Perception-Based Rewards for Robotic Shepherd Teams Maneuvering Split Flocks

Anonymous Author(s)
Affiliation
Address
email

Abstract: Underactuated system tasks, like shepherding passive agents using active coordinated robotic agent teams, require quick reactions and consistent perception and control. A recent learning-based solution demonstrated the agility required for such a task, but only accounted for single cohesive flocks. Noncontiguous flocks, on the other hand, can diffuse if not handled in a timely fashion. We address the disjoint flock case by defining novel reward schemes, based on the shepherds' visual observations. We show that policies trained on these rewards succeed at shepherding disjoint and fractious flocks to a goal region in a motion-efficient manner, and provide comparisons to state of the art learning-based and heuristic methods.

Keywords: Motion Planning, Robot Shepherding

1 Introduction

The shepherding problem asks how to efficiently get a group of reactive mobile agents (e.g. a flock of sheep) to a goal region by influencing sheep motion with a team of actively controlled and coordinated mobile guiding agents (e.g. shepherds). An interesting yet complicating feature of shepherding is that flocks may split in the process of being herded, or even start in separate clusters. Shepherds must then decide where to move to coalesce these groups of agents at a goal region. Smart, agile solutions to split flock shepherding must quickly allocate shepherds to tasks and plan motions for the shepherds, getting flocks to the goal while using minimal energy. Some heuristic methods attempt to collect separated sheep [1, 2, 3], but are not always efficient in terms of shepherd energy expended [4]. Learning based approaches have either used small numbers of herded agents [5, 6, 7] or used parameters and setups making split flocks unlikely [4]. Here we introduce a new perception-based reward scheme to [4], inspired in part by a recent occlusion-based heuristic method [3], allowing multiple-shepherd policies to effectively learn how to shepherd multiple split flocks of sheep quickly and efficiently.

2 Related Work

Shepherding has been explored with single [8, 9, 10, 1] and multiple [11, 12, 13, 4, 3] shepherd agents, environments with static obstacles [10, 11, 7], and in discrete [14, 15, 16] and continuous [10, 1, 12, 4] state and action spaces. In addition to herding animals, it is relevant for practical applications to fields such as security [17], crowd control [18], and environmental protection [19]. Few works directly address the problem of shepherding split flocks. El-Fiqi et al. examined two heuristic methods with flocks initialized in a variety of patterns [2]. Hu et al. assigned shepherds to split flocks by a coordination protocol [3]. Those methods assume global information about goal, sheep and and shepherds. We present and evaluate a limited-perception, learning-based solution to shepherding split flocks with shepherd-local observations a key component to reward.

Submitted to the 6th Conference on Robot Learning (CoRL 2022). Do not distribute.

Most shepherding works consider one of two types of dynamics for the passively guided agents. Reynolds' bird-like *boids* flocking dynamics [20] with added shepherd-avoidance terms have been studied in several shepherding works [10, 21, 7, 4]. Strömbom et al. presented dynamics in which sheep move away from shepherds, towards a local center of mass of sheep, and away from other sheep which get too close [1]. Strömbom et al. dynamics have been used heavily in shepherding problem research [12, 22, 6, 4]. We choose here to use the dynamics of Strömbom et al., as they reproduce sheep-like herding well and have a parameter n controlling flock cohesion.

3 Methods

This work extends a learning based shepherding approach [4]. Except where specified, parameters are as described in that work. Training ran for 300M timesteps, taking $\sim$ 2 days per model. Input forms a $1,536 \times 1 \times 3$ size array, consisting of three 512-ray lidar observations for the three observable types (sheep, other shepherds, and the goal region) concatenated, with a framestack of three. Network output is velocity. Training and experimental validation both use seeded random unbounded environments. A goal region of radius 2.5m is placed randomly within a 50m $\times$ 50m box. One or more flocks of sheep are placed around random points 10m to 20m from the goal center in Gausssian distributions with zero mean and standard deviation 1m. Shepherd robots are placed at random with a uniform distribution within the 50m $\times$ 50m box. Episodes of training or experimentation are 1000 time steps with a time step of 0.2s. Each shepherd agent senses and acts without explicit communication at every time step. Strömbom dynamics parameters are as in [4] with the exception of the cohesion parameter, n, which determines the nearest neighbors used to calculate the sheep attraction to the local center of mass. This value is set to 40 in training, is 40 in the varying number of starting flocks experiment, and varied in the varying flock cohesion parameter experiment. Two flocks and 1 to 6 shepherds are used in training.

We define three components of a shared global reward: an *occupancy reward,* r_{occupy}, and *distance penalty,* $r_{distance}$, which were both present in [4], and an *occlusion reward,* $r_{occlude}$, which is new to this work. Each reward has an associated weight:

$$r_{total} = w_{occupy} \cdot r_{occupy} + w_{distance} \cdot r_{distance} + w_{occlude} \cdot r_{occlude}, \tag{1}$$

where $w_{occupy} = 10$, $w_{distance} = -0.1$, as in [4], and $w_{occlude} = 0.1$, tuned empirically. r_{occupy} is the number of sheep in the goal region at a given time step divided by both the total number of sheep (fixed here to 100) and the total number of time steps in an episode (fixed here to 1000). $r_{distance}$ is $1 + d_i$, where d_i is the distance from the goal border to sheep i, summed across all sheep outside the goal. Finally, $r_{occlude}$ is calculated as:

$$r_{occlude} = \frac{\sum_p^P \sum_b^{B_p} \text{ 1 if } \exists\, s \in S \text{ s.t. } C(p,b,s) \text{ else } 0}{|\text{P}| \cdot T}, \tag{2}$$

$$\text{where } C(p,b,s) = V(p,b,g,s) \wedge D(p,b,s) \in [\alpha, min(D(p,b,g), I)]. \tag{3}$$

S is a set of sheep to be defined, P is the set of shepherds, B_p is the set of lidar beams of shepherd $p \in P$, g is the goal, T is the total timesteps of the episode, $V(p,b,g,s)$ is a Boolean function that is true when both g and $s \in S$ are visible along beam b, $D(p,b,x)$ is the distance from shepherd p to entity x's edge along beam b, α is the minimum distance from shepherd to sheep (2m), and I is the influence radius of the shepherd (10m in [4]). We define two variants of Occlusion reward: one variant, referred to as "Any Sheep", where S is the set of all sheep, and another variant, referred to as "Wild Sheep", where S is the set of sheep that have yet to reach the goal.

A shepherd satisfying eq. (3) with a sheep is properly driving that sheep towards the goal. Note that defining α is critical: without it, shepherds learned to go inside the flocks, disrupting them. The training curves seen in Figure 1 show that the Wild Sheep reward scheme converges at around 25M time steps, faster than the Any Sheep scheme at around 60M time steps. We hypothesize that this happens because Any Sheep is initially distracted from collecting more flocks by sheep that already entered the goal.

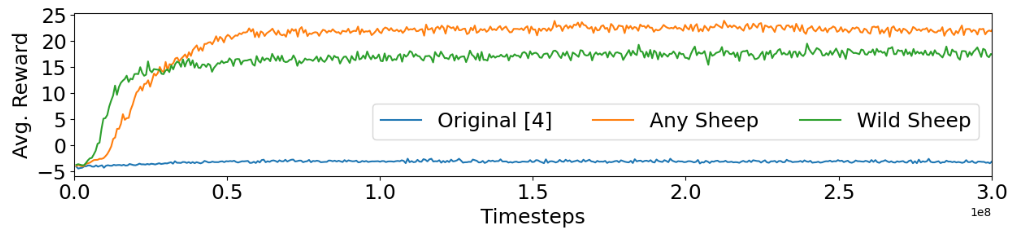


Figure 1: Mean reward curves for the original reward scheme of [4] and the Any Sheep and Wild Sheep reward schemes. Note that the range of rewards possible vary significantly between methods.

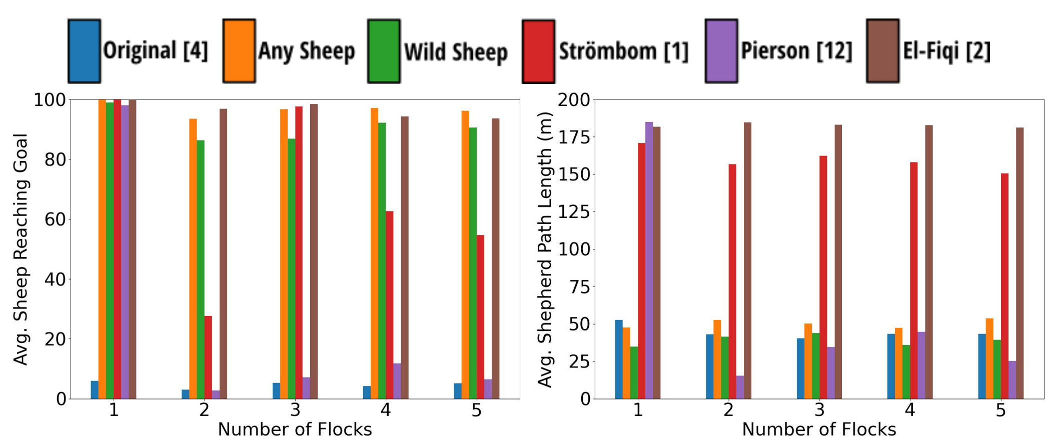
We compare against three state of the art heuristic shepherding methods chosen for being applicable to the Strömbom flock dynamics [1]. First, the Strömbom et al. heuristic (hereafter just Strömbom) switches between collecting distant sheep and driving a single coherent flock [1]. The Strömbom shepherding heuristic was originally defined for one shepherd only, but has been extended to multiple shepherds [23, 13]. Second, the Pierson and Schwager shepherding heuristic (hereafter just Pierson) forms an arc around a flock to drive the flock with unicycle-like dynamics [12]. Third, the El-Fiqi et al. shepherding heuristic (hereafter just called El-Fiqi) distributes multiple shepherds to distinct collecting and driving tasks while avoiding disturbing sheep unnecessarily [2]. The El-Fiqi algorithm has three important parameters which determine where and how shepherds travel: R1, R2 and R3. We set R1 to 5m, R2 to 4m, and R3 to equal the shepherd influence radius (10m). Finally, we additionally compare against the deep reinforcement learning model presented in [4] trained without new occlusion reward under otherwise identical conditions to the new models.

4 Experiments

4.1 Varying Number of Starting Flocks

In this experiment we evaluate the ability of the different shepherding methods to handle varying numbers of starting flocks. We vary the number of starting flocks from 1 to 5. Parameter n, which determines sheep attraction nearest neighbor count, is fixed at 40. There are three shepherds.

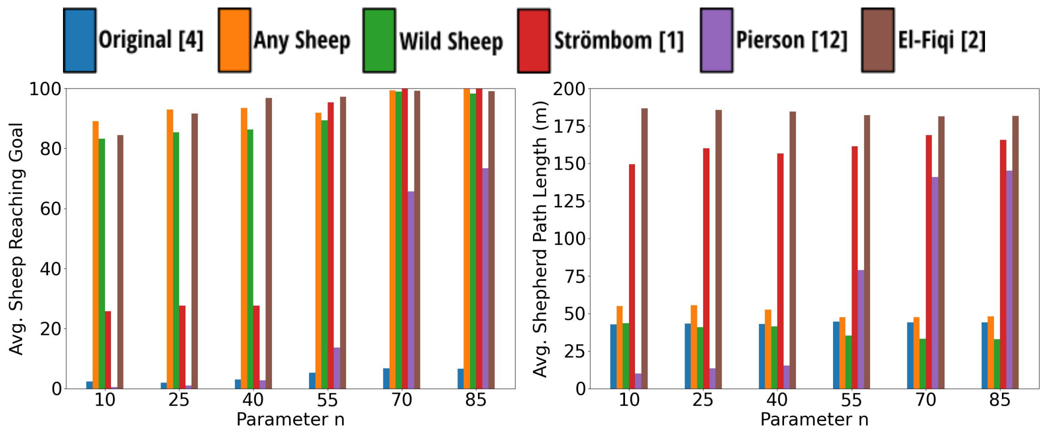
We find that the new perception-based rewards are critical to learning how to shepherd multiple flocks as well or better than existing heuristic methods. Figure 2 (a) shows the mean number of sheep arriving at the goal across 100 trials. Note that the original learning algorithm presented in [4] did not come up with a good policy for getting single or multiple flocks to the goal, or learn well with



(a) Mean number of sheep arriving at goal.

(b) Mean shepherd path lengths (meters).

Figure 2: Results for varied flock counts, 100 trials each.



(a) Mean number of sheep arriving at goal. (b) Mean shepherd path lengths (meters).

Figure 3: Results for varied cohesion parameter n, 100 trials each.

the training parameters (two flocks, $n = 40$). By contrast, Any Sheep and Wild Sheep nearly always get a single flock to the goal, and typically get all flocks to the goal with multiple flocks. These models perform comparably to the heuristic comparison algorithms for shepherding single flocks, and get high numbers of sheep to the goal in the case of multiple flocks. Moreover, as Figure 2 (b) shows, these models do so more efficiently (with travel distance a proxy for battery consumption) than the best performing heuristics, with average shepherd path lengths consistently about 100m less than well-performing heuristic methods. This shows that the models trained with perception-based rewards are effective at shepherding different numbers of flocks, successfully generalizing from the training experience of always two flocks, which is a marked improvement over [4].

4.2 Varying Flock Cohesion Parameter

Flocks that are not very cohesive may split or lose sheep. Here we evaluate the ability of the different shepherding methods to handle split flocks with more or less cohesiveness. We vary the parameter n from 10 to 85 in increments of 15. Greater n corresponds to greater cohesion. These experiments all start out with two flocks and three shepherds.

The new perception-based reward models are again effective and efficient at shepherding, generalizing to varied cohesion parameter n values. Figure 3 (a) shows the effectiveness of all methods reduces with less flock cohesion. However, even at $n = 10$, the models with perception-based reward perform as well or better than the best heuristic method. Figure 3 (b) shows that the shepherd path lengths, a metric of efficiency, are slightly higher for lower values of cohesion parameter n. However, in all cases where the heuristic methods on average delivered 50% or more sheep to the goal, average shepherd path lengths are about 100m shorter. The models which use perception-based reward for training generalize well to amounts of flock cohesion not encountered in training.

5 Conclusion

In this work presented a novel perception-based reward approach to shepherding groups of sheep to a goal, a task that requires reactive controls and continuous actions and observations. The policies learned are significantly more effective at guiding disjointed flocks than the state of the art learning method without the perception-based rewards, and are comparably as effective as state of the art heuristic shepherding methods. Moreover, they are significantly more efficient in terms of shepherd path lengths than the state of the art heuristics. The results show that the perception of goal occlusion is an effective tool for improving agile shepherding beyond what was previously possible.

References

[1] D. Strömbom, R. P. Mann, A. M. Wilson, S. Hailes, A. J. Morton, D. J. T. Sumpter, and A. J. King. Solving the shepherding problem: Heuristics for herding autonomous, interacting agents. *J. R. Soc. Interface*, 11(20140719):1–9, 2014.

[2] H. El-Fiqi, B. Campbell, S. Elsayed, A. Perry, H. K. Singh, R. Hunjet, and H. A. Abbass. The limits of reactive shepherding approaches for swarm guidance. *IEEE Access*, 8:214658–214671, 2020.

[3] J. Hu, A. E. Turgut, T. Krajník, B. Lennox, and F. Arvin. Occlusion-based coordination protocol design for autonomous robotic shepherding tasks. *IEEE Transactions on Cognitive and Developmental Systems*, 14(1):126–135, 2022.

[4] Y. Hasan, J. E. G. Baxter, C. A. Salcedo, E. Delgado, and L. Tapia. Reinforcement learning of coordinated shepherds for flock navigation and confinement. In *Proc. Workshop on Algorithmic Foundations of Robotics (WAFR)*, 2022.

[5] M. Baumann and H. Buning. *Learning shepherding behavior*. PhD thesis, University of Paderborn, 2016.

[6] H. T. Nguyen, T. D. Nguyen, M. Garratt, K. Kasmarik, S. Anavatti, M. Barlow, and H. A. Abbass. A deep hierarchical reinforcement learner for aerial shepherding of ground swarms. In *Proc. Int. Conf. Neural Information Processing (ICONIP)*, page 658–669, 2019. ISBN 978-3-030-36707-7.

[7] J. Zhi and J.-M. Lien. Learning to herd agents amongst obstacles: Training robust shepherding behaviors using deep reinforcement learning. *Robot. and Automat. Lett.*, 6(2):4163–4168, 2021.

[8] A. Schultz, J. Grefenstette, and W. Adams. RoboShepherd: Learning a complex behavior. In *Proc. Int. Symposium on Robot. and Autom.*, 1996.

[9] R. Pfeifer, B. Blumberg, J.-A. Meyer, and S. W. Wilson. *Robot Sheepdog Project achieves automatic flock control*, pages 489–493. Bradford Books, 1998.

[10] J.-M. Lien, O. Bayazit, R. Sowell, S. Rodriguez, and N. Amato. Shepherding behaviors. In *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, volume 4, pages 4159–4164 Vol.4, 2004.

[11] J.-M. Lien, S. Rodriguez, J. Malric, and N. Amato. Shepherding behaviors with multiple shepherds. In *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, pages 3402–3407, 2005.

[12] A. Pierson and M. Schwager. Bio-inspired non-cooperative multi-robot herding. In *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, pages 1843–1849, 2015.

[13] C. Aiba and K. Fujioka. A suggestion for effective shepherding models with two sheepdogs. In *Proc. Conf. IEEE Indust. Electronics Soc. (IECON)*, pages 77–81, 2020.

[14] A. S. Gadre. *Learning strategies in multi-agent systems-applications to the herding problem*. PhD thesis, Virginia Tech, 2001.

[15] C. K. Go, B. Lao, J. Yoshimoto, and K. Ikeda. A reinforcement learning approach to the shepherding task using SARSA. In *Proc. 2016 Int. Joint Conf. on Neural Networks (IJCNN)*, pages 3833–3836, 2016.

[16] M. Mahdavimoghaddam, A. Nikanjam, and M. Abdoos. Improved reinforcement learning in cooperative multi-agent environments using knowledge transfer. *Computing Research Repository (CoRR) in arXiv*, 2022.

[17] D. Shell and M. Mataric. Directional audio beacon deployment: An assistive multi-robot application. In *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, volume 3, pages 2588–2594 Vol.3, 2004.

[18] J. Kirkland and A. Maciejewski. A simulation of attempts to influence crowd dynamics. In *Proc. Int. Conf. on Systems, Man and Cybernetics*, volume 5, pages 4328–4333 vol.5, 2003.

[19] M. Fingas. *The Basics of Oil Spill Cleanup*. CRC Press/Taylor & Francis, Boca Raton, FL, 2013.

[20] C. W. Reynolds. Flocks, herds and schools: A distributed behavioral model. In *Proc. ACM SIGGRAPH*, page 25–34, 1987.

[21] S. Gade, A. A. Paranjape, and S.-J. Chung. Robotic herding using wavefront algorithm: Performance and stability. In *AIAA Guidance, Navi., and Control Conf.*, pages 1–16, 2016.

[22] J. Brulé, K. Engel, N. Fung, and I. Julien. Evolving shepherding behavior with genetic programming algorithms. *Computing Research Repository (CoRR) in arXiv*, 2016.

[23] K. Fujioka. Effective herding in shepherding problem in V-formation control. *Transactions of the Institute of Systems, Control and Information Engineers*, 31:21–27, 2018.