Bar Ilan University

Vision-Based Collective Motion: A Locust Inspired Reductionist Approach

David L. Krongauz

Submitted in partial fulfillment of the requirements for the Master's Degree in the
Computer Science Department
Bar Ilan University

Ramat-Gan, Israel 2023

This work was carried out under the supervision of Prof. Gal A. Kaminka from the Department of Computer Science, Bar-Ilan University and Amir Ayali from the School of Zoology and Sagol School of Neuroscience, Tel-Aviv University.

Contents

List of Figures

Abstract

Naturally occurring collective motion is an omnipresent, fascinating phenomenon that has been long studied within different scientific fields. Swarming individuals are believed to aggregate, coordinate and move utilizing only local social cues projected by conspecifics in their vicinity. Major theoretical studies of this phenomenon assume perfect information availability, where agents rely on complete and exact knowledge of inter-agent distances and velocities. However, the sensory modalities that are responsible for the acquisition of the environmental information in nature are often ignored. Vision plays a central role in animal perception, and in many cases of collective motion it is assumed to be the sole source of social information.

We investigate a vision-based collective motion model, i.e., a model relying on visually available parameters only, inspired by the case study of locust marching bands. We address two major challenges: the estimation of distance and velocity, and visual occlusions. We consider and compare three strategies an agent can use to interpret partially occluded visual information. *In silico* experiments were conducted in two different frameworks: the first simulating physical entities in a square arena, and the second simulating simplified two-dimensional agents moving under various geometrical conditions. The results show the feasibility of our model and its three occlusion handling approaches for overcoming occlusion. While all the models display convergence to an ordered state, they differ in the respective computational requirements they demand from an agent: the least computationally demanding approach, in which no object detection is taking place, shows slower convergence to order. This is mostly apparent in geometrically constrained environments, which may hint as to the requirements from biological swarming species in natural settings.

1 Introduction and Background

The phenomenon of collective motion in natural systems has fascinated humanity for years. The display of hundreds or in some cases thousands of collectively moving animals is a captivating event and have intrigued scientists from a diversity of disciplines including animal behavior, physics, biophysics, social sciences and computer science, all trying to better understand the fascinating emergence of flocking, swarming and schooling in groups of autonomous agents interacting between them strictly locally in a decentralized manner [55, 56, 53, 62, 79, 76, 69, 49, 34, 78, 54, 80].

A prime example of flocking behavior in nature is the emergence of marching locust hopper bands. Some species of locust show behavioral transformation, shifting from a solitary phase to gregarious [77]. Locusts in their solitary state actively avoid other locusts. In contrast, in their gregarious phase, locusts display general more active behavior and have a predisposition to march in immense bands (while in their nymph stage), later turning into huge flying swarms in their adulthood, consisting of thousands of individuals [63, 71, 28].

Various attempts have been made in order to capture the mechanisms behind the collective behavior displayed by different species, many of them in the area of theoretical agent-based modelling. The leading paradigm is that collective motion results from repeated local (myopic) interactions between individual swarm members (see, e.g., [80]). Upon perceiving its local physical and social environments, each individual acts using a motion-control procedure [35]. This approach is often modeled by '*self-propelled particles*' (SPP), where simulated agents navigate while being subjected to mutual steering forces, caused by interactions with their neighbors [4]. These interactions feed into each agent's decision-making scheme, changing its motion [79, 42, 27]. Under the appropriate conditions, this allows the generation of collectively ordered motion, where units move in approximately a single spatial direction, similar to flocking[1] animals [3, 13, 5]. Seminal models of collective motion are presented in Section 1.1. However, these models do not address the problem of information acquisition by an individual, which in nature is mainly relied on the visual sensory modality.

Recently, several studies have attempted to elucidate the problem of sensory perception in collective behavior modeling [19, 8, 60]. Some of these investigate the role of the visual field in self-organizing systems and vision-based navigation, both in natural [64, 74, 19, 58]

[1]We use *flocking* and *collective motion* interchangeably. Flocking describes the emergence of a common movement direction for all the agents out of uncorrelated initial movements. Once the coherent motion is achieved on a group level, it should persist in time. Common heading direction with a narrow spread thus emerges and defines this asymptotic flocking state.

and simulated environments, either *in silico* [43, 44, 45, 19, 73, 8, 60] or implemented via robots [50, 11]. In Section 1.2 we will elaborate on the challenges that vision-based perception carries and the ways the models mention above face them.

The main goal of this thesis is to investigate the influence of limitations of visual perception upon the process of flocking and the resultant asymptotic collective motion. We introduce the specific aspects we study, question and the inspiration for the solution in Section 1.3.

1.1 Existing Models Ignoring Vision

An early and influential model of collective motion was the *Boids* model, published by Reynolds in 1987 [62]. The model's prime objective was to simulate the visual appearance of flocking bird-like objects, termed boids. The model comprises three types of interaction rules that are run by each agent, in parallel, at every time step: firstly, separation (repulsion) from nearby neighbors. Secondly, alignment of the *boids'* heading, implemented by steering towards the average heading of the neighbors. Finally, cohesion (attraction) is carried out by steering towards the average position of the local flock members (center of mass) as illustrated in Fig.1. Neighbors are classified based on a dedicated radius for each type of interaction, i.e., if the distance from the center of one *boid* to another is equal or smaller than some radius $\mathbf{r}$, then they are considered to be neighbors. Changing the model's parameters like the radius or the steering influence, results in different swarming behaviors. The Boids model assumes that the agents are lacking the notion of volume, and thus can pass through each other.

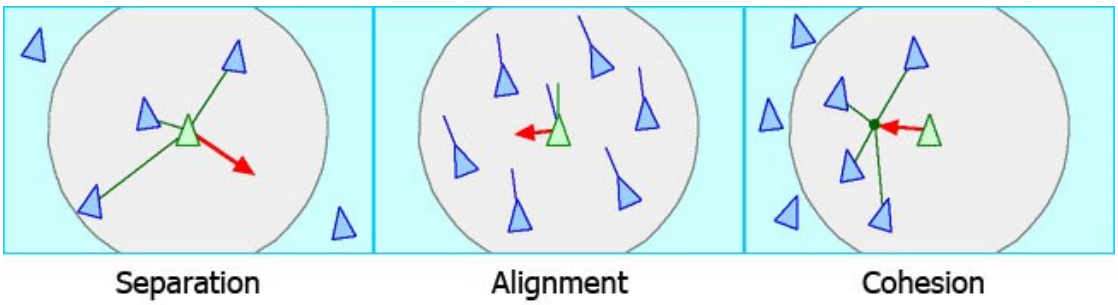


Figure 1: Boids model's three interaction rules dictating the behavior of each agent: (a) Separation – steer away neighboring *boids* to avoid crowding. (b) Alignment – steer towards the average heading direction of local flock-mates to create flocking. (c) Cohesion – Steer towards the average position of the neighbors. From *www.red3d.com/cwr/boids/* by Reynolds.

Vicsek *et al.* [79] introduced a simplistic model for collective motion in 1995. Based on ideas from statistical physics, it modelled the behavior of SPP's in the presence of perturbations. This work later inspired a wide range of follow-up works and models extending the original [17, 31, 7, 1]. The model interprets quantitatively the behaviors of huge flocks in the presence of perturbations.

The perturbations represent various factors, both stochastic and deterministic, involved in the emergence of collective motion of swarms. In practice, these perturbations were incorporated by adding a random angle to the average direction of neighboring agents. The agents move with a fixed absolute velocity and change their direction of motion, at every time step, to be the average direction of others within a fixed predefined distance. Due to its relative simplicity, using this model one can simulate large numbers of agents in each simulation, reconstructing the natural phenomenon where sometimes thousands of flock-mates are involved.

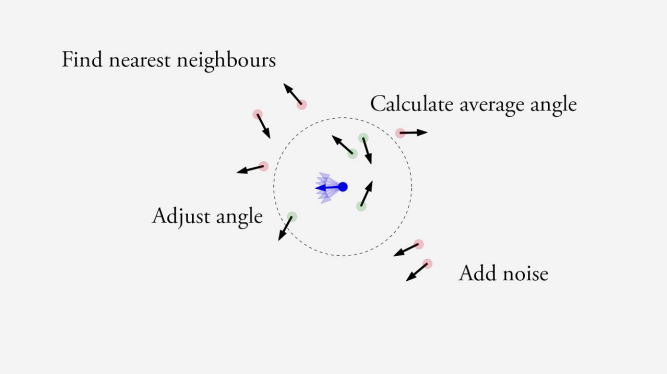


Figure 2: Vicsek Model schematic. Snapshot from https://tinyurl.com/yc4wrruy

1.2 Limitations Introduced by Visual Perception

Collective motion is dependent on the swarm members' individual capacity for acquiring complete, reliable, and real-time information of its social environment. All the models described above were based on the assumption that each individual gets its world information from global knowledge borrowing from special relativity, the lab frame of the experiment. For example, in the Vicsek model, every agent is given the exact velocities of his neighbors when carrying out the average at each time step. Furthermore, in the Boids model not only the velocities but even the exact coordinates of neighboring agents are necessary in the implementation of 'cohesion' and 'separation'. This is an idealization of natural capabilities for species relying on visual perception as the dominant

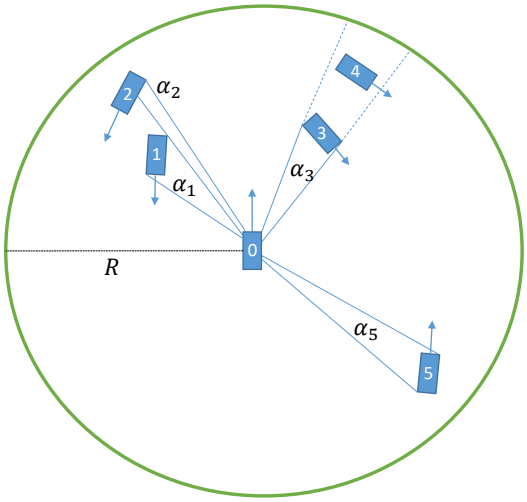


Figure 3: Occlusions within a vision radius. Focal agent (0) has vision radius R and 5 neighbors: 1 partially occludes 2. 4 is fully occluded by 3. 5 is not occluding or is being occluded by any other neighbor. The α_i angles correspond to the angles each neighbor subtends on the focal agent.

sensory modality, since it ignores two major challenges imposed by vision: First, partial occlusions of swarm members by others necessarily challenge the models that assume perfect knowledge of the social environment. Second, many species that lack stereoscopic vision, such as locusts, cannot reliably estimate distances.

1.2.1 Occlusions

When flocking of animals in Nature is observed, typically some of the neighboring subjects are occluding each other, (see Fig. 3). When the numbers of agents are higher, the occlusions become more frequent. Occlusions bear significant effect upon the accuracy of the perception of the environment, and we analyze their entailed influences on the swarming process.

Work by Da Silva *et al.* [70] titled 'Boids that See' exploited self-occlusion, meaning the assumption that an agent rear sight is occluded by itself, for purposes of computational performance improvement thus allowing the simulation of very large swarms. This work

was a step towards point-of-view analysis of an agent's perception. However, only self-occlusion was considered in the paper, while occlusions between neighbors are disregarded. Soria et al. [73] explored the effect of a limited visual field on the performance of the Reynolds flocking algorithm [62]. Rosenthal et al. [64] investigated the importance of the visual field (e.g., "angular area") in collective evasion maneuvers manifested in schooling fish. Nonetheless, they also employed globally-acquired information, including exact metric and topological distances between the swarming fish. Another extension of the Reynolds model [62] is presented by Kunz et al. [43]. Although their model deals with obstruction of further distanced neighbors, it does not account for possible partial obstruction, i.e., simplifying the perception to only seeing the closest neighbors of each agent. Moreover, neighbors' exact distances and headings are employed to calculate the motion of each agent.

1.2.2 Visual Assessment of Neighbors' Velocities and Distances

Estimation of distance based on visual perception is possible, but makes certain requirements of the sensory abilities of the observing entity. Some animals can use their binocular, stereoscopic vision capabilities for calculating distance [52]. others, lacking the anatomical-physiological required mechanisms, can try making up for this limitation by moving while obtaining the brief visual inputs, thus creating an impression of depth (in computer science, this is called generating *structure from motion*).

A key mechanism that is common in animal motion and navigation is the *optical flow*, which is in mathematical terms the vectorial difference between two frames an individual "sees" [33]. However, as described in [68], optical flow cannot be directly used to estimate distance. As seen in Fig. 4, when the visual field is covered by a single sensor there exists a multifold ambiguity in assessing the velocity-vector of a neighboring object, since infinitely many velocity-vectors are projected on the same perceived vector. Another related mechanism is sensitivity to looming objects in the vicinity [37, 32].

Several recent works tackled the issue of visual based navigation and synchronization. The work of Collignon et al. [19], though not dealing with flocking, studies zebrafish-inspired agents that move in a 2D environment. To solve the problem of range estimation, they introduce higher dimensional solid angles representing the original 3D characteristics of the simulated fish.

The vision-based model of collective motion presented by Bastien et al. [8] and extended to obstacle avoidance by Qi et al. [60] offers a different approach for perception modeling. They introduce specifically designed complex integro-differential operators

that, depending on the visual input, govern an agent's motion, which in turn creates various collective behaviors. To avoid the problem of ranging, this work and others restricted themselves to a special case of disk-shaped agents [50, 11]. However, the natural morphologies of locusts and fish are elongated, and very different from circular disks.

Wang et al. [82] demonstrated implicit cooperation between robotic swarm members using visual inputs. Utilizing an elaborate scheme of positioning poles and sets of LED lights, positioned in known distances over the poles, the robots are able to estimate the relative positioning, velocity, state and other features of their neighbors. Using the acquired information, the robots exhibit, various swarming behaviors, including flocking (collective motion), formations and foraging.

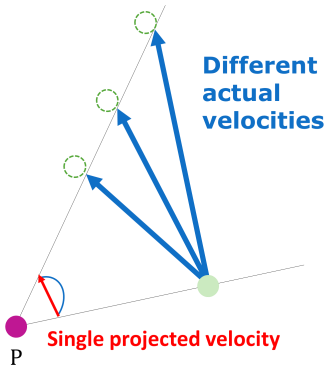


Figure 4: Ambiguity in velocity estimation when relying only on visual inputs.

1.3 Vision-Based Collective Motion: Our Approach

In contrast to most previous works, this thesis presents a reductionist approach, which on the one hand stays true to the elongated morphology of natural animals, but on the other reduces the visual inputs to the bare minimum perceivable in two dimensions (i.e., no vertical angles, no visual markers at different heights). We draw our inspiration from the marching locust nymphs, a quintessential example of collective motion ([4], and references within). In natural conditions, swarms constitute vast numbers of individuals, demonstrating coordinated and synchronized mass-motion [6, 22, 83]. This model also lends from controlled laboratory experiments that were successful in providing valuable insights into the mechanism of collective motion, and specifically, the interactions between

the individual and the group (e.g., Ariel et al. [5], Bazazi et al. [9], Knebel et al. [40]). Based on these and other reports, it is largely accepted that vision is a key sensory modality underlying the local interactions among a group of locusts. The above-mentioned challenges are exceptionally relevant in the locust model. Previous work has established that the Pause-and-Go motion scheme is central to the repeated decision-making underlying the behavior of locusts in a swarm [4, 5, 41], i.e., a representation of the local environment, utilized for deciding if and at what direction to move, is constructed by the locusts when standing. This makes both, distance and velocity assessment and occlusions, major hindrances.

From a signal processing perspective, vision is a complex, multifaceted process. Without dedicated processing, it can result in information loss during acquisition and significant computational errors introduced in the perception process. We should, therefore, limit and simplify complex properties by adopting a reductionist approach.

This thesis addresses the challenge of occlusions, whether partial or complete, detailed in Section 2. We investigate different general methods for handling occlusions and their consequences. Later, elaborated in Section 3, we propose a model of vision-based collective motion inspired by the flocking behavior of migrating locusts. We use the geometrical characteristics of visual perception to present a control algorithm solely based on the information accessible via the agent's visual field. Section 4 concludes the thesis.

2 Addressing Occlusions: Three General Approaches

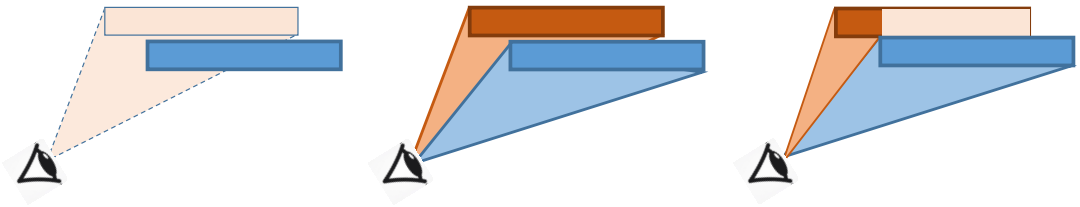
Occlusions are an inherent part of the visual modality in natural systems. Animals often swarm in dense flocks [10], and thus conspecifics located closer to the observing animal are inevitably blocking, partially or entirely, the animals standing behind them.

Figure 5: The visual social environment from the point of view of the individual locust in a swarm; a composite image following Bleichman et al. 2022 [12]. Original image by Inga Petelski.

Complete and partial occlusion of neighbors not only reduces the information available to the focal agent but can also introduce large estimation errors. Neighbors that are completely occluded are not taken into account in the collective motion model.

2.1 Models of Occlusions

In accounting for visual occlusions, we posit there are three general strategies that may be taken (illustrated in Fig. 6). Suppose the focal agent may be able to recognize peers and thus differentiate between entirely-visible individuals and parts (that is, not complete individuals). This allows it to ignore partially-visible neighbors (Fig. 6-a). It may also be able to cognitively extrapolate parts to a whole, inferring the position and orientation of the partially-occluded peer from its visible parts (Fig. 6-b). Or, as an alternative, it may perceive any visible part of a neighbor as a distinct whole individual. These strategies place very different requirements on the cognitive-computational processes of visual perception in the focal agent, as discussed in detail below.



(a) OMID (b) COMPLID (c) PARTID

Figure 6: Occlusion methods. (a) OMID – partially occluded neighbor (orange) is omitted from the field of view. (b) COMPLID – Orange neighbor is completed from the seen segment. (c) PARTID – partially seen segment is regarded as a neighbor.

2.1.1 Approach 1: Omission of the Occluded (OMID)

The first approach disregards any visual information originating in partially occluded agents (see Fig. 6a). This requires the animal to possess a dedicated *peer recognition* mechanism, i.e., be able to recognize fully-imaged conspecifics (and ignore anything else). Mechanisms of selective attention in visual perception are known to exist in humans and are achieved in the human brain in multiple stages of perception [47, 16]. Neurobiological studies showed the existence of selective attention, as well as target detecting abilities amid cluttered backgrounds, also in insects' visual neurons [26, 81]. Such visual processing mechanisms may, in principle, be utilized during collective motion.

It is not known whether locust visual perception mechanisms are able to recognize peers. Experiments reported in [12] show that an individual locust responds by walking when exposed to visual images composed of randomly-moving dots projected via computer screens to both eyes. As the dots are positioned randomly and do not mimic the shape nor the colors of locust nymphs, these results seem to indicate that the motion is triggered in the individual sans a dedicated peer recognition mechanism.

2.1.2 Approach 2: Completion of the Occluded (COMPLID)

In the second approach, partially occluded agents are "completed" as if they are fully visible to the observer. In other words, a neighbor that has even the smallest visible segment from the focal agent's perspective would be treated as if no occlusion is present when processing its visually extractable information. COMPLID utilizes *peer recognition* as in OMID. In addition, it requires that the agents will be able to assess the obscured part of their neighbors (if needed) based on the visible part of a neighbor. This completion

assumes an agent's visual extrapolation that reconstructs neighbors' outlines using their visible features.

Completing partially visible targets obscured by other objects is a long-studied process in visual perception. The filling-in of details and image regions partially obscured by interceding objects [38, 72] is an established neurophysiological process that gives the organism an ability to identify a complete form based upon observed parts of the contour and is described by the term "visual completion" [14]. This mechanism produces an internal representation called "illusory contour", which extrapolates the physical stimulus to the full geometrical shape of the object [51, 46, 48]. Visual completion of occluded objects has been shown in varied and phylogenetically distant species: birds, fishes, cephalopods, bees, etc., and is accepted as one of the fundamental components of vision in nature [51, 21, 36]. Hence, the second approach is to extrapolate partially occluded neighbors to complete the full image.

2.1.3 Approach 3: Every Part is a Full Agent (PARTID)

The third approach treats all visual stimuli related to a neighbor as if they represent a full-body conspecific. Contrary to OMID and COMPLID, in this approach, agents lack *peer recognition* abilities. Instead, the visual field is divided into segments, where each segment contains the same optical flow vectors. The agent assumes that each segment represents a different neighbor. In other words, any visual information is considered complete at face value without any additional interpretation. Hence, apart from the ability to accurately extract optical flow vectors, no further advanced visual perception mechanisms are required. Since the optical flow is essentially the vectorial difference between two consecutive frames and does not consist of any form of object recognition by itself, PARTID would be the least computationally demanding approach if implemented in real life.

PARTID takes inspiration from biological mechanisms, in which an organism performs an action based on visual stimuli originating from an object that is not recognized. For example, locusts have a pair of visually-sensitive neurons that encode looming stimuli and cause the locust to produce escape behaviors [30]. The visual stimuli directly affect the behavior of the individual without passing through object recognition mechanisms [12].

2.1.4 Requirements for Implementation in Robots

Implementing this approach in robotic swarms, one needs to relate to the machine vision requirements posed by this approach and the computational complexity that follows.

Since agents sift out incomplete agents present in the field of vision, they have to be capable of object detection abilities. There are several main methods to achieve that: one is *frame differencing*, in which the presence of moving objects in a frame is found by calculating the difference between two successive frames. Frame differencing method has strong adaptability for a range of dynamic environments, but it also shows errors in obtaining a complete outline of moving objects, as a result, the accuracy level of detection of moving objects is very low [61].

The second method is Optical Flow, mentioned already beforehand, which can be used directly for object detection. It involves calculating the image's optical flow field and doing clustering processing according to the optical flow distribution characteristics of the image. This method can get the complete movement information of an object, and it is useful for detecting the moving object from the background with 85% accuracy [18], but this method has a few disadvantages, including a large quantity of calculations and sensitivity to noise, which make it not appropriate for real-time object detection and tracking especially on simple robots often used in swarm robot research.

Another method worth mentioning is *Point detectors*, in which interest points in an image are picked and tracked based on whether they have an expressive texture in their respective localities [65]. A useful interest point is one which is invariant to changes in illumination and camera viewpoint.

The above methods do not rely on object features, meaning that they can potentially be utilized in order to implement the PARTID method that does not seek a full outline detection or completion but merely detects clusters of pixels and differentiates them from the environment. On the other hand, OMID and COMPLID require either full outline detection or visual completion, both of which are implemented via features detected on the object. A common approach that unifies various detection methods is *Edge-based features*, where the edge map of an object is extracted and later analyzed in order to identify the features of the object in terms of edges [59]. Using edges as features is advantageous over other features due to various reasons. Edges are extremely invariant to changes in illumination conditions and variations in object's colors and textures. The object boundaries are represented well, and the data is analyzed efficiently in the large spatial extent of the images [75].

2.2 Experimental Setup

The performance of the three occlusion approaches was tested and compared in a simulation. Vicsek model was used as the decision-making algorithm on top of which

the occlusion approaches are applied. At each time step, each focal agent calculates its desired velocity based on the average velocity of all its neighbors located within a radius. Additionally, Vicsek Model is simulated in its original form (without the effect of occlusions), as a baseline for comparing the influence of the approaches.

2.2.1 Simulation Environment

The experiments were conducted using ARGoS, a physics-based simulator dedicated to simulating robotic swarms [57]. It can simulate large-scale swarms of robots, while enabling the user to define the robots' parameters and features. Agents were defined as three-dimensional nonholonomic entities of rectangular shape. The kinematics of simulation are as follows: agents can change their speed and direction according to their new desired velocity. The change is not instantaneous, i.e., the agent changes its orientation with angular speed proportional to the angular difference between the current heading and the desired velocity direction. Fig. 7 shows a screenshot of the simulation.

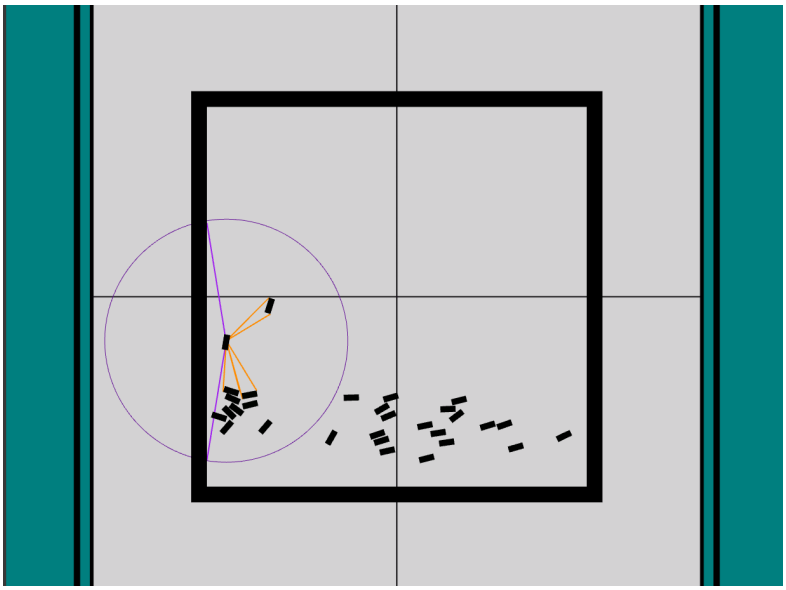


Figure 7: Snapshot of the ARGoS simulator. Pictured is a swarm of 30 agents during a simulation run. The circle represent the vision radius of an arbitrarily chosen focal agent. Its seen neighbors are denoted by the orange lines pointing to the visible edge with respect to each agent. The purple lines denote the extreme visible edge of an obstacle, a wall in this case.

Additionally, the agents simulate a basic collision avoidance algorithm with respect to obstacles and other agents. Essentially, whenever an agent is moving towards an

obstacle or neighbor, it activates a collision avoidance protocol that either lowers its speed or changes heading to a direction perpendicular to the imminent collision threat. For example, an agent moving towards a wall would steer away from the wall, reflecting its original heading direction relative to the plane of the wall.

Experiments are conducted in a square-shaped arena with immobile walls serving as the boundaries of the arena. Each experiment is initialized by uniformly distributing agents with random locations and headings. The exact initial state is determined by a seed number used in the random function.

2.2.2 Measured Order Parameter

The ideal flocking is a situation in which all agents are synchronously moving in the same direction. To compare the approaches, we use the *polarization* measure of order, denoted ϕ [79, 23]. It is defined by

$$\phi = \frac{1}{N}\left|\sum_{i}^{N}\frac{v_i}{|v_i|}\right|, \tag{1}$$

where N is the population size, and $v_i, |v_i|$ correspond to the ith agent's velocity and speed. By taking the absolute value of the average heading, we get a scalar value representing at any given time the degree of global order in the system. This parameter measures the degree of global alignment by averaging the normalized velocities of the agents (i.e., headings). From now onward, it will be defined as the order parameter. At maximal chaotic state, ϕ has the value 0, while for a fully flocked system, with all agents moving with an identical heading, it approaches the value of 1.

2.3 First Set of Experiments

We investigate the influence of the occlusion handling approaches on the flocking of a swarm, flocking using the Vicsek Model original model (termed *Principal* in the figures). A series of 50 independent experiments were performed with 50 different random seeds. In each, a swarm of varying population size was run for 3000 simulation steps. The results are shown in Fig. 8.

The panes are ordered by the population size of the swarm (increasing left to right) and by the vision radius in each series of experiments, with the first row representing the smaller radius and the second row the larger one. In each pane, the Y axis is the order parameter representing the level of alignment, and the X axis is the time measured in simulation frames.

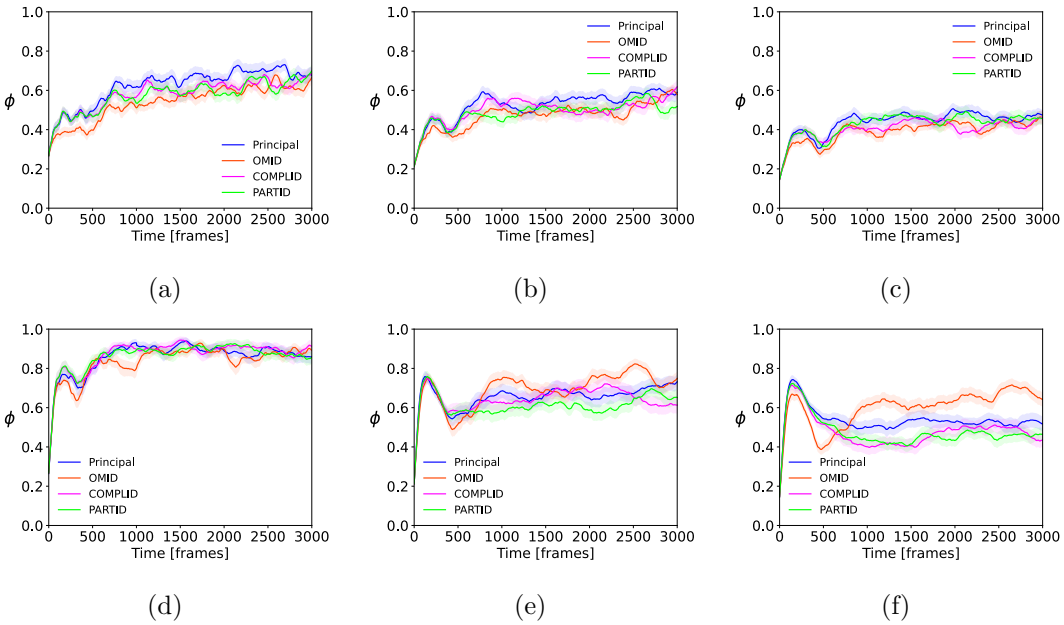


Figure 8: Time evolution of the order parameter for all occlusion approach with the Vicsek baseline for swarms of different population size. (a),(b),(c) correspond to populations of 10, 20, 30 agents, respectively, for vision radius of 3 [BL]. (d),(e),(f) correspond to a vision radius of 8 [BL].

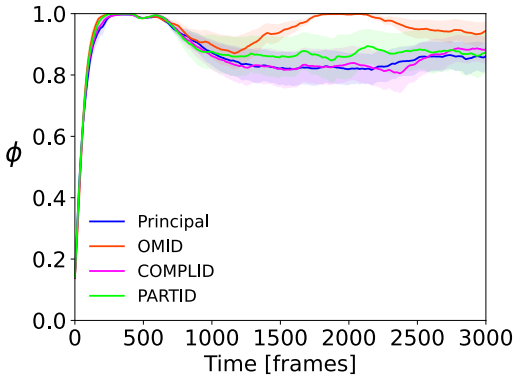


Figure 9: Time evolution of the order parameter in a setup of a 4 time larger arena. All other parameters equal those of Fig. 8f.

2.3.1 Results and Analysis

Examining the results presented in Fig. 8, several observations can be made:

1. All methods in all setups except (f) showed roughly similar convergence behavior. There is a rather simple explanation for the uniformity of convergence across the varying occlusion approaches: the original non-visual Vicsek model delivers perfect velocity information for all approaches. Their differences only appear at higher radii, at which the number of occlusions becomes considerable.

2. There is a moderate yet consistent decrease of asymptotic order with the growth of population size: from (a) to (c) and from (d) to (f). Higher population sizes result in "traffic jams", when groups of agents block the motion of each other, therefore producing systematic disturbance to the flocking process.

3. The best convergence is displayed by all approaches in Fig. (d) of high vision radius and low density. This is the only setup that attained high coherence with an order parameter around 0.95. Here there are no traffic jams to low density and numerous mutual influences due to the high vision radius, which captures a higher number of neighbors.

4. OMID in pane (f) performs best for dense and noisy environments. Since OMID ignores all partially occluded neighbors, it effectively filters out the information from further distanced neighbors. This improves the signal-to-noise ratio.

5. In all cases but one, the order parameter is poor and stays in the range of 0.4-0.7 at all times of the experiment. The arena square geometry results in multiple strong disruptions in the movement and the flocking of the collective: every collision with a wall disrupts the flock, while the corners produce a focusing and aggregating effect, causing agents to get stuck in the corners for long periods of time.

6. A pronounced peak of high order appears at early stages (before 500 frames). The first apparent peak can be attributed to the first encounter of the agents with one of the arena walls. The agents are distributed uniformly in the arena. Empirical observations of the experiments for different random seeds show a common behavioral theme. The agents are able to swarm quite fast into a coherent swarm before hitting a wall. This is more pronounced for large radii since information is exchanged more efficiently across all the agents, and the rate of alignment is higher. Then, when the initial swarm encounters a wall, the agents change direction in a non-synchronized manner, causing the order parameter to drop.

7. In pane (f), COMPLID and PARTID achieve slightly weaker flocking, as shown by their slightly lower order parameters is slightly in comparison to the Principal approach. At large R with high density, only fully occluded agents are filtered-out by the COMPLID and PARTID approaches, in contrast to OMID, which filters out every partial occlusion. Therefore, they neither enjoy the advantage of complete information (of the Principal model) nor the advantage of strong filtering of OMID. As a result, in these methods, convergence is slightly hindered with respect to the principal Vicsek model. Since OMID, which also ignores the fully occluded (in addition to the partially occluded), performs better in the same conditions, we infer that the relationship between the amount of seen neighbors and the order parameter is not linear. As a side note, the difference between COMPLID and PARTID could be attributed to the number of times a neighbor is occluded in its center by a closer neighbor, which in turn causes the focal agent to treat the "split" neighbor as two neighbors with identical velocities.

2.4 Minimizing the Influence of the Boundaries

In order to test whether the differences in the occlusion approaches displayed in pane (f) were indeed the result of the perturbations caused by the walls of the arena, another experiment was performed. In an aim to minimize the influence of the walls, we decided to examine the flock in an arena four times the size. The result is shown in Fig. 9.

As expected, the asymptotic order of the flock rises for all the approaches. Moreover, the flocking peak rises to 1, which represents the ideally flocked state.

Close inspection of pane (f) in Fig. 8 (regular arena, high density, high vision radius) and the large arena in Fig. 9 reveals that OMID approach, which ignores all partially occluded neighbors leads to higher collective order than other occlusion approaches. Moreover, OMID in these dense populations outperforms even the “x-ray vision” of the Principal approach (classic Vicsek), which includes all the neighbors within the radius. Even in medium density in pane (e) this approach outperforms others, though not as convincingly as in pane (f).

On the other hand, in the case of a smaller vision radius shown in panes (a) – (c), the performance of OMID is not distinguishably better. The same can be said for the smallest density experiments in pane (a) and (d). Either at a small R or a large R, we do not observe OMID outperforming other methods.

A possible interpretation of these effects is connected to the fact that, basically, partial occlusions rarely occur in a very small immediate vicinity of the focal agent – there, all the neighbors are directly and fully visible. In a sense, OMID acts as an effective distance filter, preventing the inclusion of further agents in the computation.

As seen in the simulations, the corners of the square arena, as well as the reflections from the walls, introduce a large disruption of directionality, reducing the ordering in the swarm. Parts of the swarm that had undergone two or more reflections are all but uncorrelated to their still "un-reflected" agents. So the inclusion of the reflected ones undermines the flocking process. And so, when OMID prevents further neighbors from influencing the sum, it effectively prevents those already reflected agents from entering the sum, and by the way of giving higher weight to closer neighbors, reduces and attenuates the disrupting influence of reflected parts of the swarm. At short vision radii, OMID has no advantage since all approaches filter-out further neighbors. At high density and large radius, the filtering advantage of OMID becomes the most pronounced.

2.5 Summary

In this section, we proposed three approaches for handling occlusions. The influence of each approach was compared against the other approaches. The results show that the difference between the approaches is significant only in high-density conditions with a large vision radius.

However, so far, we have only dealt with the effect of vision-based occlusions on the performance of the Vicsek model, essentially influencing the "neighbor-selection" part of the model. The calculation of the desired velocity of each agent was based on the exact velocities of neighboring agents. However, this precise information is not directly accessible via vision. We will tackle the latter issue in the following section.

3 A Reductionist Vision-Based Model for Velocity Estimation

Now, we move on to tackle the issue of vision-based velocity estimation. We show how velocities could be estimated with purely visually-available information. Building on top of the original Vicsek model results in a purely vision-based extension of the Vicsek model.

In Section 3.1 we present a vision-based collective motion model. The experimental setup for testing the model is detailed in Section 3.2 followed by the results in Section 3.3. Later, in Section 3.4, we incorporate our occlusion approaches and compare their performance when applied to our vision-based model.

3.1 The Principal Vision-Based Reductionist Model

Similarly to the Vicsek model, The agent's perception is assumed to be omnidirectional (360 degrees) and is limited to a constant radius R around the focal agent. The "realness" of this assumption is supported by the locust vision, which is considered almost omnidirectional [37, 30].

As depicted in Figure 10, a focal agent P heading "up" with a single neighbor n moving with velocity $\boldsymbol{v_{Total}}$ and located at a distance $r < R$ measured between the centers-of-mass of the two agents. The total velocity is combined from tangential $\boldsymbol{v_t}$ and radial $\boldsymbol{v_r}$ velocity components relating to the line of sight (LS), a line between the center-of-mass (COM) of the neighbor and the focal agent. The angular position of the neighbor relative to the heading direction is denoted as (bearing, β), and the angular area is subtended on the vision sensor α n practice. This angle is calculated by finding the vectors from the focal agent towards the vertices on the edge, that subtends the largest angle on the sensor; we shall denote it as the *effective edge*. For convenience, the full subtended angle is regarded as 2α.

Animals with stereoscopic vision perceive their environment as three-dimensional via processes in the brain as parallax that relies on both of our eyes focusing on the same object from two slightly different positions or motion parallaxes, where the observed object remains static while the eye moves. However, for each of our eyes, light rays are reflected from a three-dimensional object and hit a two-dimensional retina. In other words, visual stimuli is one dimension less than the object it originated from. Applying the same logic to our world of two-dimensional agents, visual input would have to be one-dimensional. Essentially, a whole dimension of information about the environment

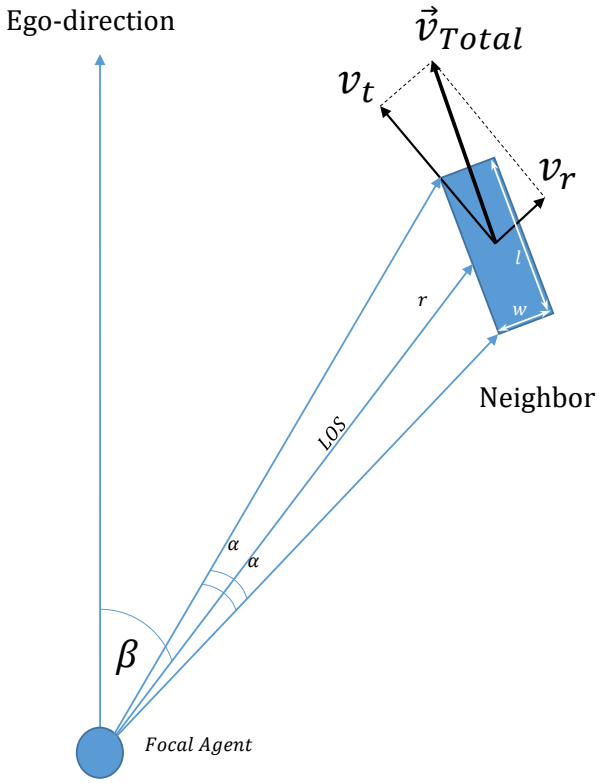


Figure 10: Schematic representation of a focal agent, a neighboring agent, and its visual parameters as observed by the focal agent.

is lost when employing vision. The same goes for our simulated two-dimensional world and how we define vision in the simulation: since the agents are two-dimensional, the visual information an agent perceives, using its non-stereoscopic visual sensor, is one-dimensional, as shown in Fig. 11. The geometric angles explained above constitute, by definition, this one-dimensional type of vision.

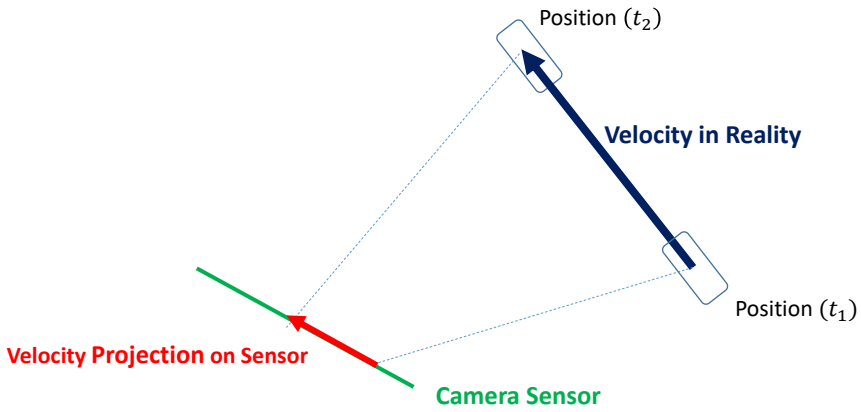


Figure 11: Velocity projection.

Due to the loss of information, the principal challenge when relying on vision is estimating the velocity heading and magnitude of each neighbor. This requires knowledge of the distance to a neighbor [68]. However, it is generally impossible to infer inter-agent distances from projections without additional information. As shown in Fig. 4, different actual velocities may be projected to identical observed velocity vectors and distances. Previous works have often assumed a special case where the agents are circular. This case allows precise calculation of inter-agent distances from projective information since, for disks, distances are in a one-to-one correspondence to the projected visual area. However, as the realistic morphologies of swarming locusts are elongated, we do not make this assumption here.

In this study, we deliberately set our agents to be of elongated shape, in our case, rectangular. We consider a group of N rectangular agents with width w and length l, moving in a two-dimensional environment at speed v with heading parallel to their length axis. The maximal speed the agents can accumulate is v_{max}. The position coordinates $\boldsymbol{x_i}$ of the agent i are updated at discrete time steps according to the motion equation,

$$x_i\left(t+\Delta t\right) = \ x_i\left(t\right) + \ v_i\left(t\right) \cdot \Delta t, \tag{2}$$

with velocity $v_i(t)$ updated at each time-step causing the agent to steer towards a desired velocity with steering-parameter factor η,

$$v_i\left(t+\Delta t\right) = v_{desired}\left(t\right)\cdot\eta + \ v_i\left(t\right)\cdot\left(1-\eta\right), \tag{3}$$

where $v_{desired}$ is calculated based on a decision algorithm similar to the Vicsek Model. The algorithm's objective is to estimate the velocities of the neighbors at each time frame based on visually available geometrical information. Then, the average of these velocities will constitute as $v_{desired}$.

We start our calculation of $\boldsymbol{v}$ by estimating the distance r. Since the orientation of the neighbor is unknown to the observer, we deliberately use a simplistic approximation: the neighbor's effective edge is always perpendicular to the LS, which means the triangle comprised of the focal point P and the two vertices of the effective edge d (see Fig. 10) is assumed to be an equilateral. Thus, the LS constitutes both median and altitude to effective edge and bisector of the subtended angle 2α (See Fig. 10). Therefore, the distance r is given by

$$r = \frac{1}{2}d\cot\alpha, \tag{4}$$

differentiating Eq. 4 we get the radial velocity to be

$$v_r = \frac{d}{dt}r = -\frac{1}{2}\frac{d\,\dot{\alpha}}{\sin^2(\alpha)} \tag{5}$$

where $\dot{\alpha}$ denotes the time derivative of the subtended angle. Expressing $d = 2r\tan\alpha$ from Eq. 4 and substituting into Eq. 5 results in

$$v_r = -\frac{2\dot{\alpha}r}{\sin 2\alpha}, \tag{6}$$

the tangential velocity is obtained using the time derivative of the bearing angle

$$\begin{aligned} \dot{\beta} = \frac{d}{dt}\beta = \omega = \frac{v_t}{r} \\ \Rightarrow v_t = -\dot{\beta}\cdot r, \end{aligned} \tag{7}$$

where $\dot{\beta}$ is the derivative in time of the bearing angle, which is equivalent to the instantaneous angular velocity ω. Time-derivatives are approximated by the difference of parameter's values in two consecutive frames. By calculating both components, we get $v = v_r + v_t$. This process is repeated for all the neighbors, resulting in $v_{desired}$ of the

focal agent.

We assume every flock member possesses knowledge of the conspecific's size (effective edge d is used in the calculation), i.e., an animal has an inherent representation of the typical dimensions of individuals from its species. Combining this knowledge with the angular area angle (α) one can estimate the distance (as shown in Eq. 4). The α angel is also present in loom calculations [30, 29, 66, 24].

We emphasize that this is a baseline model. It assumes all the neighbors are fully visible and does not account for possible obstructions of sight. In other words, agents are presumed to be "transparent", in the sense that no obstruction of further neighbors occurs. Obviously, this assumption is a gross oversimplification of the fundamental capabilities and limitations of visual perception in nature. In Section 3.4 we incorporate the occlusion approaches from Section 2 to address this issue as well.

3.2 Experimental Setup

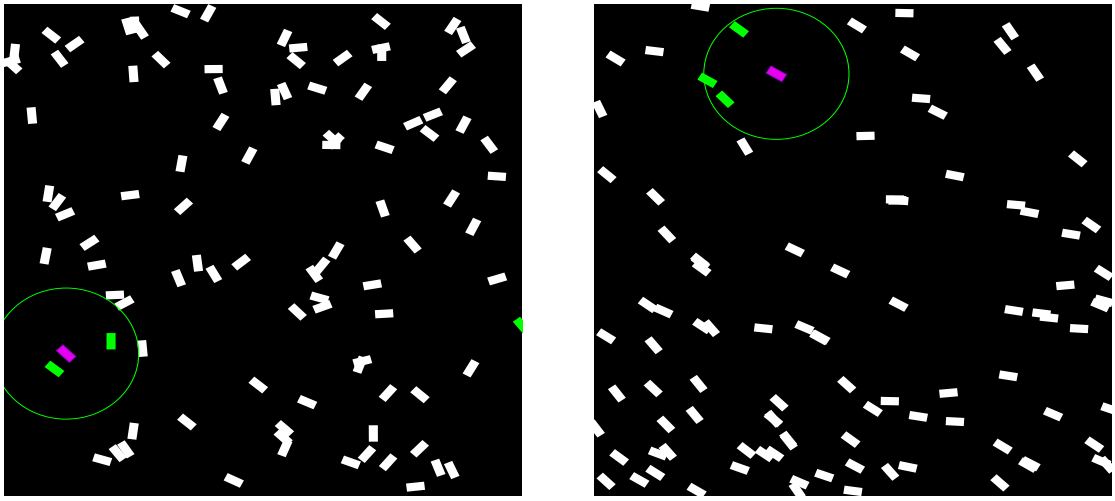
In order to evaluate our vision-based model, we developed a two-dimensional (2D) collective motion simulator based on a basic simulation engine [67] (see Fig. 12). The principal and the occlusion-handling approaches were implemented in the simulation, as detailed below. As described in Section 3.1 the agent's steering is based upon the mean of the neighbors' velocities. These velocities, in turn, are derived from the angular measurements of each perceived neighbor: the subtended angle and the angular velocity of each neighbor. These inputs serve as the agent's subjective perception.

3.2.1 Simulating Kinematics

The agents move in 2D by updating their coordinates at each iteration according to their current velocities. The location and orientation of each rectangular agent are computed from the coordinates of its COM. It is assumed in our model that velocity heading is always along the long axis of the body. The velocity magnitude can vary between 0 and a fixed maximal speed value, i.e., the agents can accelerate up to a maximal speed. Additionally, a steering parameter is introduced to moderate the influence of $v_{desired}$ at each point in time, thus reducing the sharpness of turns and accelerations performed by the agents. This serves as the motion backbone for all the approaches described below.

3.2.2 Simulating Perception

We compare the emergent collective motion resulting from the different occlusion-handling models. The perception of each agent is calculated from the exact values stored in the



(a) Toroidal arena snapshot at $t = 10[frames]$. Agents are initialized at random positions and random velocities. The purple colored agents is an arbitrary marked focal agent with its respective neighbors colored green.

(b) Toroidal arena snapshot at $t = 2000[frames]$. An apparent flocking behavior is displayed, with agents moving roughly in a single direction.

Figure 12: Simulator snapshots.

simulation. This allows emulating different perception models by varying the visual parameters of the agents and the effects of occlusions. Each simulated focal agent is given the information it would have perceived in the 2D environment, in each of the following perception models.

Simulating the Principal Model. The α angle is calculated using the neighbor's vertices of the edge that subtends the largest angle on the agent. The angle between the two vectors pointing from the focal agent's COM to the respective vertices equals to α. The β angle is simply the angle of between the focal agent's velocity vector and the neighbor's COM. The calculation of β is the same for the rest of the approaches. However, since occlusions are non-existent in the principal model, the focal agent receives visual parameters of all the neighbors, including those completely occluded by other more closely distanced neighbors.

Simulating OMID. We simulate this capacity by calculating the effective α for each neighbor, i.e., the subtended angle on the visual field after canceling occluded individuals, meaning that the occluded agent is ignored in its entirety during the calculation of $v_{desired}$. The effective α calculation starts by sorting all the neighbors by their exact distances, then sequentially for each neighbor, finding the angle

"edges" and storing them in a 1D array representing the visual field, spanning from $-\pi$ to π. In case of an angle segment overlaps previously-stored segments, the neighbor is ignored, and the calculation proceeds to the next neighbor.

Simulating COMPLID. We simulate this capacity by calculating the neighbor's "edges" similarly to the process in the OMID approach. Then, subtended angles that are fulling contained in others (i.e., capturing neighbors completely occluded by others in the visual field) are eliminated.

Simulating PARTID. The implementation of the last approach in the simulation start similarly to the OMID and OCLID approaches. We iterate over the neighbors, from the closest to the furthest. Each neighbor effective edges are calculated and then checked against an array of edges. If a partial overlap (i.e., partial occlusion) exist with the current edge and one or two of already checked edges, the effective α is calculated using only the non-overlapping segment.

3.2.3 Controlled (independent) simulation variables

In the simulation experiments, we controlled the following independent variables:

Population size (N) number of simulated agents participating in each experiment.

Agents length/width sets the agent's dimensions, that in turn change the effective edge (d).

Vision radius (R) the radius inside which agents are considered as neighbors (measured in the same units as agent length).

Steering-parameter (η) sets the weight of $v_{desired}$ relative to the current velocity (v) of an agent.

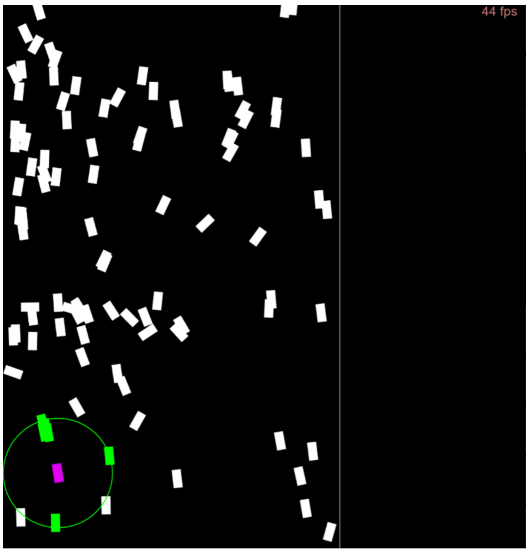
Maximal speed (v_{max}) sets the maximal speed of the agents. In a case the magnitude of $v_{desired}$ exceeds v_{max} at the end of the model's calculation, the simulator will limit it to a magnitude of v_{max} while keeping the same vector direction.

We utilized different areas (*arenas*) to test the agents. The key difference between the arenas is the periodicity of the boundaries:

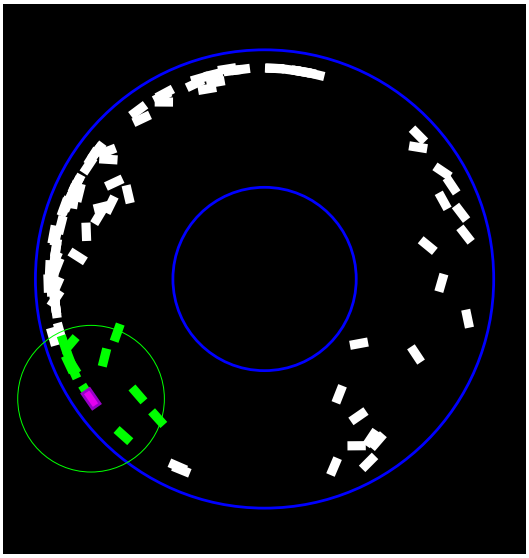
- The first type (shown in Fig. 12) is a square arena with periodic boundary conditions (torus), i.e., once an agent's COM passes the maximal/minimal coordinates or the X/Y axes, it reappears on the other side respectively. The arena dimensions are L, W, which stand for the length and width of the arena.

- The second type is the "infinite corridor" arena (shown in Fig. 13a). In this arena, the vertical boundaries bounce incoming agents by reflecting their velocity's x-component, while the horizontal boundaries are set to be periodic, similarly to the toroidal arena. The distances between the boundaries are varied in the experiments and are detailed below.

- The third is the ring arena, modeled after the ring-shaped arena widely used in *in vivo* locusts flocking experiments (e.g., [15, 2, 41], shown in Fig. 13b). The agents are repelled from the boundaries with varying repelling force, depending on the size of the radial velocity component (relative to the arena center), i.e., an agent traveling to the external circular boundary will be repelled from with a force proportional to the size of the agent's radial velocity component. This was designed in a way to mimic the behavior of living locusts that align themselves to the ring walls so as to avoid possible collision with them [2]. The dimensions of the ring arena are characterized by the two radii of the inner and outer circles that form the ring's boundaries.



(a) Snapshot of corridor arena. The vertical boundaries are repelling, while the horizontal ones are periodic.



(b) Ring arena snapshot.

Figure 13: Corridor and ring arenas.

For the first set of experiments, we consider a group of N agents, with random initial positions and velocities, positioned in a square arena of size $W \times H$ with periodic

boundaries. We studied the dynamics for a range of independent variables' values. We present the results for representative parameters' values with arena size of 500×600, typical agent's main axis length (l) of 20, number of swarm members of $N = 60, 120, 180$, and maximal velocity normalized to $v_{max} = 1$. Vision radii are measured in units of body-length (BL), ranging from $R = 1$ to $R = 6[BL]$. These ranges for group size and vision radius were chosen to achieve flock density corresponding to available biological literature of common flock densities in nature [20].

3.2.4 Measured (dependent) simulation variables

We employ the same order parameter used in Section 2. For detailed description, see Section 2.2.

3.3 Results for the Principal Model

We begin by testing the principal model in a toroidal arena (in many aspects, close to open-field conditions), varying independent simulation variables chosen with accordance to the existing biological literature. We conducted experiments with representative independent variables' values of $N = 120, R = 3$ body lengths [BL], $\eta = 0.01$, and *agent length* set to 20. The population size was set based on observed locust marching-bands densities [10], the sensing radius was set according to empirical observations of locust nymphs not reacting to visual stimuli located farther than 2-3 BLs [5], and the steering-parameter was chosen according to a sensitivity analysis, detailed below).

Figure 14 shows the time-evolution of a simulated swarm's mean polarization (the order parameter ϕ). The order parameter rises from $\phi \approx 0.1$ to an asymptotic value of $\phi \approx 0.9$, which approximately approaches an ideally parallel ordered state.

3.3.1 Parameter Sensitivity Analysis

We investigate the principal model's sensitivity to various parameter values, i.e., the dependence of order parameter ϕ on major controlled quantities: the effects of the steering-parameter factor η, the vision radius R, the population size N, and body length/width ratio. For each of the listed simulation independent parameters, we present two types of graphs. The first is the asymptotic parameter sensitivity, obtained by measuring the order parameter ϕ at a fixed time of $t = 3000$ [frames] for different parameter values. By empirical evidence, at that point in time, the swarm is believed to have reached its asymptotic converged state and no significant improvement of the order-parameter later to follow. The second type of graph is the time-evolution of the order parameter

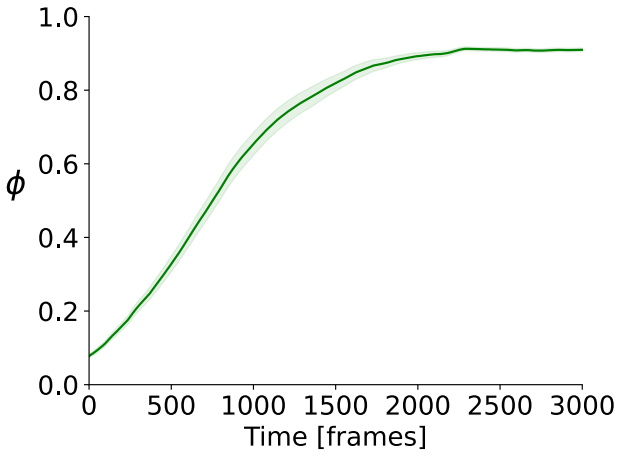


Figure 14: Swarm's alignment level vs time. Data points are the mean order parameter of the swarm at each simulation frame, with a standard error margin. Experiments performed in torus arena.

measured for various parameter values, measured at each frame until the end of the experiment at $t = 3000$ [frames]. For each global parameter, all other parameters' values were fixed arbitrarily (typical fixed values were $N = 180, R = 3 - -4$ [BL], $\eta = 0.01$, agent length$=20 - -30$). For each combination of parameters, the experiments were repeated for 50 independent runs and averaged.

We report on the sensitivity to the steering-parameter η. Fig. 15 indicates that a value of $\eta = 0.01$ yields the maximal value of the order parameter, approaching 1. This is where the swarm is nearly fully aligned. Notably, no convergence occurs for smaller values, meaning agents are too apathetic to the environment and retain their original heading directions. On the other hand, we notice a significant drop in the order parameter magnitude for large η values, i.e., the agents' convergence worsens due to over-sensitivity to the external steering parameter.

Figure 16a displays the time evolution of the order parameter for different η values. As can be seen, there is a bounded zone of values beyond which the algorithm fails to reach any convergence. A further examination of at the simulated flock for these values is shown in figure 16b. It shows that agents aggregate in small tight clusters and constantly change their headings, unable to reach either local or global uniform moving direction.

Based on these results, we fixed the steering-parameter factor parameter to be $\eta = 0.01$ for the rest of our investigations.

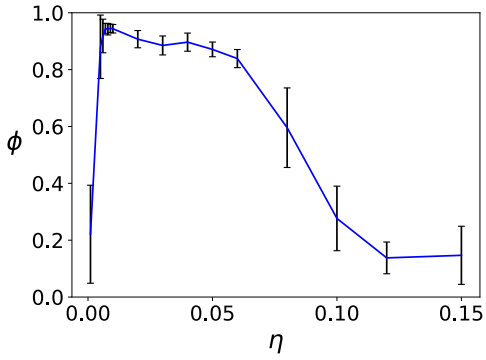
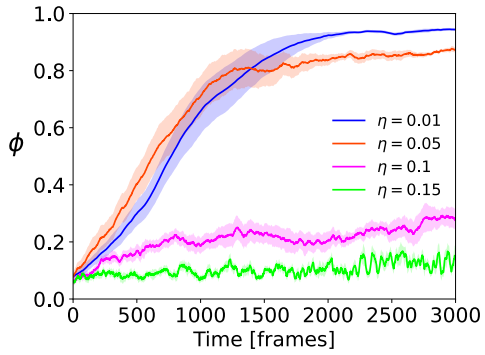
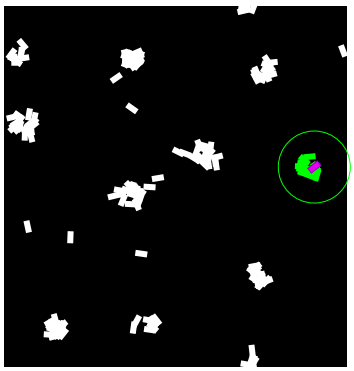


Figure 15: Asymptotic steering parameter's sensitivity analysis. Data points represent the order parameter at t=3000 [frames] for varying η values with a standard error margin. Torus arena.



(a) Time-evolution of the order parameter of the swarm for different steering parameter values. Data points are the mean order parameter of the swarm at each simulation frame, with standard error margin.



(b) Diamond pattern clusterization. High η value.

Figure 16: Collective motion for high values of η. Experiments performed in the torus arena.

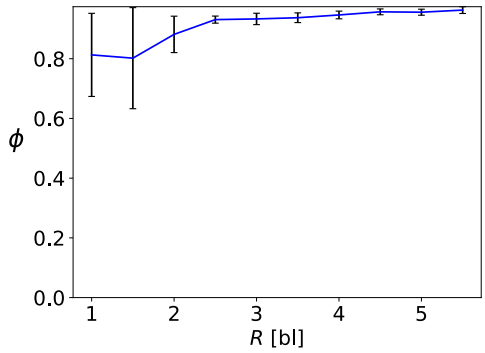
3.3.2 Influence of Vision Radius and Population Size

Fig. 17 shows the dependence of the order parameter on the vision radius R and the population size N, respectively. As seen in Fig. 17a the order parameter increases monotonically as the vision radius grows. For radii $R \geq 50 \sim 2.5\,[BL]$, at $R = 90$ the improvement of the convergence process reaches saturation and the evolution towards a fully flocked state becomes the steepest.

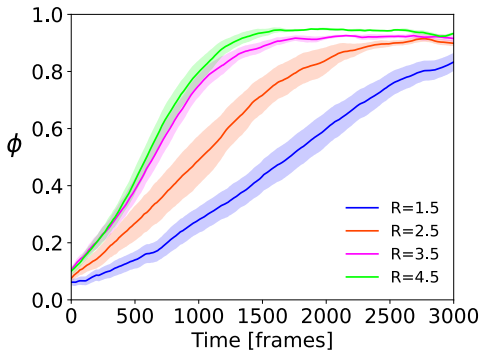
Figures 17c,17d depict the population size sensitivity analysis. As seen in Fig. 17c for relatively small N sizes, increasing the population leads to higher asymptotic order parameter, though it reaches saturation for $N \geq 100$. Moreover, in Fig. 17d a significant qualitative difference is seen between the convergence of the order parameter for $N = 20$ compared with bigger population sizes.

3.3.3 Influence of Agent's Dimensions

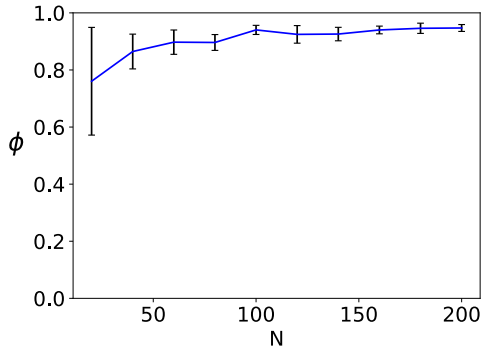
We set to test the role of an agent's dimensions in our model by comparing the parameter for increasing agent's lengths, making the agents more and more elongated. As shown in Fig. 18, the model behaves practically the same for more elongated agent's morphologies. In Fig. 18a we see that the asymptotic order parameter is relatively the same for different dimensions, and even increases a bit for longer agents. Fig. 18b strengthens the above by pointing the effective ratios of agents' dimensions (length / width) and the respective time-evolution for each body-length.



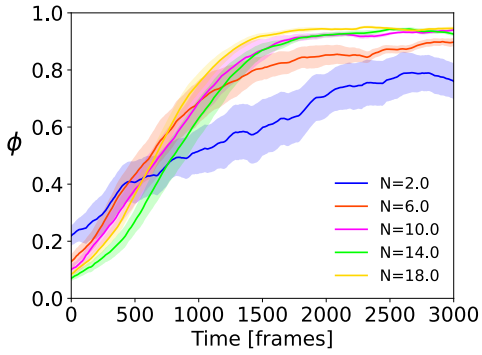
(a) Asymptotic vision radius (R) sensitivity analysis.



(b) Time-evolution of the order parameter of a swarm for different vision radii.

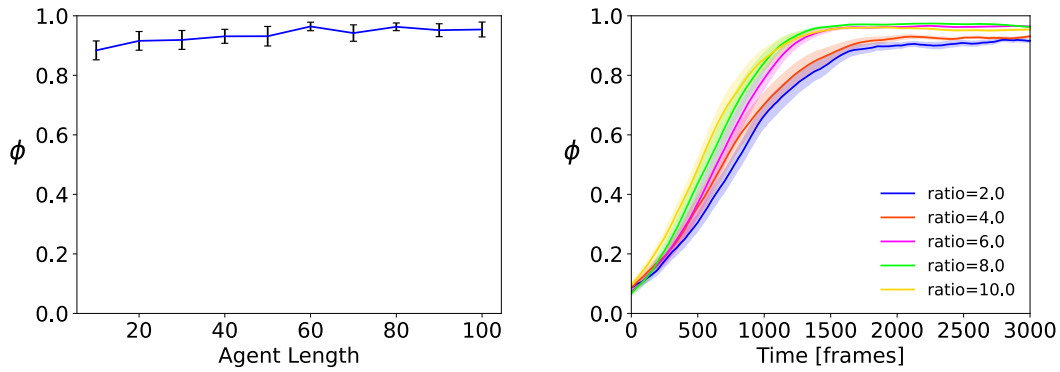


(c) Asymptotic population size (N) sensitivity analysis.



(d) Time-evolution of the order parameter of a swarm for different population sizes.

Figure 17: Vision radius and population size sensitivity analysis. For asymptotic analysis: data points represent the order parameter at t=3000 [frames] for varying R and N respective varying values, with standard error margin. For time-evolution analysis: data points are the mean order parameter of the swarm at each simulation frame, with standard error margin. Torus arena.



(a) Asymptotic sensitivity analysis for different agent lengths.

(b) Time-evolution of the order parameter for different agent length. The legend states the various ratios between the increasing agent's length and its constant width.

Figure 18: Agent's length sensitivity analysis. For asymptotic analysis: data points represent the order parameter at t=3000 [frames] for varying l values with standard error margin. For time-evolution analysis: data points are the mean order parameter of the swarm at each simulation frame, with standard error margin. Torus arena.

3.4 Incorporating the three occlusion approaches

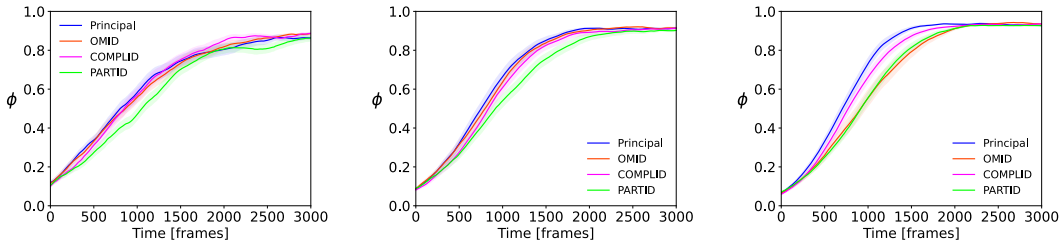
We now turn to investigate the role of occlusions by neighbors when used on top of the reductionist distance and velocity estimation model. The three approaches are evaluated in comparison with the *principal* model (which does not account for occlusions). By way of a reminder: the first approach (OMID) ignores any neighbor that is partially obstructed by another neighbor; the second approach (COMPLID) extrapolates any partially visible neighbor to its full size as if the occluding neighbors do not occlude; and finally, the last approach (PARTID), assumes that any unique set of visible parameters $(\alpha, \dot{\alpha},\ \beta,\ \dot{\beta}\ )$ originates in a completely visible neighbor, so the visible parts of partially occluded neighbors are treated as if they are a complete neighbor. PARTID produces the highest potentially erroneous estimations of distance and velocity of a single neighbor, since the subtended angle $(\alpha, \dot{\alpha})$ parameters are greatly affected by occlusions.

Partially-occluded neighbors generate additional errors, as the projected area of the subtended angle is used as a proxy for distance. For example, suppose a neighbor is partially occluded, such that only a small portion of it is observed, and thus it may be perceived to be distant. If the occluding animal moves to uncover it, its full length is now revealed, and it will now be seen as being close within a very short time, implying high velocity towards the observer and a potential collision. The accumulation of such frequent errors may disturb the stability of the swarm.

3.4.1 Torus Arena

In Fig. 19 time-evolution in the *torus* arena of the order-parameter ϕ is compared between the four approaches. The graphs show the mean order parameter for each point in time. Data is averaged over 30 runs. Three population sizes of $N = 60, 120, 180; R = 3\,[BL], l = 30, \eta = 0.01$. As seen from the figures, all three occlusion approaches, alongside the original model, reach similar asymptotic order-parameter values ($\phi \sim 0.9$), indicative of reaching similar degrees of synchronization after flocking. It is seen from Fig. 19b and Fig. 19c that PARTID has a slower rate of convergence. It can be viewed as additional evidence that this approach results in excessively noisy perception. At higher densities (Fig. 19c) the rate of convergence of all three perceptive approaches lag behind the Principal model. At higher densities, rates of convergence become steeper. At the same time, the asymptotic order parameter remains very close for all the methods, and even different densities. Finally, the ordering of the rates of convergence at 180 indicates that COMPLID converges faster than OMID. Completing parts of neighbors, instead of omitting them, leads to an effectively larger number of neighbors, which leads to stronger

alignment.



(a) (b) (c)

Figure 19: Comparison of the occlusion approaches in the torus arena. Data points are the mean order parameter of the swarm at each simulation frame, with a standard error margin. Compared population sizes: (a) 60 agents, (b) 120 agents, (c) 180 agents. Convergences are practically indistinguishable between approaches. A higher population leads to a slightly steeper transition to a flocked state.

3.4.2 Infinite Corridor and Ring Arenas

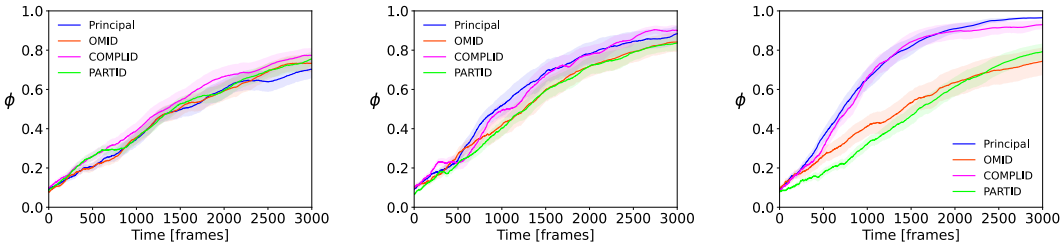
In order to further test the model's robustness, we investigate the behavior of the different approaches (including the principal model, which does not handle occlusions), examining flocking in other geometrical environments. Specifically, we examine the behavior in an infinite corridor and in a ring environment (Figure 20) similar to ring arenas present in biological locust experiments, as noted above.

Three versions of each arena type are tested: *wide, intermediate, and narrow*. The geometry of the arenas will be characterized by the ratio of arena-width to single agent body-length. In our chosen measurement units of body-length (BL) this parameter is simply the arena width. A swarm of 100 agents was run in each arena configuration. For the infinite corridor, the distance between the periodical boundaries (length) was 20 [BL] for all the experiments. The widths were: 10, 23, 33 [BL], respectively. For the ring arena, the radius of the inner circle is 1.66 [BL] and the outer circle radii tested were: 5, 8.33, 11.66 [BL].

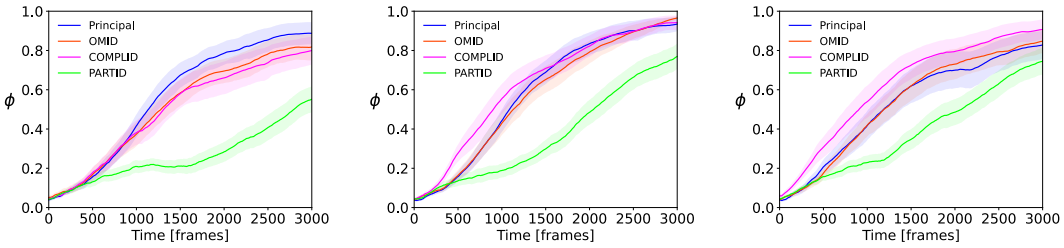
Analyzing the infinite corridor arena, we observe an interesting distinction displayed in the *narrow* corridor. All perceptive approaches show a rise in the ordering parameter: The Principal model and the COMPLID approach converge significantly faster and to a higher asymptotic value than OMID and PARTID. The fact that this difference is significant only in the narrow environment indicates that the harsher geometrical constraints cause some approaches to struggle. A potential reason is that in narrow arenas, interactions

with the boundaries are much more frequent. Principal and COMPLID converge better even in highly perturbed geometry, this indicates that acquiring accurate information from the maximal amount of seen neighbors is beneficial in constrained environments. As a reminder, COMPLID is able to essentially approximate the information available to the principal model apart from the agents that are fully occluded. Thus, the information it perceives is quite similar to the all-seeing Principal approach.

In the ring arena, we observe a common theme in all the different widths. Although the rest of the approaches seem to converge successfully to an ordered state with similar rates, the PARTID seems to struggle with the constraints imposed by the circular boundaries. The agents in the ring arena are faced with external perturbations by the boundaries that interfere with the alignment of the model itself in a way that is critically detrimental to the PARTID approach.



(a) Corridor - Wide (b) Corridor - Intermediate (c) Corridor - Narrow



(d) Ring - Wide (e) Ring - Intermediate (f) Ring - Narrow

Figure 20: (a),(b),(c): Infinite corridor arena. Wide / Intermediate / Narrow arena dimensions are 20 × 33 / 23 / 10 [BL], respectively (agent BL 20-30). Data points are the mean order parameter of the swarm at each simulation frame, with a standard error margin. (d),(e),(f): Ring arena. Wide / Intermediate / Narrow ring external border radii are 11.66 / 8.33 / 5, respectively. The internal ring border is constant for all three types and equals 50. (agent BL 20–30). Data points are the mean order parameter of the swarm at each simulation frame, with a standard error margin.

4 Discussion and Conclusions

In this work, we studied the limitations and implications of vision as the key source of information for flocking agents. In the first part, we investigated the effect of partial occlusions (of one neighbor by another) on flocking-related decision-making. To this end, visual cues from partially occluded objects constitute lost or distorted information. Thus, the problem of reaching a collective state becomes more challenging. Previous attempts to account for occlusions and their role in vision-based flocking have been very limited. To the best of our knowledge, the present work constitutes one of the first attempts to investigate this question thoroughly.

The influence of partial occlusions was tested in a series of simulations, where each approach was applied on top of the Vicsek model, which was used as the backbone for agents' motion policy. The first (COMPLID) completes the outline of a partially hidden neighbor. Such abilities were demonstrated in various species, including insects [51, 36]. However, this approach is the most complex of the three (cognitive-wise) as it requires object detection combined with extrapolating capabilities. The second (OMID) ignores any partial information altogether. Thus, it calls for differentiating between fully vs. partially observed neighbors. This approach requires object detection; however, it is somewhat simpler than COMPLID since it only filters out erroneous visual stimuli rather than correcting them in the observer's mind. The last one (PARTID) treats each segment of a neighbor as if it represents a full-length body. Hence, the latter is the simplest of the three since it requires minimal cognitive processing from the individual — neither object detection nor object recognition, each set of distinguishable visual parameters is regarded as a neighbor. This results in PARTID being the most erroneous perception of the surrounding agent, e.g., segments change due to closer neighbors revealing or occluding them, which in turn is perceived as neighbors moving away or towards the focal agent.

The difference in the required computational power under the different approaches is very significant, as in nature, organisms demonstrating collective motion are very often limited in this respect (small brains, simple neuronal substrates). Hence, finding the least computationally demanding algorithm still capable of reaching flocking can potentially explain the actual mechanisms involved in the flocking of these relatively simple species.

The three occlusion approaches were tested in a series of simulations, where each approach was applied on top of the Vicsek model, which was used as the backbone for agents' motion policy. The difference between the three proposed approaches only shows in Fig. 8f, where agents' population and the vision radius were the largest out of the

numbers tested. OMID higher resilience to the perturbation induced by the square arena walls prevailed even in the omniscient Vicsek model. When tested in a significantly larger arena, where the influence of the walls was substantially diminished, again, OMID excelled over the others.

In the second part of this thesis, we have introduced a non-stereoscopic perception framework and developed a novel, purely vision-based flocking algorithm. Taking a reductionist approach, we considered 2D agents, aiming to generate a collective motion with the bare minimum of information. This model employs geometrical aspects of vision, such as subtended visual angle, observable angular velocity, and other derived parameters. The emergence of the collective state and its dynamics were achieved in several arenas. Further on, we examined the model with the occlusion approaches described in the first part, applied on top of the vision-based motion model. The flocking process showed to be robust enough to emerge with occlusions.

The vision-based motion model does not assume direct measurement of inter-agent distances or velocities. Instead, it infers distances and velocities from observed angles and their rates of change. Such measurements are fully available in non-stereoscopic vision. This was one of the main motivations of the current study, and as noted, it is essential for modeling natural flocking organisms that lack stereoscopic vision, the prime examples being insects or fish, as well as similar artificial agents. The inability to directly assess range hinders even the acquisition of velocity direction of the neighbors, presenting a significant challenge to efficient flocking. To our knowledge, this is the first time a non-stereoscopic vision-based model, estimating neighbors' velocities only from angular information, has demonstrated flocking.

The collective state is reached by the gradual alignment of the agent's velocity vector with the velocity vectors of its neighbors. The algorithm computes the average of the velocities of an agent's neighbors that are located within a vision radius. It was observed that locusts [5] mostly react to their neighbors up to a limited visual distance. Visual radius thus limits the agent's knowledge about his neighbors. We expect a smaller vision radius to reduce collective coordination between the flock members while a larger one to increase their coordination. A natural assertion would expect better convergence to occur when using higher vision radii. Experimental results have confirmed this dependence of convergence rates on R, as shown in Figs. 17a,17b. Rates of convergence were higher for larger R values. Also, the final asymptotic degrees of the order parameter were higher for larger radii. A biologically relevant vision radius significance is evident in Fig. 17a. The asymptotic order parameter reaches saturation at approximately 2-3 BL. In other words, increasing R beyond 3 BL does not improve asymptotic order.

Next, we combined the vision-based flocking model with the occlusion handling approaches. We tested the model with the various approaches in a series of simulations performed in several arenas and different conditions.

In the torus arena, all three perception approaches of occlusions have successfully demonstrated the flocking transition from a chaotic initial state to an ordered collective state. Analyzing in more detail, we saw slight differences in convergence rates, where PARTID consistently performs slower than the others since it creates a particularly high misperception rate. This deficiency of PARTID is more pronounced in higher densities when a higher number of neighbors are present around the focal agent. This, in turn, leads to an increased prevalence of occlusions and, consequently, to the increased relative noisiness of PARTID.

We tested our approaches in geometrically constrained environments (arenas): corridors of varying width and rings of varying thickness, where the boundaries lead to velocity reflections. The convergence process becomes significantly perturbed by rebounds of agents colliding with the boundaries.

Proceeding to examine the corridor arena, in the wide corridor, all four behaved statistically similarly as seen in Fig. 20a. In the intermediate width, a slight separation between the approaches emerges 20b. Finally, at the narrow corridor, a strongly pronounced difference is displayed between COMPLID (at faster convergence) and OMID and PARTID group with slower convergence. It implies that COMPLID is the more robust method, withstanding frequent disruptions by the boundary and performing just as well as the baseline Principal model.

Moving on to examine the ring arena, the results demonstrate a distinct weakness of the PARTID approach, especially pronounced in the wide ring. The other two perceptive approaches behaved similarly, though not identical. One can surmise that ring geometry exudes high geometric influence on the collective motion, thereby leveling out the consequences of different methods, with the sole exception of PARTID. It could be suggested that the frequent change of direction present in the ring arena causes agents to re-appear and disappear from occluding neighbors, therefore strongly perturbing PARTID perception, while COMPLID and OMID cope well with these changes due to the better treatment of occluded neighbors.

Comparing the results of the approaches from the first part, where exact velocities were used, with the results from the second part, we see a common theme. Only in highly constrained environments do the approaches become distinguishable. However, we should point out several divergences in the results of the two parts. First, the vision radius in the first part, sufficiently big enough to express the differences between the approaches,

was 8 BL rather than 3 BL in the second part. This could be due to the difference of the simulations. While in the first part, all agents featured physical dimensions preventing them from colliding through to each other, in the second simulation, this was allowed, causing agents to move "on top" of one another. The latter allowed for a bigger amount of agents fitting within the vision radius and creating occlusions.

Another difference in the results can be found in the effect arena boundaries have on the flocking process. The behavior of the order parameter is much noisier in the square arena experiments than in the ring and corridor arenas. There is a major geometrical difference between them. In the ring and corridor arenas, the interactions with the boundaries guide the agent into order, the ring eventually prods the agents to move along its perimeter (either clockwise or counter-clockwise), and the corridor blocks free movement along one axis, which eventually causes the agents to flock either upwards or downwards. On the other hand, the square arena used in the experiment of the first part introduces recurring shocks to each agent, practically forcing the agent to undergo a radical direction shift.

To examine the possible implications of our findings for the real world, one ought to consider the biological plausibility of the three different approaches; in other words, the likeliness that the approaches describe the actual visual processing of animals. The topography of natural terrain has creeks, valleys, ridges, and other lateral constraints resulting in effectively constrained geometry. As is well established (see [25] and references within), marching locusts bands keep flock formation successfully despite such constraints. Considering our results from constrained arenas, we infer that PARTID is an oversimplification of the perceptive mechanisms in locust vision. Advanced capabilities are required for coping with occluded neighbors, such as the proposed OMID or COMPLID approaches. COMPLID, as referenced in Sec. 2, has empirical evidence of similar capabilities in animals that show the ability to infer complete outlines from only partial visual cues.

The present work studies different aspects of vision-based collective motion in swarms. The biological inspiration was to study purely visual and non-stereoscopic inputs without direct distance measurements and to account for occlusions. We showed that the reductionist model is sufficient in principle to achieve collective behavior in a swarm for different geometries. However, its reliability is dependent on the mechanisms used to address occlusions. This is an important area of continued investigation. The results here also are relevant for navigation and flocking in robotic swarms, where often robots lack stereo vision and ranging sensory capabilities (e.g., LIDAR devices) and are expensive in energy use and computation [39]. As directions for further study, the minimalist model

could serve as the basis both for the development of more robust mechanisms and for solving more challenging environments. It also informs the use of alternative interaction modalities observed in nature or technologically applicable.

This work establishes that a non-stereoscopic, purely vision-based model can achieve flocking without direct kinematic information — neither of distances nor velocities. It is an essential step for modeling natural flocking organisms that lack stereoscopic vision, e.g., insects or fish, as well as similar artificial agents. Also, the stability of the model for some simplified approaches to neighbor occlusion was proven. Analyzing the differing emergence rates of collective motion supports the necessity of higher-level visual processes such as outline completion and conspecific recognition.

References

[1] M. Aldana, V. Dossetti, C. Huepe, V. M. Kenkre, and H. Larralde. Phase Transitions in Systems of Self-Propelled Agents and Related Network Models. *Physical Review Letters*, 98(9):095702, Mar. 2007.

[2] G. Amichay, G. Ariel, and A. Ayali. The effect of changing topography on the coordinated marching of locust nymphs. *PeerJ*, 4:e2742, 2016.

[3] I. Aoki. internal dynamics of fish schools in relation to inter-fish distance. *Nippon Suisan Gakkaishi*, 50(5):751–758, 1984.

[4] G. Ariel and A. Ayali. Locust collective motion and its modeling. *PLOS Computational Biology*, 11(12):e1004522, 2015.

[5] G. Ariel, Y. Ophir, S. Levi, E. Ben-Jacob, and A. Ayali. Individual pause-and-go motion is instrumental to the formation and maintenance of swarms of marching locust nymphs. *PLOS ONE*, 9(7):e101636, 7 2014.

[6] A. Ayali. The puzzle of locust density-dependent phase polyphenism. *Current opinion in insect science*, 35:41–47, 2019.

[7] G. Baglietto and E. V. Albano. Nature of the order-disorder transition in the Vicsek model for the collective motion of self-propelled particles. *Physical Review E*, 80(5):050103, Nov. 2009.

[8] R. Bastien and P. Romanczuk. A model of collective behavior based purely on vision. *Science Advances*, 6(6), 2 2020.

[9] S. Bazazi, J. Buhl, J. J. Hale, M. L. Anstey, G. A. Sword, S. J. Simpson, and I. D. Couzin. Collective motion and cannibalism in locust migratory bands. *Current biology*, 18(10):735–739, 2008.

[10] L. Bennett and P. Symmons. A review of estimates of numbers in some types of desert locust (schistocerca gregaria (forsk.)) populations. *Bulletin of Entomological Research*, 61(4):637–649, 1972.

[11] F. Berlinger, M. Gauci, and R. Nagpal. Implicit coordination for 3d underwater collective behaviors in a fish-inspired robot swarm. *Science Robotics*, 6(50), 1 2021.

[12] I. Bleichman, P. Yadav, and A. Amir. Visual processing and collective motion-related decision making in desert locusts. *Proc. R. Soc. Lond. B*, 2022. see also bioRxiv ; https://doi.org/10.1101/2022.09.19.508462.

[13] N. W. Bode, J. J. Faria, D. W. Franks, J. Krause, and A. J. Wood. How perceived threat increases synchronization in collectively moving animal groups. *Proceedings of the Royal Society B: Biological Sciences*, 277(1697):3065–3070, 10 2010.

[14] V. Bruce, P. R. Green, and M. A. Georgeson. *Visual perception: Physiology, psychology, & ecology.* Hove & London: Psychology Press, 4th edition, 2003.

[15] J. Buhl, D. J. Sumpter, I. D. Couzin, J. J. Hale, E. Despland, E. R. Miller, and S. J. Simpson. From disorder to order in marching locusts. *Science*, 312(5778):1402–1406, 6 2006.

[16] R. L. Canosa. Real-world vision: Selective perception and task. *ACM Transactions on Applied Perception (TAP)*, 6(2):1–34, 2009.

[17] H. Chaté, F. Ginelli, G. Grégoire, and F. Raynaud. Collective motion of self-propelled particles interacting without cohesion. *Physical Review E*, 77(4):046113, Apr. 2008.

[18] A. K. Chauhan and P. Krishan. Moving object tracking using gaussian mixture model and optical flow. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(4):243–246, 2013.

[19] B. Collignon, A. Séguret, and J. Halloy. A stochastic vision-based model inspired by zebrafish collective behaviour in heterogeneous environments. *Royal Society Open Science*, 3(1), 1 2015.

[20] I. D. Couzin, J. Krause, N. R. Franks, and S. A. Levin. Effective leadership and decision-making in animal groups on the move. *Nature*, 433(7025):513–516, 2005.

[21] M. A. Cox, M. C. Schmid, A. J. Peters, R. C. Saunders, D. A. Leopold, and A. Maier. Receptive field focus of visual area v4 neurons determines responses to illusory surfaces. *Proceedings of the National Academy of Sciences of the United States of America*, 110(42):17095–17100, 10 2013.

[22] D. A. Cullen, A. J. Cease, A. V. Latchininsky, A. Ayali, K. Berry, J. Buhl, R. De Keyser, B. Foquet, J. C. Hadrich, T. Matheson, et al. From molecules to management: mechanisms and consequences of locust phase polyphenism. In *Advances in Insect Physiology*, volume 53, pages 167–285. Elsevier, 2017.

[23] A. Czirók, A. L. Barabási, and T. Vicsek. Collective motion of self-propelled particles: Kinetic phase transition in one dimension. *Physical Review Letters*, 82(1):209, 1 1999.

[24] S. E. De Vries and T. R. Clandinin. Loom-sensitive neurons link computation to action in the drosophila visual system. *Current Biology*, 22(5):353–362, 3 2012.

[25] J. Dkhili, U. Berger, L. M. I. Hassani, S. Ghaout, R. Peters, and C. Piou. Self-organized spatial structures of locust groups emerging from local interaction. *Ecological Modelling*, 361:26–40, 2017.

[26] J. R. Dunbier, S. D. Wiederman, P. A. Shoemaker, and D. C. O'Carroll. Facilitation of dragonfly target-detecting neurons by slow moving features on continuous paths. *Frontiers in Neural Circuits*, 0(OCTOBER 2012):1–29, 10 2012.

[27] C. Escudero, C. A. Yates, J. Buhl, I. D. Couzin, R. Erban, I. G. Kevrekidis, and P. K. Maini. Ergodic directional switching in mobile insect groups. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 82(1):011926, 7 2010.

[28] N. Geva, M. Guershon, M. Orlova, and A. Ayali. Memoirs of a locust: Density-dependent behavioral change as a model for learning and memory. *Neurobiology of learning and memory*, 93(2):175–182, 2010.

[29] M. A. Goodale, C. G. Ellard, and L. Booth. The role of image size and retinal motion in the computation of absolute distance by the mongolian gerbil (meriones unguiculatus). *Vision research*, 30(3):399–413, 1990.

[30] J. R. Gray, E. Blincow, and R. M. Robertson. A pair of motion-sensitive neurons in the locust encode approaches of a looming object. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 196(12):927–938, 12 2010.

[31] G. Grégoire and H. Chaté. Onset of Collective and Cohesive Motion. *Physical Review Letters*, 92(2):025702, Jan. 2004.

[32] B. B. Guest and J. R. Gray. Responses of a looming-sensitive neuron to compound and paired object approaches. *Journal of neurophysiology*, 95(3):1428–1441, 2006.

[33] T. Hamada. Vision, action, and navigation in animals. *Visual Navigation: From Biological Systems to Unmanned Ground Vehicles*, 2:1, 1997.

[34] D. Helbing, I. Farkas, and T. Vicsek. Simulating dynamical features of escape panic. *Nature*, 407(6803):487–490, 2000.

[35] J. E. Herbert-Read, A. Perna, R. P. Mann, T. M. Schaerf, D. J. Sumpter, and A. J. Ward. Inferring the rules of interaction of shoaling fish. *Proceedings of the National Academy of Sciences of the United States of America*, 108(46):18726–18731, 11 2011.

[36] G. A. Horridge, S. Zhang, and D. O'Carroll. Insect perception of illusory contours. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 337(1279):59–64, 1992.

[37] S. J. Judge and F. C. Rind. The locust DCMD, a movement-detecting neurone tightly tuned to collision trajectories. *Journal of Experimental Biology*, 200(16):2209–2216, 8 1997.

[38] G. Kanizsa, P. Renzi, S. Conte, C. Compostela, and L. Guerani. Amodal completion in mouse vision. *Perception*, 22(6):713–721, 6 1993.

[39] M. Keidar and G. A. Kaminka. Efficient frontier detection for robot exploration. *IJRR*, 33(2):215–236, 2014.

[40] D. Knebel, A. Ayali, M. Guershon, and G. Ariel. Intra-versus intergroup variance in collective behavior. *Science advances*, 5(1):eaav0695, 2019.

[41] D. Knebel, C. Sha-ked, N. Agmon, G. Ariel, and A. Ayali. Collective motion as a distinct behavioral state of the individual. *iScience*, 24(4):102299, 4 2021.

[42] A. Kolpas, J. Moehlis, and I. G. Kevrekidis. Coarse-grained analysis of stochasticity-induced switching between collective motion states. *Proceedings of the National Academy of Sciences of the United States of America*, 104(14):5931–5935, 4 2007.

[43] H. Kunz and C. K. Hemelrijk. Simulations of the social organization of large schools of fish whose perception is obstructed. *Applied Animal Behaviour Science*, 138(3-4):142–151, 5 2012.

[44] B. H. Lemasson, J. J. Anderson, and R. A. Goodwin. Collective motion in animal groups from a neurobiological perspective: The adaptive benefits of dynamic sensory loads and selective attention. *Journal of Theoretical Biology*, 261(4):501–510, 12 2009.

[45] B. H. Lemasson, J. J. Anderson, and R. A. Goodwin. Motion-guided attention promotes adaptive communications during social navigation. *Proceedings of the Royal Society B: Biological Sciences*, 280(1754), 3 2013.

[46] I. R. Lin and C. C. Chiao. Visual equivalence and amodal completion in cuttlefish. *Frontiers in Physiology*, 8(FEB):40, 2 2017.

[47] S. J. Luck and M. A. Ford. On the role of selective attention in visual perception. *Proceedings of the National Academy of Sciences*, 95(3):825–830, 1998.

[48] E. Mascalzoni and L. Regolin. Animal visual perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(1):106–116, 1 2011.

[49] A. Mogilner and L. Edelstein-Keshet. A non-local model for a swarm. *Journal of mathematical biology*, 38(6):534–570, 1999.

[50] N. Moshtagh, N. Michael, A. Jadbabaie, and K. Daniilidis. Vision-based, distributed control laws for motion coordination of nonholonomic robots. *IEEE Transactions on Robotics*, 25(4):851–860, 2009.

[51] A. Nieder. Seeing more than meets the eye: processing of illusory contours in animals. *Journal of Comparative Physiology A 2002 188:4*, 188(4):249–260, 4 2002.

[52] V. Nityananda and J. C. Read. Stereopsis in animals: evolution, function and mechanisms. *Journal of Experimental Biology*, 220(14):2502–2512, 2017.

[53] A. Okubo. Dynamical aspects of animal grouping: swarms, schools, flocks, and herds. *Advances in biophysics*, 22:1–94, 1986.

[54] J. K. Parrish, S. V. Viscido, and D. Grunbaum. Self-organized fish schools: an examination of emergent properties. *The biological bulletin*, 202(3):296–305, 2002.

[55] B. L. Partridge. Internal dynamics and the interrelations of fish in schools. *Journal of comparative physiology*, 144(3):313–325, 1981.

[56] B. L. Partridge. The structure and function of fish schools. *Scientific american*, 246(6):114–123, 1982.

[57] C. Pinciroli, V. Trianni, R. O'Grady, G. Pini, A. Brutschy, M. Brambilla, N. Mathews, E. Ferrante, G. Di Caro, F. Ducatelle, M. Birattari, L. M. Gambardella, and M. Dorigo. ARGoS: a modular, parallel, multi-engine simulator for multi-robot systems. *Swarm Intelligence*, 6(4):271–295, 2012.

[58] D. Pita, B. Collignon, J. Halloy, and E. Fernández-Juricic. Collective behaviour in vertebrates: A sensory perspective. *Royal Society Open Science*, 3(11), 11 2016.

[59] D. K. Prasad. Survey of the problem of object detection in real images. *International Journal of Image Processing (IJIP)*, 6(6):441, 2012.

[60] J. Qi, L. Bai, Y. Xiao, Y. Wei, and W. Wu. The emergence of collective obstacle avoidance based on a visual perception mechanism. *Information Sciences*, 582:850–864, 1 2022.

[61] R. S. Rakibe and B. D. Patil. Background subtraction algorithm based human motion detection. *International Journal of scientific and research publications*, 3(5):2250–3153, 2013.

[62] C. W. Reynolds. Flocks, herds and schools: A distributed behavioral model. In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pages 25–34, 1987.

[63] P. Roessingh, S. J. Simpson, and S. James. Analysis of phase-related changes in behaviour of desert locust nymphs. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 252(1333):43–49, 1993.

[64] S. B. Rosenthal, C. R. Twomey, A. T. Hartnett, H. S. Wu, and I. D. Couzin. Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion. *Proceedings of the National Academy of Sciences of the United States of America*, 112(15):4690–4695, 4 2015.

[65] R. K. Rout. *A survey on object detection and tracking algorithms*. PhD thesis, Department of Computer Science and Engineering, National Institute of Technology Rourkela, 2013.

[66] R. D. Santer, F. C. Rind, R. Stafford, and P. J. Simmons. Role of an identified looming-sensitive neuron in triggering a flying locust's escape. *Journal of Neurophysiology*, 95(6):3391–3400, 6 2006.

[67] E. Schmidt. Ernst schmidt - coding.

[68] J. R. Serres and F. Ruffier. Optic flow-based collision-free strategies: From insects to robots. *Arthropod structure & development*, 46(5):703–717, 2017.

[69] N. Shimoyama, K. Sugawara, T. Mizuguchi, Y. Hayakawa, and M. Sano. Collective motion in a system of motile elements. *Physical Review Letters*, 76(20):3870, 1996.

[70] A. R. D. Silva, W. S. Lages, and L. Chaimowicz. Boids that see: Using self-occlusion for simulating large groups on GPUs. *Computers in Entertainment*, 7(4):1–20, Dec. 2009.

[71] S. J. Simpson, A. R. McCAFFERY, and B. F. Hägele. A behavioural analysis of phase change in the desert locust. *Biological reviews*, 74(4):461–480, 1999.

[72] M. Singh. Modal and amodal completion generate different shapes. *Psychological Science*, 15(7):454–459, 7 2004.

[73] E. Soria, F. Schiano, and D. Floreano. The influence of limited visual sensing on the reynolds flocking algorithm. *Proceedings - 3rd IEEE International Conference on Robotic Computing, IRC 2019*, pages 138–145, 3 2019.

[74] A. Strandburg-Peshkin, C. R. Twomey, N. W. Bode, A. B. Kao, Y. Katz, C. C. Ioannou, S. B. Rosenthal, C. J. Torney, H. S. Wu, S. A. Levin, and I. D. Couzin. Visual sensory networks and effective information transfer in animal groups. *Current Biology*, 23(17):R709–R711, 9 2013.

[75] C. Sukanya, R. Gokul, and V. Paul. A survey on object recognition methods. *International Journal of Science, Engineering and Computer Technology*, 6(1):48, 2016.

[76] J. Toner and Y. Tu. Flocks, herds, and schools: A quantitative theory of flocking. *Physical review E*, 58(4):4828, 1998.

[77] S. B. Uvarov et al. Grasshoppers and locusts. a handbook of general acridology. volume i. anatomy, physiology, development, phase polymorphism, introduction to taxonomy. *GRASSHoppers and locusts. A handbook of general acridology. Volume I. Anatomy, physiology, development, phase polymorphism, introduction to taxonomy.*, 1966.

[78] T. Vicsek. A question of scale. *Nature*, 411(6836):421–421, 2001.

[79] T. Vicsek, A. Czirk, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, 75(6):1226, 8 1995.

[80] T. Vicsek and A. Zafeiris. Collective motion. *Physics reports*, 517(3-4):71–140, 2012.

[81] H. Wang, J. Peng, and S. Yue. A directionally selective small target motion detecting visual neural network in cluttered backgrounds. *IEEE Transactions on Cybernetics*, 50(4):1541–1555, 4 2020.

[82] X. Wang, F. Wang, Z. Nie, Y. Ai, and T. Hu. optiswarm: Optical swarm robots using implicit cooperation. *IEEE Sensors Journal*, 2022.

[83] L. Zhang, M. Lecoq, A. Latchininsky, and D. Hunter. Locust and grasshopper management. *Annu. Rev. Entomol*, 64(1):15–34, 2019.