and therefore might still be able to accurately determine which color is more represented. The one exception in the literature is [23], in which perception of absolute density is assessed as an auxiliary contribution (the primary focus is completion of a targeted action). In the reported results, the swarm's perception of density diverges widely from the ground truth. In short, collective perception of absolute conditions still requires further study. In this study, we have tested some of the most accurate approaches for relative conditions in a new experimental setup in order to benchmark the approximate error (incorporating both bias and variance) present in fully decentralized approaches when perceiving an absolute condition.

Paper structure

The remainder of this paper is organized as follows. In Materials and Methods, we first discuss statistical uncertainty in collective perception and identify two sources of uncertainty that are not present in all spatial sampling and inference problems, but are crucial when sampling with mobile robots. We then introduce and describe the proposed self-organizing hierarchy approach, the three fully decentralized approaches selected as benchmarks for comparison, and the design and setup of the experiments. We report the results of the comparative experiments in Results, discuss the results and future work in Discussion, and finally summarize our conclusions in Conclusion.

Materials and Methods

Uncertainty in collective perception

Collective perception by ground robots is generally a two-dimensional (2D) spatial sampling and inference problem (cf. [35]). In other words, robots need to collect samples of information that varies spatially in two dimensions and use the samples to make some estimate, such as a mean or total value for an area (e.g., mean temperature, mean noise level, total daylight coverage, or total soil toxicity) or the positions of some objects within a relative coordinate system.

Spatial sampling and inference

In estimation of spatial data, there are three points from which uncertainty and error can originate, following [35]:

1. features of the stochastic field (\mathcal{R}), i.e., the distribution and variability of the spatial data,
2. the spatial sampling method (\mathcal{T}), and
3. the statistical inference method (ψ).

The main task is to minimize biases that originate in the sampling method \mathcal{T} and inference ψ . Note that biased \mathcal{T} can also be compensated for during ψ .

Collective perception with mobile robots

In principle, collective perception approaches should use parallel sample collection with multiple robots to overcome the bias that would be present in a single robot perceiving by itself.

In collective perception, robots do not have any prior knowledge of the stochastic field. A sampling method \mathcal{T} that introduces the least possible uncertainty would have a sampling ratio of 1 with perfectly evenly distributed sampling sites (see [35]). Therefore, parallel sampling in collective perception can minimize bias by increasing the total number of samples that are available to one robot, as well as increasing the spatial dispersion and potentially the density uniformity of the sampling sites (see Fig. 1).

However, because samples are taken by mobile robots that sweep an area over time, not all aspects of \mathcal{T} can be directly defined. For instance, simple random sampling cannot be programmed directly, but instead must be targeted indirectly by designing, e.g., a random walk and sampling time protocol. Furthermore, if the stochastic field \mathcal{R} is time-varying, spatio-temporal correlations and variability aspects that are not representative of \mathcal{R} could be introduced during sampling due to the robot's sweeping motion. Also, in collective perception, robots need to incorporate samples collected by their peers in parallel. In most approaches, robots sample the decisions of their peers, in addition to sampling the stochastic field directly, and fuse decisions during the inference process ψ .

Therefore, in collective perception with mobile robots, there are five points at which uncertainty and error can originate

- 1.–3. \mathcal{R} , \mathcal{T} , ψ (following [35]),
4. the spatial decision sampling method (\mathcal{G}), and
5. the spatiotemporal sweeping method (\mathcal{P}).

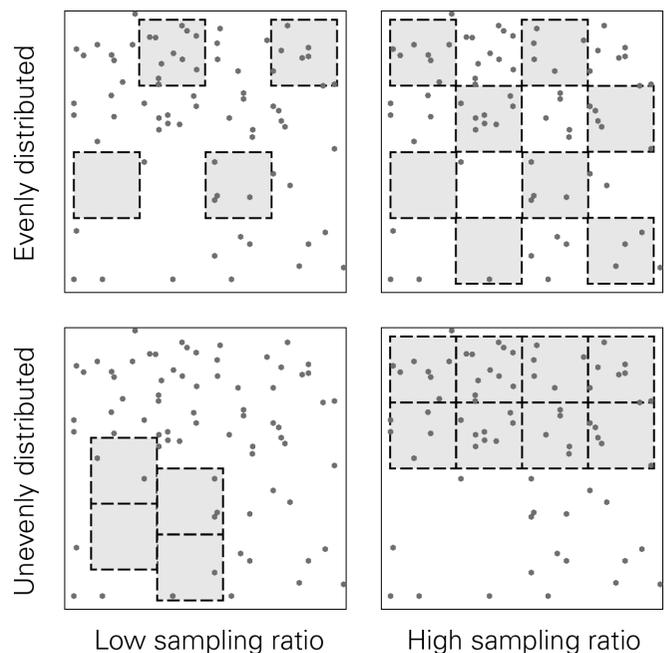


Fig. 1. When prior knowledge about a stochastic field is not known, higher sampling ratios and more evenly distributed samples can reduce the potential for bias to be introduced during the sampling method \mathcal{T} .

In collective perception, a random walk is typically used to sweep the environment because robots are assumed to have no or very little prior knowledge of the layout of the environment or the stochastic field \mathcal{R} . Random walks can be tuned or modulated to improve the sampling distribution (see Fig. 1), but some interference from obstacles, environment layout, or other robots is unavoidable. An optimal random walk sweep \mathcal{P} would reliably result in a uniform random distribution of sampling sites in \mathcal{T} and \mathcal{G} —i.e., \mathcal{T} and \mathcal{G} would be simple random sampling. However, simple random sampling often leads to large gaps and therefore bias [35], so even if an optimal random walk free of interference were practical, substantial uncertainty could still originate during \mathcal{T} and \mathcal{G} . Consider that although biased \mathcal{P} , \mathcal{T} , and \mathcal{G} can be compensated for during inference ψ , it is difficult or impossible to do so without prior knowledge of the stochastic field \mathcal{R} [35]. Therefore, we conclude that it might be intractable to substantially improve the performance of fully decentralized approaches to collective perception beyond the current state of the art (e.g., [7,8]).

In this paper, we assert that by reducing or eliminating key sources of uncertainty that are usually present in collective perception, a well-designed self-organizing hierarchy approach could provide much more accurate estimates without sacrificing scalability and fault tolerance, as compared to fully decentralized approaches. We tested this assertion empirically.

Problem statement

We have developed a self-organizing hierarchy approach to collective perception based on the MNS concept and tested it against fully decentralized collective perception in simulated

experiments. In the **self-organizing hierarchy approach** (HIER), a robot fulfilling the temporary role of a “brain” forms one collective opinion on behalf of the group, using collective sensor information merged hierarchically by all robots. In the fully decentralized approaches, each robot takes part (either explicitly or implicitly) in a collective decision-making process to form its own opinion and reach decentralized consensus with its peers.

We tested the HIER approach against the following three fully decentralized approaches, as benchmarks for comparison.

- **Voter decision model** (VOTE): Each robot selects new opinions randomly from among the current opinions of itself and its neighbors (based on [8]).
- **Mean decision model** (MEAN): Each robot averages the opinions of itself and its neighbors (based on [8]).
- **Stigmergy** (STIG): Robots do not communicate explicitly but leave cues for each other to observe in the environment (based on [36]).

In all approaches, simulated ground robots use short-range onboard sensing to detect some objects that are distributed randomly in an arena of unknown size. Their collective goal is to form an accurate opinion on the density of objects in the whole arena.

In this paper, robots collect samples by detecting individual objects and then infer either the mean or distribution of the density of objects in the arena using odometry and knowledge of their own sensor ranges.

Absolute object density $\lambda = b/a$ is defined as the number of objects b per unit area a . The true density λ in the environment

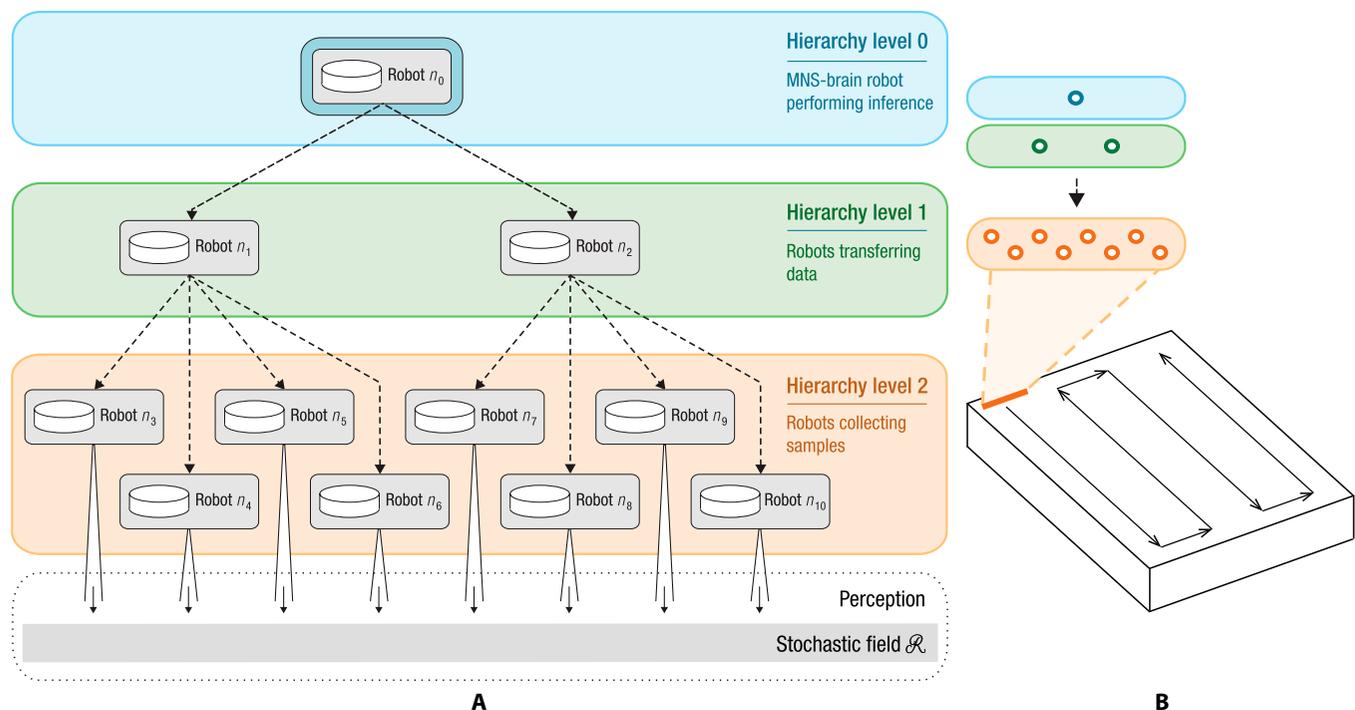


Fig. 2. In the HIER approach, a dynamic ad hoc hierarchical network self-organizes using local communication and relative positioning. (A) Roles of robots in the collective perception task according to their hierarchy level. The 8 robots in level 2 (orange) collect samples of the stochastic field \mathcal{R} , the MNS-brain robot in level 0 (blue) performs inference, and any robots between them (green) transfer data. (B) Reactive boustrophedonic motions to sweep the environment. The MNS-brain robot in level 0 (blue) is responsible for determining the reactions and sending motion instructions downstream, the robots in level 2 (orange) are responsible for obstacle avoidance, and any robots between them (green) monitor their children’s positions and calculate their motion instructions.

is denoted by λ^{true} . Robots are not given any information about the size and shape of the objects or arena. Robots only know the dimensions of their own fields of view and must use this knowledge to infer the number of objects per unit area.

Experimental design

To study the baseline performance of the tested methods, we constructed stochastic fields \mathcal{R} with relatively low spatial variability by distributing the features (objects to be detected) uniformly randomly in a regular polygonal area. Inference ψ is still not negligible, even in the case of optimal sampling, because the robots cannot detect the target value (i.e., density) directly and must instead infer it from representative information (i.e., short-range Boolean detection of objects).

We tested the approaches under several time-invariant and time-varying \mathcal{R} . Each approach was assessed in terms of perceptual accuracy (i.e., error in the collective opinion with respect to the true value), consistency of the accuracy, and reaction time. The scalability and fault tolerance of each approach was also assessed in time-invariant \mathcal{R} in terms of the change in accuracy under robot failures and group size variations.

Methods for collective perception

The robots run two parallel processes.

- Process A: Robot r individually counts detected objects and infers the density in its own field of view.
- Process B: Robots influence the collective opinion of the group via either the inputs or outputs of Process A, and robot r forms its opinion on the absolute object density (λ) in the arena.

Robots use only local or indirect communication to influence the collective opinion of the group, so the motion routines and communication rules of each approach are presented in the descriptions of Process B.

Process A is the same in each approach (HIER, VOTE, MEAN, and STIG), except for the tuning parameters. Process B is different in each approach.

Process A: Robots make individual interpretations

In all approaches, robot r counts sensed objects and infers the density in its (direct or indirect) field of view, outputting the value ϵ_r . The sensing and inference output ϵ_r , which is equal to 0 in the first time step, is defined as the average number of objects per unit area in the robot’s field of view in a given time window, calculated as follows for time t :

$$\epsilon_r = \frac{\sigma_t - \sigma_{t-k_t}}{\nu \cdot s \cdot k_t^{\text{max}}} \cdot P, \tag{1}$$

where σ_t is the cumulative number of objects seen from the first time step to the current time t by robot r , k_t is the time window at time t , k_t^{max} is the maximum time window, and the remaining terms are tuning parameters: ν is the view area of robot r , s is the average speed of robot r , and P is a parameter related to the motion routines of the different approaches. For details on the tuning parameters, see the “Simulation setup” section.

The time window k_t keeps track of the elapsed time since the robot started the mission, up to the maximum time window k_t^{max} , i.e.,

$$k_t = \begin{cases} t, & \text{if } t < k_t^{\text{max}} \\ k_t^{\text{max}}, & \text{otherwise} \end{cases} \tag{2}$$

At each time step, a new value σ_t is saved to a circular buffer σ of length k_t^{max} in the memory of robot r . In other words, after σ reaches k_t^{max} elements (i.e., when $t \geq k_t^{\text{max}}$), then at every subsequent update of σ_t (at position $i = k_t^{\text{max}}$), each element at $i: i \in \{1, \dots, k_t^{\text{max}} - 1\}$ is replaced by the element at $i + 1$. The buffer σ is defined as

$$\sigma = (\sigma_{t-k_t}, \sigma_{t-k_t+1}, \dots, \sigma_t), \quad \sigma_t = \sum_{i=1}^t b_i, \tag{3}$$

where b_i is the number of objects detected (directly or indirectly) at the current time step t by robot r . Note that, to calculate σ_t , only σ_t and the most recent b_{t-1} values are required; it is not necessary to maintain a memory of all previous b_i values.

In summary, at each time step t , Process A: inputs b_t , the number of objects the robot detects, and outputs ϵ_r , the inferred density in the robot’s field of view during the time window.

Process B: Robots influence the collective opinion

In all approaches, all robots influence the collective opinion through either the inputs or outputs of Process A. Also, each robot r uses the outputs of Process A to form its opinion λ_r^{app} of the apparent absolute density in the arena.

In the two decision model approaches (VOTE and MEAN), robots modulate the input b_t individually and process the output ϵ_r collectively. In the other two approaches (STIG and HIER), robots do the opposite—they modulate the input b_t collectively and process the output ϵ_r individually. Thus, in all four approaches, the opinion λ_r^{app} is influenced by a collective process.

Decision model approaches (VOTE and MEAN): Process B

We tested two fully decentralized approaches that use a collective decision-making process to reach consensus about the apparent density in the arena. Both approaches use a basic stochastic motion routine and explicit communication among neighbors. The robots coordinate their opinions using either a voter or mean decision model.

Stochastic motion routine

The VOTE and MEAN approaches use a stochastic motion routine based on RANDOM BILLIARDS [37,38]. Each robot moves forward at a constant velocity unless it detects the boundary line of the arena. A robot can detect a line’s angle relative to its own heading by driving over the line and, likewise, can detect the “inside” or “outside” of the arena by driving partially over the respective area. When a robot detects an arena boundary line, it turns away from the line to face a random direction towards the “inside” of the arena. For a boundary line with detected angle θ_1 and the “inside” of the arena towards the direction $\theta_1 + \frac{\pi}{2}$, the robot turns to face a random direction with uniform distribution $U(\theta_1, \theta_1 + \pi)$. A robot pauses its RANDOM BILLIARDS motion and performs obstacle avoidance when it meets an object or another robot. Robots use line-of-sight sensing and communication to detect (and distinguish between) objects and robots that are within a short-range radius ρ_1 and within $\pm \frac{\pi}{3}$ of the heading angle θ_h . When a robot sees an object or robot, it turns away from it until it is

no longer visible within $\theta_1 + \frac{\pi}{3}$. If an object was detected, then the input of Process A is updated as $b_t \leftarrow b_t + 1$.

Voter decision model (VOTE)

In this approach, robots share their individual inference outputs ϵ_r (i.e., outputs of Process A) using explicit communication. Each robot then uses a voter model process to form its opinion λ_r^{app} of absolute object density. Note that we use average opinion (i.e., taking the mean) instead of the majority opinion approach used in [8] (i.e., taking the mode), because we consider a continuous decision rather than a discrete decision task, and the sample size the robots use at each time step is very small. It would often not be possible for a robot to find a unique mode among its available samples, so the performance of the majority model would be poor.

For explicit communication, each robot r_n maintains and shares a matrix ϵ as follows:

$$\epsilon = \begin{pmatrix} r_1 & \epsilon_{r_1} & t_{r_1} \\ r_2 & \epsilon_{r_2} & t_{r_2} \\ \vdots & \vdots & \vdots \\ r_m & \epsilon_{r_m} & t_{r_m} \end{pmatrix}, \tag{4}$$

where r_j is the robot ID, ϵ_{r_j} is the most up-to-date ϵ_r value known for robot r_j by robot r_n , and t_{r_j} is the time stamp of the known ϵ_{r_j} . At each time step, robot r_n updates its own ϵ_{r_n} and t_{r_n} according to Eq. 1 and then sends its matrix ϵ to its current neighbors. If a robot receives a matrix that contains a higher t_{r_j} than its current entry for that robot, it updates its row for r_j accordingly. Also, if $t - t_{r_j} < k_t^{\text{max}}$ for any t_{r_j} entry, the robot removes the corresponding row. In this way, the ϵ of robot r_n always contains the most up-to-date ϵ_r values for its peers that r_n has seen within the time window k_t .

At each time step, robot r decides λ_r^{app} by randomly selecting one ϵ_{r_j} entry from its matrix. In other words,

$$\lambda_r^{\text{app}} = \epsilon_{x \sim U(r_1, r_m)} \in \epsilon \tag{5}$$

such that opinion λ^{app} of absolute density is the result of a voter decision model.

Mean decision model (MEAN)

In this approach, robots also share their individual inference outputs ϵ_r using matrices ϵ , as defined in Eq. 4.

At each time step, robot r decides λ_r^{app} by averaging the ϵ_{r_j} entries in its matrix. In other words,

$$\lambda_r^{\text{app}} = \overline{\epsilon_{r_j}} \in \epsilon \tag{6}$$

such that opinion λ^{app} of absolute density is the result of a mean decision model.

Stigmergy approach (STIG): Process B

In this approach, robots influence the collective opinion via the inputs of Process A—the number of objects detected (b_t)—not the outputs of Process A.

This approach uses stigmergic (indirect) communication through artificial pheromones—i.e., cues left in the environment

that are observable by the robots within a certain range. (In real robots, a stigmergic approach could be implemented using communication-enabled smart blocks, such as those in [39].) Robots use the pheromones deposited in the environment to reach a consensus about the apparent density in the arena.

When a robot detects an object within short-range radius ρ_1 , it counts the object and deposits pheromone at the object location. Robots can detect and differentiate pheromone sources within long-range radius ρ_2 , which allows them to count objects already found by their peers in a much larger vicinity than that in which they can detect objects directly. Each robot counts sensed pheromone sources the same as sensed objects b_t (see Eq. 3 of Process A). Robots also understand that pheromones are cues that the immediate area has already been explored by another robot, and therefore, they turn away to search for new unexplored areas.

Each robot can sense a pheromone in any direction and can sense the relative direction and distance of its source. When a robot senses a pheromone source within long-range radius $\rho_2 - \delta_2$ and in a direction close to that of its heading, it turns away from the source to a random direction according to Algorithm 1.

Algorithm 1: Pheromone reaction

Input:

- θ_x // Randomized angle, see Eq. 7
- \vec{s} // Vector from robot to pheromone source
- \vec{h} // Vector of robot heading
- t_s // Time since last pheromone reaction

$\theta_{sh} \leftarrow$ angle between \vec{s} and \vec{h}

while $|\theta_{sh}| \leq |\theta_x|$ **and** $(t_s \leq 50$ **or** $t_s \geq 250)$ **do**
 [robot turns heading in the $\text{SIGN}(\theta_x)$ direction ;
 move forward ;

The angle θ_x is selected randomly once every 200 time steps according to the following uniform distributions:

$$\theta_{x \sim} \begin{cases} U\left(\frac{\pi}{6}, \frac{\pi}{2}\right), & \text{if } \lfloor \frac{t}{200} \rfloor \text{ is even} \\ U\left(-\frac{\pi}{2}, -\frac{\pi}{6}\right), & \text{otherwise} \end{cases} \tag{7}$$

If a robot is simultaneously within range of multiple pheromone sources, it logs them in a list with an arbitrary order and reacts to the first pheromone source in its list for that time step. After a robot reacts to a given pheromone source, it continues moving forward until it encounters a different pheromone, an object, or a boundary line.

Recall that robots have already influenced the collective opinion via the inputs of Process A. Therefore, at each time step, robot r simply takes its own inference output ϵ_r as its opinion λ_r^{app} of absolute density. In other words,

$$\lambda_r^{\text{app}} = \epsilon_r. \tag{8}$$

Hierarchical approach (HIER): Process B

In this approach, robots influence the collective opinion via the inputs (not the outputs) of Process A, and only the robot occupying

the dynamic leadership position, i.e., the MNS-brain robot r , performs Process A.

The HIER approach is based on the existing MNS concept, which is a general framework for constructing and reconstructing self-organizing hierarchy [4]. Under the MNS framework, robots can self-organize a dynamic ad hoc control network in which robots temporarily and interchangeably occupy certain positions in a leadership hierarchy, including an MNS-brain position (i.e., highest hierarchy position) [5,37,40,41]. In an MNS control network, each robot communicates only with its direct neighbors to prevent the type of bottleneck that would occur at the communication hub in a fully centralized system. According to task specifications and system constraints, sensor information can be merged as it is passed upstream, control information can be unmerged as it is passed downstream, and the balance of individual versus collective behaviors can be actively managed. This flexibility can be used to reduce or eliminate the potential for bottlenecks throughout the hierarchical network (see Fig. 2).

In this study, we used the MNS implementation of [5], in which camera-equipped unmanned aerial vehicles (UAVs) are responsible for sensing the relative positions and orientations of the ground robots for the purpose of keeping the robot formation together during sweeping. (Note that UAVs in the HIER approach in this study cannot sense objects directly so do not increase the total sensing range available in the approach. If we were using ground robots that were capable of sensing each other's relative positions and orientations, the UAVs could be removed, and this removal would not have an impact on the collective perception results.) We used the sweeping technique of [40] and applied it to the task of collective perception.

In the motion routine of Process B of the HIER approach, robots self-organize into a roughly linear formation to sweep the environment (for details, see [40]). In the HIER approach, the MNS-brain robot r detects arena boundary lines and reacts to them using a deterministic process. The dimensions of the MNS's collective ground robot sensor range, calculated by the MNS based on the number of ground robots in the formation, are used as parameters in the deterministic process. The MNS reactively sweeps the unknown environment using standard back-and-forth boustrophedonic motions [42]. The reactive boustrophedonic motions are based on the MNS's knowledge of its own dimensions such that the collective sensor ranges of the ground robots can cover the environment as completely as possible.

As it sweeps the arena, the MNS-brain robot sends control information downstream to maintain the formation (for full details, see [43]). Each robot in the MNS receives motion instructions from its parent that include the targeted relative linear velocity \mathbf{v} and angular velocity $\boldsymbol{\omega}$ as well as the current orientation quaternion \mathbf{q}_p , following [43]. To send motion instructions, each parent in the MNS senses its child's displacement \mathbf{d}_t and relative orientation \mathbf{q}_p , determines the new targeted values based on the most recent motion instructions it received from its own parent, and calculates motion instructions for its child as follows [40]:

$$\mathbf{v} = k_1 \left(\frac{\mathbf{d}_{t+1} - \mathbf{d}_t}{\|\mathbf{d}_{t+1} - \mathbf{d}_t\|} \right), \quad \boldsymbol{\omega} = k_2 \cdot \left\| f(\mathbf{q}_{t+1}^{-1} \times \mathbf{q}_t) \right\|, \quad (9)$$

where k_1 and k_2 are speed constants and function $f(x)$ converts a quaternion to an Euler angle. In the HIER approach, when a ground robot detects an object within short-range radius ρ_1

and within $\pm \frac{\pi}{3}$ of the heading angle θ_h , it temporarily ignores the motion instructions received from its parent to circumvent the object (in a predefined arc trajectory relative to its current position) until the object is no longer within $\theta_h \pm \frac{\pi}{3}$. Complementarily, if the parent of a ground robot detects that its child is behind another ground robot within $\rho_1 + \delta_1$, then the parent will temporarily ask the child to stop moving until the two ground robots are no longer within $\rho_1 + \delta_1$ of each other. The child will accept the request as long as it detects an object within short-range radius ρ_1 .

In the HIER approach, ground robots essentially act as temporary remote sensors of the MNS-brain robot r . Each child robot sends its sensor readings upstream to the MNS-brain robot r via its parent, which calculates one inference output ϵ_r for the whole MNS. Note that, to reduce the per-step simulation time needed by CPU and GPU solvers, some of the upstream and downstream data transfers within the MNS are simulated as a single-step rather than multi-step process. In this study, the approaches are assessed according to the simulation steps, not the per-step time. This implementation strategy negligibly impacts the number of simulation steps needed to collect and pass information (adds a maximum of 1 step), so it does not undermine the analysis of the reported results.

At each time step, the MNS-brain robot r takes its own inference output ϵ_r as its opinion λ_r^{app} of absolute density, i.e.,

$$\lambda_r^{\text{app}} = \epsilon_r. \quad (10)$$

Simulation setup

The experiments were conducted in the ARGoS simulator [44], with robot models implemented using an existing plugin [45,46]. The experiments were conducted with the kinematics of small differential-drive ground robots based on the extended e-puck robot [47–49] and, for the HIER approach, of quad-camera UAVs based on the S-drone quadrotor [50]. For more implementation details of all four approaches and the experimental setup in ARGoS, please see the open-source code repository (https://github.com/BlueDiamond07/Collective_perception).

In all approaches and all setups, only ground robots have the capability to directly sense objects. In all setups except scalability, each approach has 8 ground robots. All robots in all approaches have the same average linear velocity (7.5 cm/s). The experiments begin when the robots start to sweep the arena and end after 50,000 time steps (2,000 s).

Tuning parameters

Regarding the parameters in Eq. 1, in the MEAN and VOTE approaches, the view area ν of robot r is based on the sensor range of onboard object sensing, and in the STIG approach, ν is based on the sensor range of the onboard pheromone sensing. In the HIER approach, ν is based on the maximum bounds of the combined sensor ranges of all ground robots. Note that in all approaches, all ground robots have the same sensor range for detecting objects (short-range radius ρ_1). The time window used in the experiments is $L_t^{\text{max}} = 1,000$ time steps. In other words, in all approaches, a robot's memory of a sensing input lasts for 1,000 time steps (40 s). Parameter P was tuned separately for each approach during a manual testing phase to reduce observable bias in the output of Eq. 1 in each approach. To prevent a reliance on prior knowledge, the tuning phase was not adjusted to specific density conditions. Tuning P is intended

to compensate for the stochasticity of the motion routines in the decentralized approaches, relative to the deterministic sweep of the MNS approach. In the HIER approach, the motion trajectory of the MNS-brain robot is deterministic, so $P = 1$. For each fully decentralized approach, we tuned P through a trial-and-error testing phase to obtain the highest observable performance from Eq. 1 for each approach. After the empirical tuning phase, we set P as follows: voter model $P = 0.48$, mean model $P = 0.55$, and stigmergy $P = 1$. Note that P could, in principle, be optimized in all approaches (see Discussion).

Variations

The experimental variations are as follows. In the basic experiments, we varied the stochastic fields \mathcal{R} by distributing 50, 100, 200, or 300 small objects ($5 \times 5 \times 5 \text{ cm}^3$ cubes) uniformly randomly in a square $6 \times 6 \text{ m}^2$ arena, but with a minimum distance of 20 cm between object center points such that the robots can always pass between them. (In all variations, the objects are rotated in 2D around their center points according to angles defined uniformly randomly.) The true densities for these fields are $\lambda^{\text{true}} = 1.38, 2.7, 5.5, \text{ and } 8.3$ objects/ m^2 , respectively. For accuracy under time-invariant \mathcal{R} , we tested three variations with different true densities.

Basic uniform environments can provide a performance benchmark, but it is also important to test stochastic fields \mathcal{R} in which the data have significantly nonuniform spatiotemporal features, such as spatial heterogeneity (high variance between a few subregions). We therefore ran two types of variations of the basic experiments: temporal and spatial.

For temporal variation, we tested accuracy under time-varying \mathcal{R} , with both minor and major density changes that occur both frequently and infrequently. Specifically, we tested fields that fluctuate between $\lambda^{\text{true}} = 1.38$ and either 2.7 or 8.3 objects/ m^2 over two different rates of change (either 40 s or 400 s), for a total of four possible variations.

For spatial variation, we tested accuracy under spatially non-uniform \mathcal{R} with high variance between subregions. Specifically, we tested fields in which 100 objects are distributed according to different bivariate unimodal and bimodal skew-normal probability distributions. The distributions were generated by defining the standard skew-normal shape parameter α uniformly randomly for x and y (separately) in the unimodal and bimodal distributions, in addition to, for the bimodal distributions, defining the location parameter ξ across the arena dimensions uniformly randomly for x and y (separately). For the bimodal distributions, α and ξ were defined separately for each peak, and each ξ was constrained to half of the arena. Note that the minimum distance between objects was maintained in these setups. In the environments generated from these distributions, the mean density of the whole arena is $\overline{\lambda^{\text{true}}} = 2.7$ objects/ m^2 , but the distribution of λ^{true} is of course nonuniform, and large portions of the arenas do not contain any objects (see Fig. 3).

When testing scalability and fault tolerance, all fields are uniform and time-invariant with $\lambda^{\text{true}} = 2.7$ (i.e., 100 objects in a $6 \times 6 \text{ m}^2$ arena). In the experiments testing scalability, we varied the number of ground robots (4, 8, 12, 16, 20, or 24 ground robots). In the experiments testing fault tolerance, we varied the percentage of ground robots that arbitrarily fail (0%, 25%, 50%, or 75%). Failed ground robots continue their motion

routines, communication, and calculations as normal but experience sensor failure such that they cannot directly count objects—i.e., in Eq. 3, they always add $n = 0$ to $b_t = b_t + n$ as the number of objects detected. Note that, in the STIG approach, this implies that a failed robot never deposits an artificial pheromone but can still detect pheromones left by other robots.

We conducted 10 runs per variation, except for the spatially nonuniform \mathcal{R} experiments with bimodal skew-normal distributions, for which we conducted 20 runs because of the increased complexity in the environment.

Analysis

Recall that λ^{true} is the true density of the environment and that robots produce the values ϵ_r (i.e., individual inferences that robot r makes about the density in its own field of view) and λ_r^{app} (i.e., the opinion of robot r on the apparent density in the whole arena based on the collective influence of all robots). We report the experimental data in Tables S1 to S5 in the Supplementary Materials, and in an open-access data repository (<https://doi.org/10.5281/zenodo.7244384>). We report all data, but given that the time window for sampling inference in the experiments is $k_t^{\text{max}} = 1,000$, we only considered time steps 1,000 (40 s) and later when making assessments about performance.

We assessed the perception accuracy of the approaches according to the error of the robot opinions on apparent density λ^{app} with respect to true density λ^{true} . The overall error (i.e., incorporating both bias and variance) is measured by the mean squared error (MSE) of the opinion λ_r^{app} of robots r over time:

$$\text{MSE}(\lambda_{rt}^{\text{app}}) = \frac{1}{nm} \sum_{r=1}^n \sum_{t=1}^m (\lambda_t^{\text{true}} - \lambda_{rt}^{\text{app}})^2. \quad (11)$$

The instantaneous error is measured by $\text{MSE}(\lambda_{rt}^{\text{app}})$, calculated as in Eq. 11, but without the terms related to t .

Spatially nonuniform environments

For the spatially nonuniform environments, error is measured by the root MSE (RMSE) of the opinion $\text{RMSE}(\lambda_{rt}^{\text{app}}) = \sqrt{\text{MSE}(\lambda_{rt}^{\text{app}})}$ instead of by MSE, because RMSE is better for assessing performance when the errors are large. The measured error in these experiments is high because the true density λ^{true} varies substantially across the arena, but the error was measured according to the mean of λ^{true} , not the distribution of λ^{true} . We used the mean of λ^{true} for the error calculation because the fully decentralized approaches lack any localization or relative positioning capabilities and therefore cannot spatially coordinate their samples among multiple robots in a way that would enable direct comparison with the λ^{true} distribution.

We make the assumption that, in spatially nonuniform environments, the goal of collective perception will be either (a) to estimate the density distribution λ^{true} in a given instance of an environment or (b) to estimate the overall probability distribution of λ^{true} in a class of environments. To fulfill this goal in practice, additional methods would need to be implemented to aggregate the robots' samples and, e.g., perform probability distribution fitting. These additional layers are beyond the scope of this study, but we assessed the similarity of the opinion distribution of each approach to the actual density distribution in

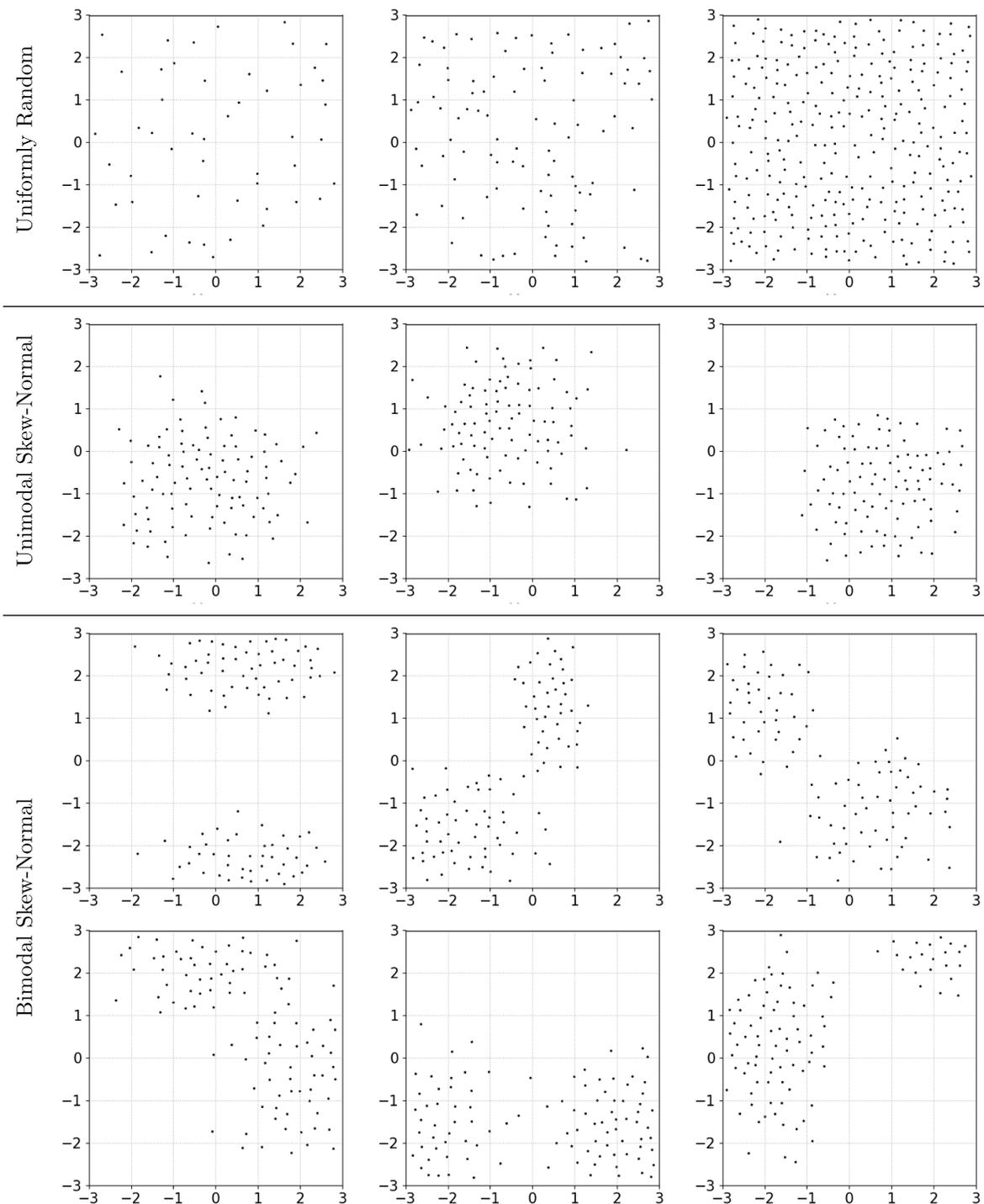


Fig. 3. Example spatial distributions of objects in the arena (in meters) for different experimental variations. The basic time-invariant and time-varying experiments, as well as the scalability and fault-tolerance experiments, used uniformly random distributions. The spatially nonuniform experiments used bivariate skew-normal distributions (both unimodal and bimodal).

the environment as an indication of how feasible various distribution estimation methods might be.

We estimated the λ^{true} distribution in the environment by taking a set of 50,000 density measurements λ^{REF} of the reference samples (i.e., objects) using rectilinear windows R of

variable area, where the xy positions of the vertices of R are defined uniformly randomly, but with R having minimum length and width of 1.5 m. To make a more direct comparison between the density distributions of the reference samples λ^{REF} and the distributions of opinions λ_r^{app} of the four approaches,

we calculated $\text{RMSE}(\lambda^{\text{REF}})$ according to the mean density $\overline{\lambda^{\text{true}}}$. All estimates were then sorted according to RMSE, and we calculated the mutual information (MI) between the $\text{RMSE}(\lambda_r^{\text{app}})$ of the four approaches and the reference $\text{RMSE}(\lambda^{\text{REF}})$. (Specifically, we calculated MI between continuous datasets based on entropy estimation from k -nearest neighbor distances [51,52], using $k = 3$ or 500 neighbors.)

Results

The results show that, under both time-invariant and time-varying stochastic fields \mathcal{R} , the HIER approach has higher accuracy, more consistent accuracy, and faster reaction times than the fully decentralized benchmark approaches.

Under all the tested time-invariant fields \mathcal{R} (see Fig. 4), the HIER approach shows minimal error with only minor spikes, and stable estimates are reached in less than 50 s. The MEAN and VOTE approaches are somewhat less accurate but converge

very quickly and rather stably—when error does spike, the spikes are moderately high and last for less than 200 s. The VOTE approach is less accurate than the MEAN approach and also less consistent (i.e., shows more variance in error across trials). The STIG approach is approximately as accurate and consistent as the VOTE approach, but it converges much more slowly and at higher $\text{MSE}(\lambda_r^{\text{app}})$. Overall, it shows more variance in the error over time. Across all approaches, time-invariant fields with higher true densities λ^{true} seem to be more challenging than those with lower λ^{true} . This is reasonable because the objects being detected while perceiving density also create physical obstructions that the robots must avoid or circumnavigate, which implies more interference and therefore more uncertainty originating during the sweeping method \mathcal{P} . In the HIER approach, however, the difference in accuracy between lower and higher λ^{true} is extremely small.

Under time-varying fields \mathcal{R} , the HIER approach is the only approach that consistently reaches very low error after all temporal shifts. Under slow fluctuations (see Fig. 5B and D), each

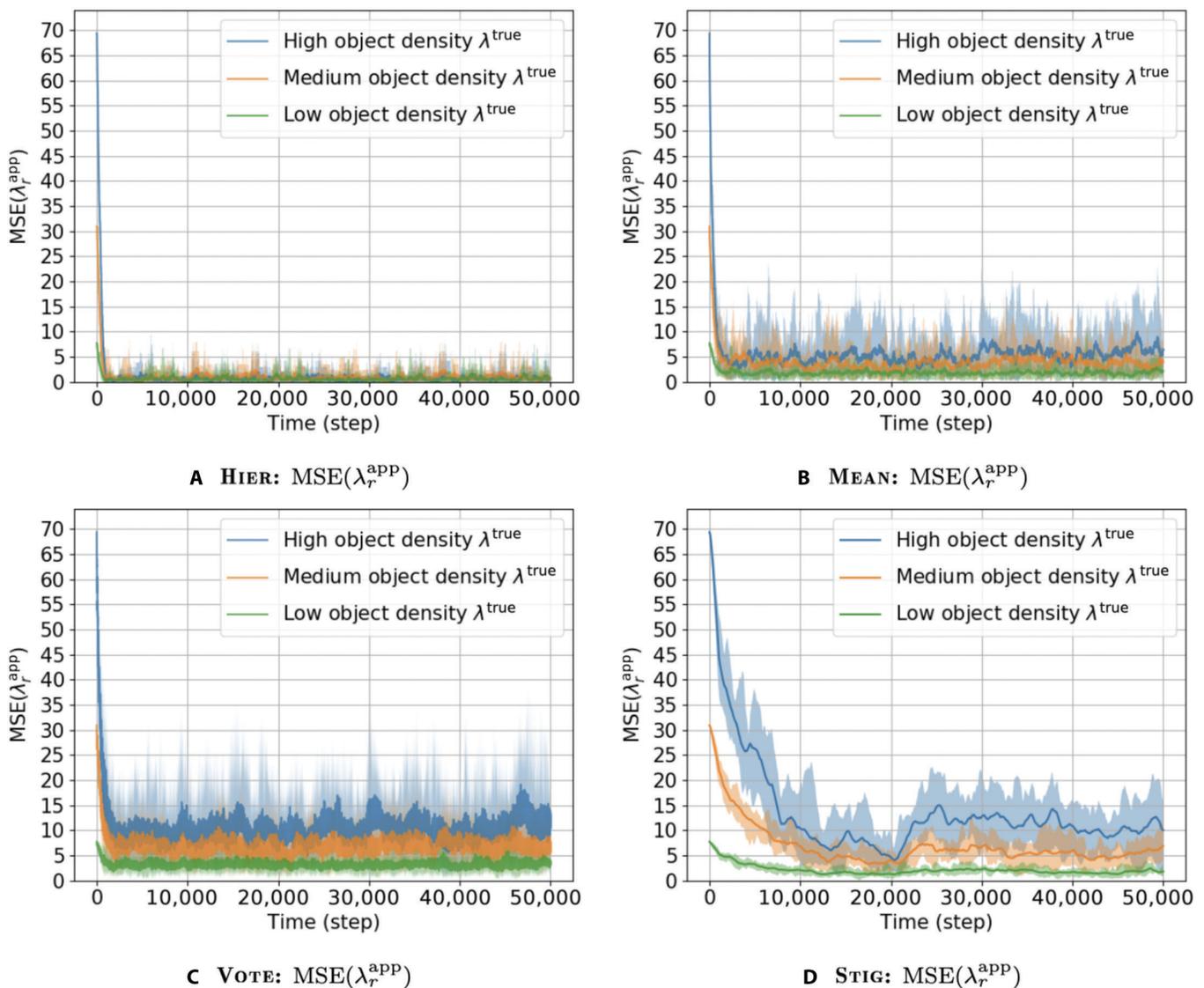


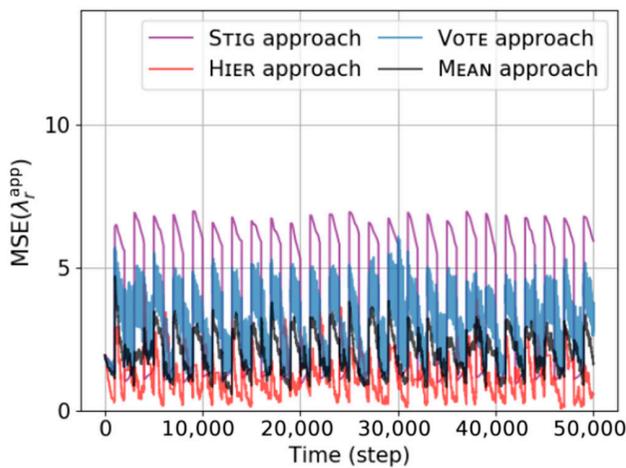
Fig. 4. Error under time-invariant fields \mathcal{R} . (A to D) Lines show mean and shaded areas show minimum and maximum of $\text{MSE}(\lambda_r^{\text{app}})$ of all runs for three variations of true density λ^{true} (objects/m²): low density $\lambda^{\text{true}} = 2.7$, medium density $\lambda^{\text{true}} = 5.5$, and high density $\lambda^{\text{true}} = 8.3$.

approach is able to return to its lowest respective error after a shift to a lower density λ^{true} . In these periods of lowest respective error, the HIER, MEAN, and STIG approaches reach a similar minimal error, while the VOTE approach converges on a slightly higher error. However, after a shift to a higher density λ^{true} , in all approaches the reaction is slower and error is higher (compared to a shift to lower λ^{true}).

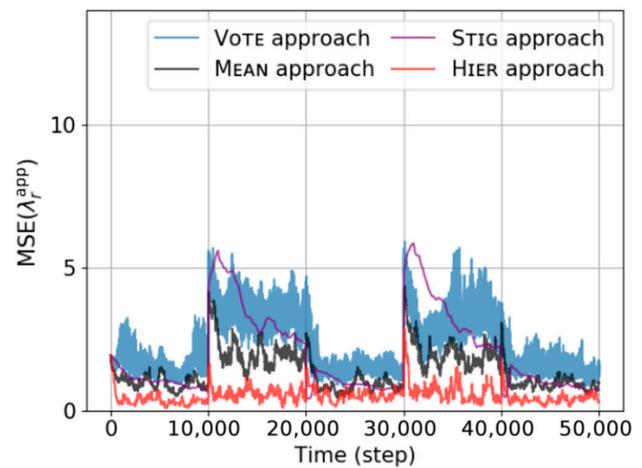
Under time-varying fields, when the fully decentralized approaches are able to converge, all three show noticeably more error than the HIER approach. The HIER approach also reacts much more quickly than the other approaches under all variations of time-varying stochastic fields. The MEAN and VOTE approaches converge relatively quickly, but the STIG approach converges relatively slowly, and under major fluctuations. It cannot converge within 400 s (see Fig. 5D). Under fast fluctuations, these patterns are exacerbated (see Fig. 5A and C). The MEAN and VOTE approaches have barely enough time to

converge under fast, minor fluctuations and cannot converge at all under fast, major fluctuations. The STIG approach never has sufficient time to come close to convergence after a shift to higher λ^{true} . When the fully decentralized approaches do converge under fast fluctuations, their accuracy is the same as that under slow fluctuations.

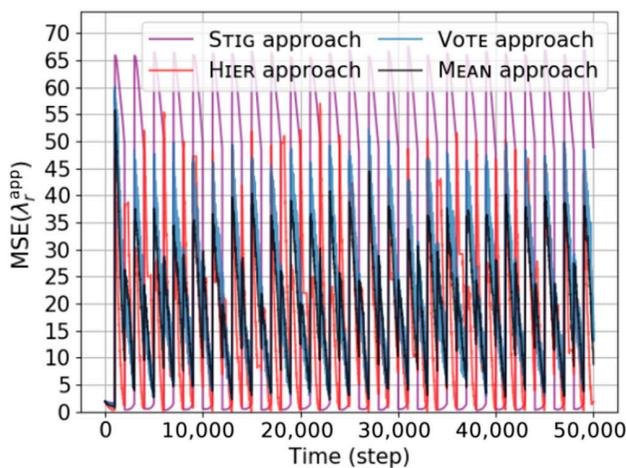
Overall, the HIER approach shows the lowest error (see Tables S1 to S5 for details). Under time-invariant fields (see Fig. 4), mean error in the HIER approach is very minor, almost negligible. $\text{MSE}(\lambda_r^{\text{app}})$ does spike occasionally in some runs but generally recovers within 20 to 40 s. The spikes in error in the HIER approach are much less frequent than in the other approaches, and the largest spikes are comparatively minor. Under time-varying fields (see Fig. 5), error in the HIER approach is also overwhelmingly much lower than in the other approaches, under all rates and magnitudes of field fluctuations. Note that, under time-varying fields with major fluctuations, immediately



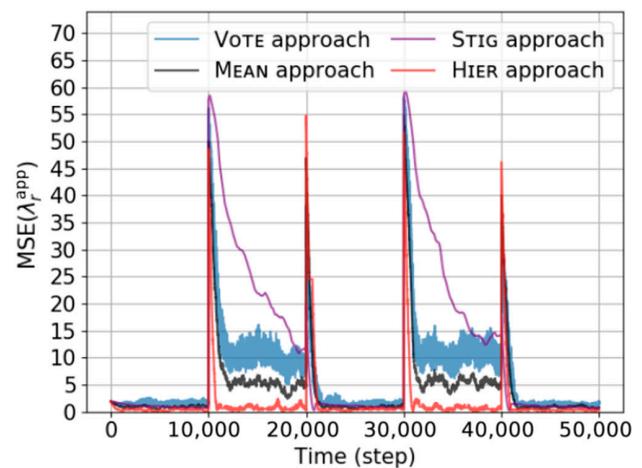
A Fast, minor fluctuations: field alternating every 40 s between densities that differ slightly: $\lambda^{\text{true}} = 1.38$ or 2.7 objects/m².



B Slow, minor fluctuations: field alternating every 400 s between densities that differ slightly: $\lambda^{\text{true}} = 1.38$ or 2.7 objects/m².



C Fast, major fluctuations: field alternating every 40 s between densities that differ substantially: $\lambda^{\text{true}} = 1.38$ or 8.3 objects/m².



D Slow, major fluctuations: field alternating every 400 s between densities that differ substantially: $\lambda^{\text{true}} = 1.38$ or 8.3 objects/m².

Fig. 5. Error under time-varying fields \mathcal{R} . (A to D) Mean $\text{MSE}(\lambda_r^{\text{app}})$ of all runs for four temporal variations of fluctuating true density λ^{true} .

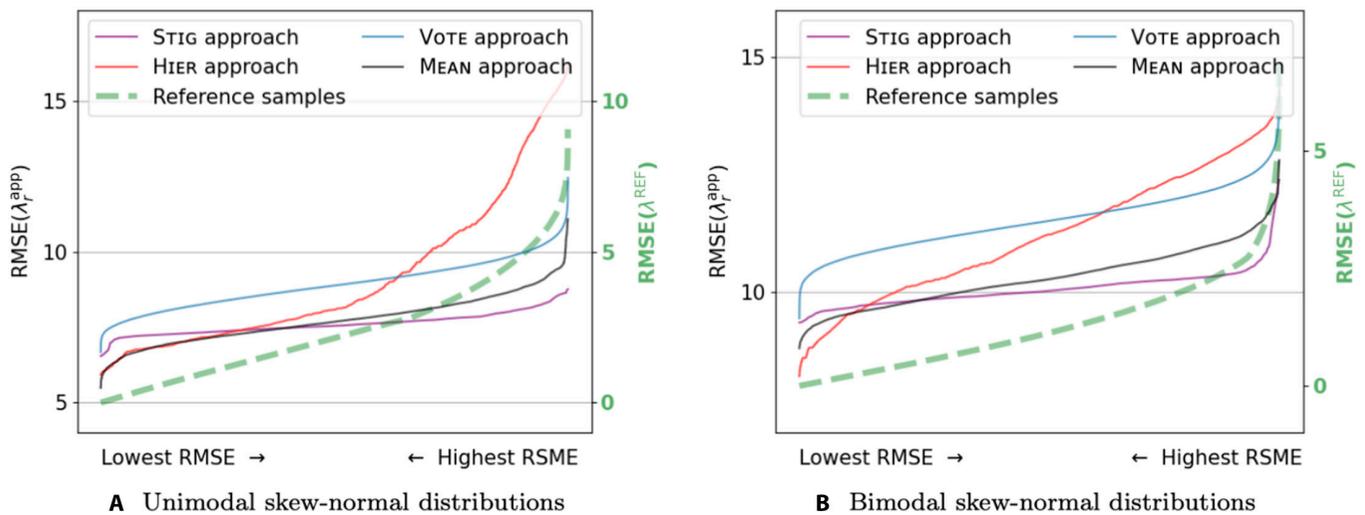


Fig. 6. Error under spatially nonuniform fields \mathcal{R} compared to the reference samples. (A and B) Mean $\text{RMSE}(\lambda_r^{\text{app}})$ of the four approaches (left-hand y axis) compared to the mean $\text{RMSE}(\lambda_r^{\text{REF}})$ from the reference samples (right-hand y axis) of all runs for different fields with spatially nonuniform true density λ^{true} . All RMSE values were calculated according to the mean density $\lambda^{\text{true}} = 2.7$ (for the details of the analysis approach, see the “Spatially nonuniform environments” section). Estimates are sorted on the x axis according to RMSE. To make visual comparison easier, $\text{RMSE}(\lambda_r^{\text{REF}})$ has been realigned on the y axis in relation to $\text{RMSE}(\lambda_r^{\text{app}})$; see right- and left-hand y-axis labels. For the estimated MI, see Table S5.

after some of the shifts in the stochastic field, error in the HIER approach spikes more aggressively than in some of the other approaches. However, the HIER approach recovers very quickly after these spikes and, in most cases, returns quickly to a lower error than that shown by the other approaches. Overall, we can conclude that the HIER approach has higher and more consistent accuracy and reacts accurately and more quickly than the three benchmark approaches.

Spatially nonuniform environments

The results show that, under the spatially nonuniform fields \mathcal{R} , the distributions of opinions produced by all four approaches have some basic similarity to the true density distributions in terms of shape (see Fig. 6).

For both unimodal and bimodal environments, the mean $\text{RMSE}(\lambda_r^{\text{app}})$ of the VOTE and MEAN approaches shows extremely similar distribution shapes, with the VOTE approach having higher error overall. In other words, the VOTE approach’s distribution curve is effectively a displaced (higher-error) duplication of the MEAN approach’s distribution curve. For both environment types, the STIG approach’s distribution curve is much flatter than those of the VOTE and MEAN approaches, but otherwise, it is fairly similar to the distribution curve of the MEAN approach, in terms of both shape features and overall error. For both environment types, the HIER approach’s distribution curve is much steeper than those of the three decentralized approaches, and in the unimodal environments, it is also quite different from the other three in terms of shape features.

In comparison to the reference samples, all four approaches have much higher overall error but show a fair amount of similarity in their distribution curves. For both unimodal and bimodal environments, the three decentralized approaches have approximately the same similarity between their mean $\text{RMSE}(\lambda_r^{\text{app}})$ distributions and the mean $\text{RMSE}(\lambda_r^{\text{REF}})$ distribution. All three have (a) a sharp increase at the lowest end that the reference distribution does not have at all, (b) a sharp increase at the highest end that is much more gradual in the reference distribution,

and (c) a gradual increase in the middle that is flatter than the increase of the reference distribution. Also, in all three approaches, the distance between the lowest and highest values of each respective distribution is noticeably smaller than that of the reference distribution. Of the three, the STIG approach’s distribution is the least similar to the reference distribution in terms of rate of increase and overall shape. In comparison to the three decentralized approaches, the HIER approach’s distribution in both environment types is more similar to the reference distribution in terms of the difference between the lowest and highest values. The HIER approach’s distribution also seems to potentially be more similar in terms of overall shape, but this is difficult to assess from the information available, so we consider this finding to be inconclusive overall.

The estimated MI between the mean $\text{RMSE}(\lambda_r^{\text{app}})$ distribution and the mean $\text{RMSE}(\lambda_r^{\text{REF}})$ distribution is also very similar for all approaches (see Table S5). Overall, the HIER approach is slightly worse in terms of MI than the other three, while the STIG approach is slightly better. The STIG approach therefore has the highest dependency (based on MI) but the least curve similarity.

Overall, the accuracy differences between the opinion distribution curves of the four approaches are inconclusive, but all four approaches display some dependency and approximate curve similarity with the reference distribution.

Scalability

In the scalability setup, in the tested group sizes, none of the approaches show a decrease in accuracy as the group size increases (see Fig. 7), implying that the threshold at which inter-robot interference could negatively affect accuracy has not been reached. (Note that this also implies that none of the approaches, including the HIER approach, display a bottleneck at these sizes. For more discussion of bottlenecks in the MNS approach, see [43].) Rather, we see moderate decreases in performance in some approaches as the group size decreases. The accuracy difference between group sizes in the MEAN approach is the most noticeable but is still relatively minor;

however, its accuracy is also much less consistent in the smaller group sizes, displaying larger and more frequent spikes. The accuracy difference between group sizes in the HIER and STIG approaches is extremely minor, and the VOTE approach shows no difference between group sizes. Overall, all four approaches show good scalability of accuracy, which stays the same or improves as the group size increases. The VOTE approach also shows relatively good resiliency to smaller group sizes. However, it is important to note that although the accuracy of the VOTE approach does not worsen with smaller groups, it shows more error in all group sizes than the HIER or STIG approaches in their worst group size.

Overall, the HIER approach shows less error than all other approaches for all tested group sizes and shows extremely minor (almost no) error in the largest group size of 24 ground robots.

Fault tolerance

In the fault tolerance setup, all four approaches show a noticeable decrease in accuracy as a greater percentage of robots fail (see Fig. 8). This is expected because the fault tolerance setup

is quite challenging—failed robots continue to move and communicate but always record that they have directly detected zero objects, thereby introducing extra bias during sampling method \mathcal{G} .

Some of the fault tolerance variants can be considered to have matching respective scalability variants—e.g., the fault tolerance condition of 50% failure leaves the swarm with 4 failing and 4 correctly working ground robots, which is a match to the scalability condition of 4 ground robots. When comparing accuracy under 50% failure to the matching scalability variant, the HIER and MEAN approaches show only slightly more error, while the VOTE approach shows a more noticeable increase in error. In contrast, the STIG approach shows no noticeable difference. All four approaches show a similarly substantial increase in error from 50% to 75% failure. Under 75% failure (the highest failure rate), the HIER approach shows noticeably less error than the MEAN and VOTE approaches. Under 75% failure, the HIER approach shows slightly less error than the STIG approach, but with less consistency (noticeably higher spikes).

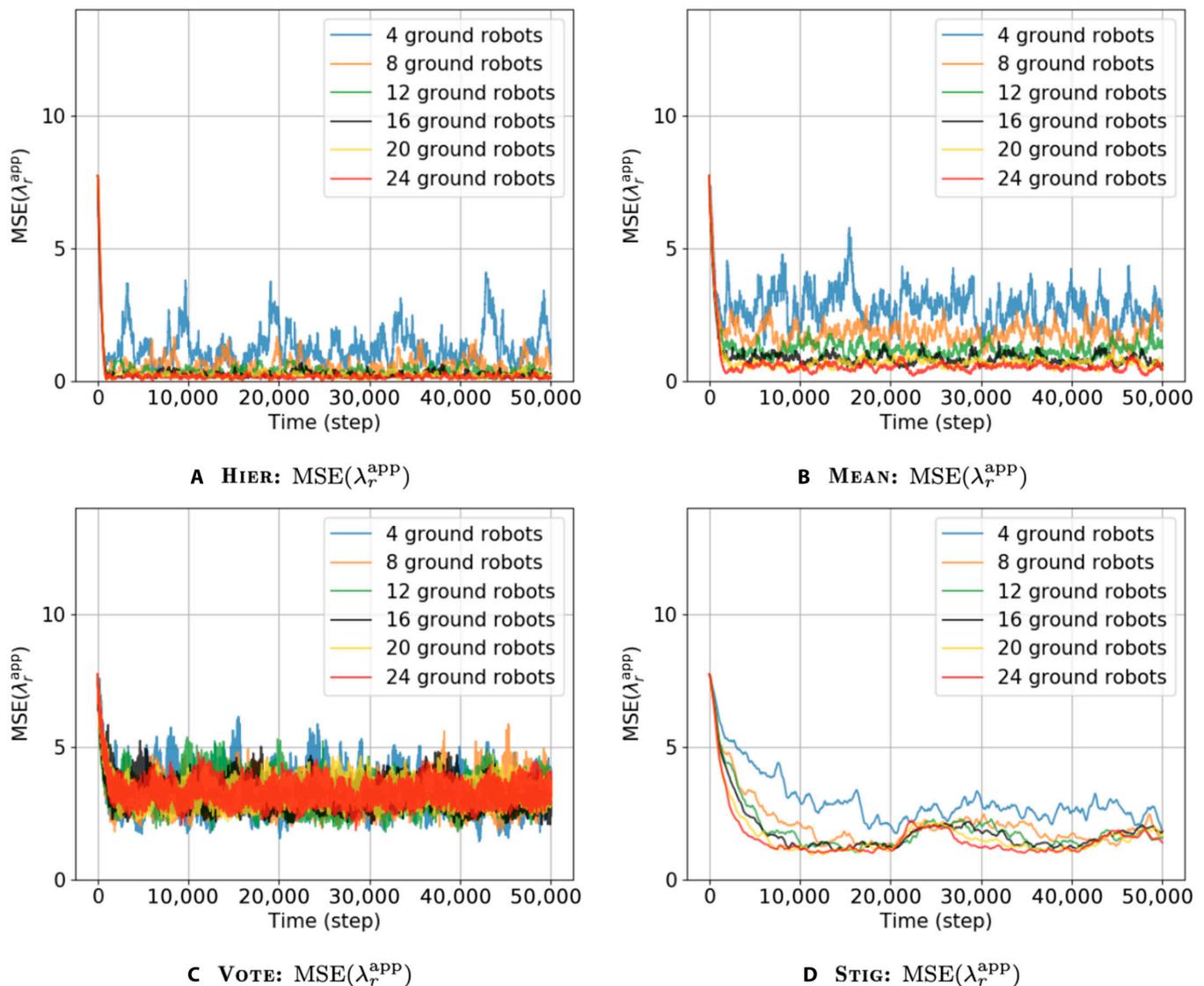


Fig. 7. Error when testing scalability. (A to D) Mean $MSE(\lambda_r^{app})$ of all runs for six variations of group size: 4, 8, 12, 16, 20, or 24.

Overall, the STIG approach shows the least increase in error from no failure to 75% failure, with the other approaches showing similar amounts of increase. However, the HIER approach has a lower mean error than the STIG approach under no, 25%, and 50% failure and a comparable (but still slightly lower) mean error under 75% failure. The MEAN and VOTE approaches show more error than the other two approaches under all failure rates. Therefore, the HIER approach shows the least overall error in the fault tolerance setups.

Overall, all four approaches are approximately comparable in terms of fault tolerance of accuracy, with the HIER approach having the least overall error and STIG approach showing the most consistency in its tolerance.

Summary

The experimental results can be summarized as follows: (a) the HIER approach shows higher and more consistent accuracy and reacts accurately more quickly than the other three approaches; (b)

the VOTE approach shows the least change in accuracy under changes in group size, but the HIER approach shows the highest overall accuracy under all group sizes; and (c) the STIG approach shows the least change in accuracy under changes in failure rates, but the HIER approach shows the highest overall accuracy under all failure rates.

Discussion

Based on the assertions made in this paper, under conditions without robot failures, the HIER approach should be able to conduct collective perception with very low uncertainty. In the HIER approach, uncertainty originating during the sweeping method \mathcal{P} is expected to be greatly reduced because \mathcal{P} is a deterministic method. Likewise, the uncertainty that can originate at the sampling method \mathcal{T} is expected to be greatly reduced because the deterministic \mathcal{P} is also able to provide a nearly optimal sampling distribution (i.e., very evenly distributed sampling

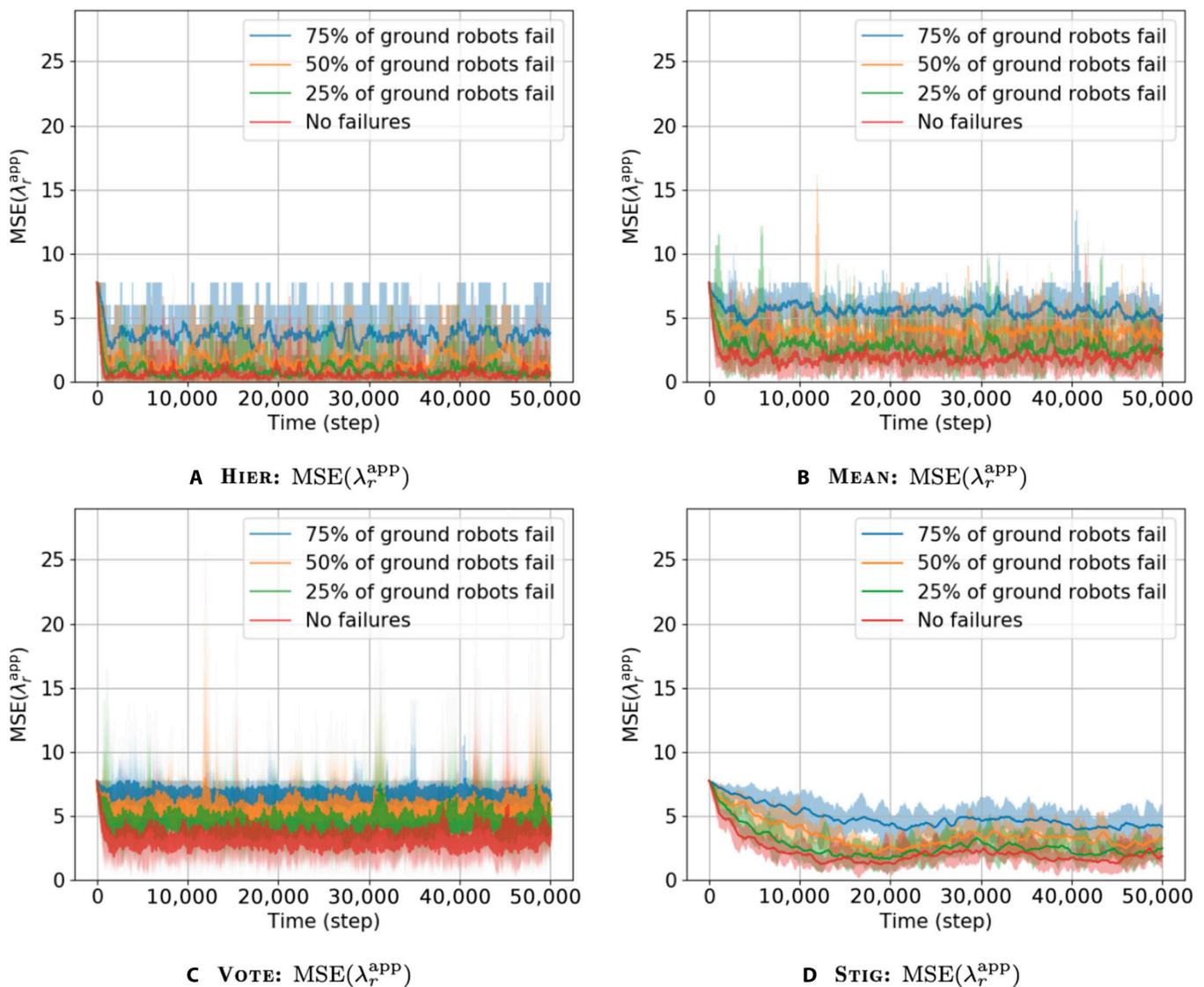


Fig. 8. Error when testing fault tolerance. (A to D) Lines show mean and shaded areas show minimum and maximum of $MSE(\lambda_r^{app})$ of all runs for four variations of arbitrary failures: 0% (0 of 8), 25% (2 of 8), 50% (4 of 8), and 75% (6 of 8) ground robots failing.

sites and a sampling ratio approaching 1; see Fig. 1). (See [40] for demonstration of uniform and complete spatial coverage using the MNS.) Also, because the MNS network is a connected graph and all sensor information is fused at one node in that graph, no uncertainty should originate during decision sampling \mathcal{G} , and uncertainty that can originate during inference ψ , beyond the uncertainty that is inherent without prior knowledge of the stochastic field \mathcal{R} , is expected to be minimal.

When the results of the four approaches can be directly compared with the overall true density λ^{true} , the empirical results align well with these expectations (e.g., see the low error in Fig. 4A). In terms of accuracy, the HIER approach indeed outperforms the others under time-varying and time-invariant stochastic fields. These results therefore support the assertion that self-organizing hierarchy can improve perceptual accuracy by integrating aspects of centralized control that reduce uncertainty.

Under the spatially nonuniform environments, where the results are compared to a reference set of density measurements, the opinion distributions of all four approaches display similarity to the reference distribution, with the performance differences between the four approaches being inconclusive. Overall, the empirical results indicate that distribution estimation (either of the density distribution λ^{true} in a given instance of an environment or of the overall probability distribution of λ^{true} in a class of environments) would in principle be feasible in any of the four approaches, and further study would be required to assess the performance differences between them. However, in practice, fully decentralized approaches to distribution estimation (e.g., probability distribution fitting) would give each robot access either to only a subset of the total samples or to an asynchronously updated opinion on the overall distribution, introducing an extra source of uncertainty. By contrast, the information that would be available for distribution estimation in the HIER approach would be the same as that shown in Fig. 6, without additional sources of uncertainty.

Also based on the assertions of this paper, in setups testing scalability and fault tolerance, the HIER approach should be able to conduct collective perception without any meaningful increase in error, as compared to the fully decentralized approaches. Indeed, in all scalability and fault tolerance setups, the error in the HIER approach does not increase beyond a level comparable to that of any of the other approaches, and in most of these variants, the error of the HIER approach is lower than that of all other approaches. Therefore, the empirical results again align with expectations: the scalability and fault tolerance of the HIER approach can be considered commensurate with those of the other approaches. The results therefore support the assertion that self-organizing hierarchy, despite introducing some aspects of centralized control, can maintain the beneficial aspects of decentralized control.

Poor performance of VOTE approach

Although the voter decision model has been shown to be relatively accurate with discrete best-of- n decision-making (e.g., choosing one of a few color options), this is because it works well at accurately converging on the most common opinion in a group. When dealing with high-variance samples of a continuous stochastic field, it has no mechanism to compensate for bias in the sampling or reduce the variance—it essentially shuffles several highly biased opinions among group members, so the distributions of opinions before and after the voter decision model process are statistically indistinguishable. This notion is

confirmed empirically by the similarity between $\text{MSE}(\epsilon_{rt})$ and $\text{MSE}(\lambda_{rt}^{\text{app}})$, i.e., the collective error in robot opinions before and after the voter decision model process (see Tables S1 to S5).

The VOTE approach results therefore represent roughly what a single robot would perceive on its own, which explains the poor accuracy.

Tuning parameters

Parameter tuning is a difficult process, whether using simulated or real testing. Here, P was tuned during an initial manual testing phase. Results should be interpreted with the understanding that parameter P could in principle be optimized further in every approach but that this optimization could improve only the mean bias of the opinions, not the variance (refer to Eq. 1). It should also be acknowledged that optimization of P would require prior knowledge of the stochastic field \mathcal{R} . For example, if P was optimized to a very high-density field \mathcal{R} , the performance under low density would presumably suffer, and vice versa.

Future work

Future work on collective perception should study the robustness of sampling methods under other types of challenging conditions, such as environments where it is difficult for robots to have good coverage over a spatial area due to large obstacles or boundary irregularity.

More advanced inference methods could be investigated, such as weighting samples by inclusion probability. However, most of these inference methods require some prior knowledge of the stochastic field [35], which might make them unsuitable for deployment-ready self-organized robot systems. General optimized performance could also be investigated, including for situations in which it is not possible to have prior knowledge of the stochastic field. Optimization without such prior knowledge would likely need to be based on online adaptation to or learning of the spatial information being sampled, which could in principle be done in any of the approaches, but would certainly be easier to design in the HIER approach than in the fully decentralized approaches.

It would also be useful to study other types of failures, such as motion control errors (e.g., robots get stuck in corners), odometry errors (e.g., robots believe they have traveled far less distance than they actually have), other sensor errors (e.g., robots believe they are always detecting an object), and full robot shutdown, as well as the timing of such failures and other factors that could exacerbate their impact.

Improvements to the HIER approach

Although it might be simple to design and implement system-wide adaptivity in the HIER approach using the MNS, we have not added such behaviors in this study because it is much more difficult to implement such adaptivity in the fully decentralized approaches. For as fair a comparison as possible, we limited the capabilities of the HIER approach to be similar to those of the fully decentralized approaches. For instance, the fault tolerance results of the HIER approach reflect the lack of adaptability currently implemented—e.g., the MNS-brain robot believes its indirect field of view to be that of eight fully functioning robots, although it only has two fully functioning robots remaining. Minor additions to the approach could be made to allow parents to detect malfunctioning sensor readings from children and substantially improve fault tolerance. This would be more straightforward to implement and design in the HIER approach than in

the fully decentralized approaches because of the aspects of centralization that are integrated into self-organization when using the MNS framework. More broadly, advanced behaviors are much simpler to implement using the self-organizing hierarchy capabilities of the MNS than when using strictly decentralized approaches, so the addition of even more advanced abilities, such as self-awareness, can be considered. The self-reconfiguration capabilities of the MNS [5] would also make it relatively straightforward to implement additional behaviors that help keep the robot formation together when sweeping challenging environments, such as changing formation or temporarily splitting into multiple smaller formations that sweep independently until they have the opportunity to re-merge.

Conclusion

We have identified the sources of uncertainty that are present in the collective perception problem, especially when perceiving an absolute condition without prior knowledge, and detailed why this uncertainty is reduced by using self-organizing hierarchy. We have supported this assertion empirically by showing that a proof-of-concept self-organizing hierarchy approach (based on the MNS framework) is generally more accurate, more consistent, and faster than fully decentralized benchmark approaches. We have also shown that the self-organizing hierarchy approach to collective perception, besides producing more accurate estimates of spatial data, does not suffer substantial scalability or fault tolerance disadvantages compared to fully decentralized benchmark approaches. We have also tested spatially nonuniform environments with high heterogeneity and shown that the self-organizing hierarchy approach is comparable to the other approaches in terms of accuracy compared to reference samples. Therefore, the comparative ease of designing system-wide behaviors under self-organizing hierarchy can be taken advantage of, without a reduction in the performance benefits that are often associated with swarm robotics approaches.

Acknowledgments

We thank Weixu Zhu and Michael Allwright for their advice regarding the external code libraries used in the experiments. **Funding:** This work was supported by the Program of Concerted Research Actions (ARC) of the Université libre de Bruxelles, the Natural Sciences and Engineering Research Council of Canada under Discovery Grants, and the Ontario Trillium Scholarship Program through the University of Ottawa and the Government of Ontario, Canada. M.D. and M.K.H. acknowledge support from the Belgian F.R.S.-FNRS, of which they are a Research Director and a Postdoctoral Researcher, respectively. **Author contributions:** All authors contributed to conceiving the study. A.J. conducted the experiments and collected the data. A.J. and M.K.H. led the experimental design and analysis. M.K.H. led the writing of the manuscript, and all authors read and approved the final version of the manuscript. **Competing interests:** The authors declare that they have no competing interests.

Data Availability

The code used in the experiments is open-source and available on GitHub: https://github.com/BlueDiamond07/Collective_perception. All experimental data collected during the study are open-access and available on Zenodo: <https://doi.org/10.5281/zenodo.7908113>.

Supplementary Materials

Tables S1 to S5

References

1. Khaluf Y, Simoens P, Hamann H. The neglected pieces of designing collective decision-making processes. *Front Robot AI*. 2019;6:16.
2. Dorigo M, Theraulaz G, Trianni V. Reflections on the future of swarm robotics. *Sci Robot*. 2020;5(49):abe4385.
3. Dorigo M, Theraulaz G, Trianni V. Swarm robotics: Past, present, and future [point of view]. *Proc IEEE*. 2021;109(7):1152–1165.
4. Mathews N, Christensen AL, O'Grady R, Mondada F, Dorigo M. Mergeable nervous systems for robots. *Nat Commun*. 2017;8:439.
5. Zhu W, Allwright M, Heinrich MK, Oğuz S, Christensen AL, Dorigo M. Formation control of UAVs and mobile robots using self-organized communication topologies. In: *Swarm Intelligence – Proceedings of ANTS 2020 – Twelfth International Conference, Lecture Notes in Computer Science*, vol. 12421; 2020. p. 306-314.
6. Valentini G, Ferrante E, Dorigo M. The best-of-n problem in robot swarms: Formalization, state of the art, and novel perspectives. *Front Robot AI*. 2017;4:9.
7. Valentini G, Hamann H, Dorigo M. Self-organized collective decision making : the weighted voter model. In: *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS '14) 2014*. p. 45-52.
8. Valentini G, Hamann H, Dorigo M. Collective perception of environmental features in a robot swarm. In: *Swarm Intelligence – Proceedings of ANTS 2016 – Tenth International Conference, Lecture Notes in Computer Science*, vol. 9882; 2016. p. 65-76.
9. Heinrich MK, Wahby M, Dorigo M, Hamann H. *Swarm robotics*. In: Cangelosi A, Asada M, editors. *Cognitive Robotics*. MIT Press; 2022.
10. Yan Z, Jouandeau N, Cherif AA. A survey and analysis of multi-robot coordination. *Int J Adv Robot Syst*. 2013;10(12):399.
11. Sun S, Lin H, Ma J, Li X. Multi-sensor distributed fusion estimation with applications in networked systems: A review paper. *Inform Fusion*. 2017;38:122–134.
12. Rizk Y, Awad M, Tunstel EW. Cooperative heterogeneous multi-robot systems: A survey. *ACM Comput Survey*. 2019;52(2):1–31.
13. Li B, Xian Y, Zhang D, Su J, Hu X, Guo W. Multi-sensor image fusion: A survey of the state of the art. *J Comput Commun*. 2021;9(6):73–108.
14. Strobel V, Castello Ferrer E, Dorigo M. Managing byzantine robots via blockchain technology in a swarm robotics collective decision making scenario. In: *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS '18); 2018*. p. 541-549.
15. Ebert JT, Gauci M, Nagpal R. Multi-feature collective decision making in robot swarms. In: *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS '18); 2018*. (pp. 1711-1719).
16. Shan Q, Mostaghim S. Collective decision making in swarm robotics with distributed bayesian hypothesis testing. In: *Swarm Intelligence – Proceedings of ANTS 2020 – Twelfth International Conference, Lecture Notes in Computer Science*, vol. 12421; 2020. p. 55-67.

17. Bartashevich P, Mostaghim S. Benchmarking collective perception: New task difficulty metrics for collective decision-making. In: Progress in Artificial Intelligence: 19th EPIA Conference on Artificial Intelligence (EPIA 2019), Proceedings, Part I; 2019. p. 699-711.
18. Shan Q, Mostaghim S. Discrete collective estimation in swarm robotics with distributed bayesian belief sharing. *Swarm Intell.* 2021;15(4):377-402.
19. Prasetyo J, De Masi G, Ranjan P, Ferrante E. The best-of-n problem with dynamic site qualities: Achieving adaptability with stubborn individuals. In: Swarm Intelligence – Proceedings of ANTS 2018 – Eleventh International Conference, Lecture Notes in Computer Science, vol. 11172; 2018. p. 239-251.
20. Prasetyo J, De Masi G, Ferrante E. Collective decision making in dynamic environments. *Swarm Intell.* 2019;13(3):217-243.
21. Khaluf Y. Edge detection in static and dynamic environments using robot swarms. In: Proceedings of the 11th International Conference on Self-Adaptive and Self-Organizing Systems (SASO); 2017. p. 81-90.
22. Wahby M, Petzold J, Eschke C, Schmickl T, Hamann H. Collective change detection: Adaptivity to dynamic swarm densities and light conditions in robot swarms. In: Proceedings of ALIFE 2019: The 2019 Conference on Artificial Life; 2019. p. 642-649.
23. Khaluf Y, Allwright M, Rausch I, Simoens P, Dorigo M. Construction task allocation through the collective perception of a dynamic environment. In: Swarm Intelligence – Proceedings of ANTS 2020 – Twelfth International Conference. Lecture Notes in Computer Science, vol. 12421; 2020. p. 82-95.
24. Capitan J, Spaan MT, Merino L, Ollero A. Decentralized multi-robot cooperation with auctioned POMDPs. *Intl J Robot Res.* 2013;32(6):650-671.
25. Mazo M, Speranzon A, Johansson KH, Hu X. Multi-robot tracking of a moving object using directional sensors. In: Proceedings of the 2004 IEEE International Conference on Robotics and Automation (ICRA '04); 2004 Vol. 2, p. 1103-1108.
26. Mirzaei FM, Mourikis AI, Roumeliotis SI. On the performance of multi-robot target tracking. In: Proceedings of the 2007 IEEE International Conference on Robotics and Automation (ICRA '07); 2007. p. 3482-3489.
27. Rodrigues T, Duarte M, Oliveira S, Christensen AL. Beyond onboard sensors in robotic swarms. In: ICAART 2015: Proceedings of the International Conference on Agents and Artificial Intelligence; 2015. Vol. 2, p. 111-118.
28. Stroupe AW, Martin MC, Balch T. Distributed sensor fusion for object position estimation by multi-robot systems. In: Proceedings of the 2001 IEEE International Conference on Robotics and Automation (ICRA '01); 2001. Vol. 2, p. 1092-1098.
29. Sasaoka T, Kimoto I, Kishimoto Y, Takaba K, Nakashima H. Multi-robot SLAM via information fusion extended Kalman filters. *IFAC-PapersOnLine.* 2016;49(22):303-308.
30. Zadorozhny V, Lewis M. Information fusion based on collective intelligence for multi-robot search and rescue missions. In: Proceedings of the: 2013 IEEE 14th International Conference on Mobile Data Management; 2013. Vol. 1, p. 275-278.
31. Czarnetzki S, Rohde C. Handling heterogeneous information sources for multi-robot sensor fusion. In: Proceedings of the IEEE Conference on Multisensor Fusion and Integration; 2010. p. 133-138.
32. Otte M. Collective cognition and sensing in robotic swarms via an emergent group-mind. In: International Symposium on Experimental Robotics, Proceedings; 2016. p. 829-840.
33. Kornienko S, Kornienko O, Constantinescu C, Pradier M, Levi P. Cognitive micro-agents: Individual and collective perception in microrobotic swarm. In: Proceedings of the IJCAI-05 Workshop on Agents in Real-time and Dynamic Environments; 2005. p. 33-42.
34. Giusti A, Nagi J, Gambardella L, Di Caro GA. Cooperative sensing and recognition by a swarm of mobile robots. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2012. p. 551-558.
35. Wang J-F, Stein A, Gao B-B, Ge Y. A review of spatial sampling. *Spatial Stat.* 2012;2:1-14.
36. Hunt ER, Jones S, Hauert S. Testing the limits of pheromone stigmergy in high-density robot swarms. *R Soc Open Sci.* 2019;6(11):190-225.
37. Jamshidpey A, Wahby M, Heinrich MK, Allwright M, Zhu W, Dorigo M. Centralization vs. decentralization in multi-robot coverage: Ground robots under UAV supervision. IRIDIA, Universite Libre de Bruxelles, Brussels, Belgium, Tech. Rep. TR/IRIDIA/2021-008 2021.
38. Comets F, Popov S, Schütz GM, Vachkovskaia M. Billiards in a general domain with random reflections. *Arch Ration Mech Anal.* 2009;191(3):497-537.
39. Allwright M, Zhu W, Dorigo M. An open-source multi-robot construction system. *HardwareX.* 2019;5:e00050.
40. Jamshidpey A, Zhu W, Wahby M, Allwright M, Heinrich MK, Dorigo M. Multi-robot coverage using self-organized networks for central coordination. In: Swarm Intelligence – Proceedings of ANTS 2020 – Twelfth International Conference, Lecture Notes in Computer Science, vol. 12421; 2020. p. 216-228.
41. Zhang Y, Oğuz S, Wang S, Garone E, Wang X, Dorigo M, Heinrich MK. Self-reconfigurable hierarchical frameworks for formation control of robot swarms. *IEEE Trans Cybern.* 2023.
42. Choset H, Pignon P. Coverage path planning: The boustrophedon cellular decomposition. In: Field and Service Robotics. Springer London; 1998.
43. Zhu W, Allwright M, Heinrich MK, Oğuz S, Christensen AL, Dorigo M. Formation control of UAVs and mobile robots using self-organized communication topologies. In: Swarm Intelligence – Proceedings of ANTS 2020 – Twelfth International Conference, Lecture Notes in Computer Science, vol. 12421; 2020. p. 306-314.
44. Pinciroli C, Trianni V, O'Grady R, Pini G, Brutschy A, Brambilla M, Mathews N, Ferrante E, Di Caro G, Ducatelle F, et al. ARGoS: A modular, parallel, multi-engine simulator for multi-robot systems. *Swarm Intell.* 2012;6(4):271-295.
45. Allwright M, Bhalla N, Pinciroli C, Dorigo M. ARGoS plugins for experiments in autonomous construction. IRIDIA, Universite Libre de Bruxelles, Brussels, Belgium, Tech. Rep. TR/IRIDIA/2018-007 2018.
46. Allwright M, Bhalla N, Pinciroli C, Dorigo M. Simulating multi-robot construction in ARGoS. In: Swarm Intelligence – Proceedings of ANTS 2018 – Eleventh International Conference, Lecture Notes in Computer Science, vol. 11172; 2018. p. 188-200.
47. Mondada F, Bonita M, Raemy X, Pugh J, Christopher M, Cinci A. The e-puck, a robot designed for education in engineering. In: Proceedings of the 9th Conference on Autonomous Robot Systems and Competitions; 2009. Vol. 1, p. 59-65.
48. Gutiérrez Á, Campo A, Dorigo M, Donate J, Monasterio-Huelin F, Magdalena L. Open E-puck Range & Bearing

- miniaturized board for local communication in swarm robotics. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '09); 2009. p. 3111-3116.
49. Millard AG, Joyce R, Hilder JA, Fleşeriu C, Newbrook L, Li W, McDaid LJ, Halliday DM. The Pi-puck extension board: A Raspberry Pi interface for the e-puck robot platform. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2017. p. 741-748.
 50. Oğuz S, Heinrich MK, Allwright M, Zhu W, Wahby M, Garone E, Dorigo M. S-drone: An open-source quadrotor for experimentation in swarm robotics. IRIDIA, Université Libre de Bruxelles, Brussels, Belgium, Tech. Rep. TR/IRIDIA/2022-010 2022.
 51. Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. *Phys Rev*. 2011;69:066138.
 52. Ross BC. Mutual information between discrete and continuous data sets. *PLOS ONE*. 2014;9(2):Article e87357.