Efficient Q-Learning over Visit Frequency Maps for Multi-agent Exploration of Unknown Environments

Xuyang Chen, Ashvin N. Iyer, Zixing Wang, and Ahmed H. Qureshi
{chen4007, aniyer, wang5389, ahqureshi}@purdue.edu

OCTOBER 1 - 5, 2023
IEEE/RSJ International Conference on Intelligent Robots and Systems

PURDUE UNIVERSITY

CORAL LAB
Cognitive Robot Autonomy & Learning Lab

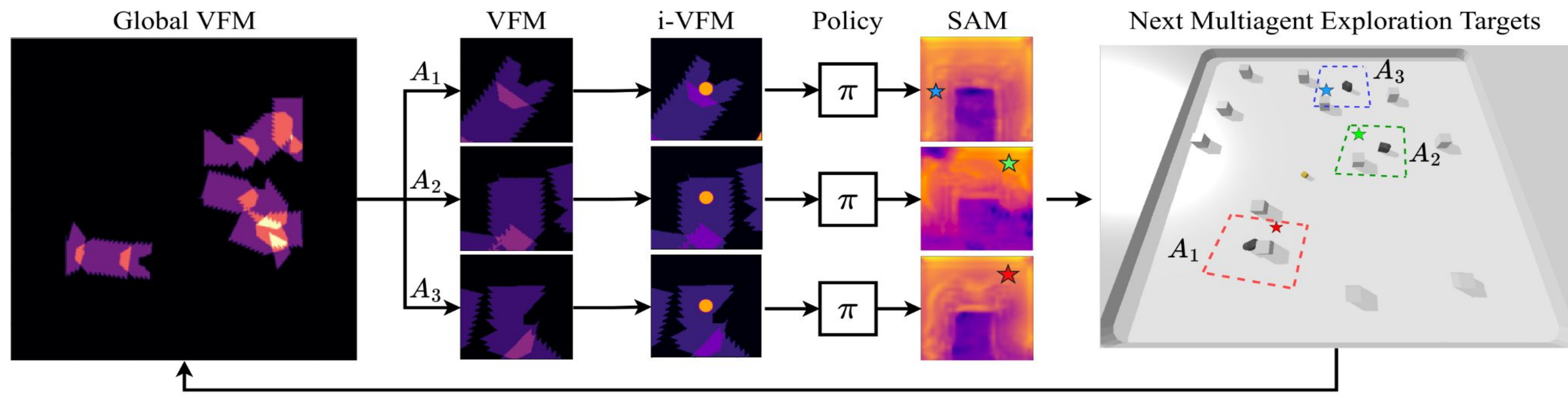
Introduction

We introduce a multiagent exploration system that is sensitive to bandwidth usage and generalizes to different environments and swarm sizes.

We also propose i-VFM, a novel state representation that uses only half the memory of past representations and encodes the same information.

Our methods demonstrate **zero-shot** generalization to a swarm setting despite no explicit multi-agent training.

Methodology

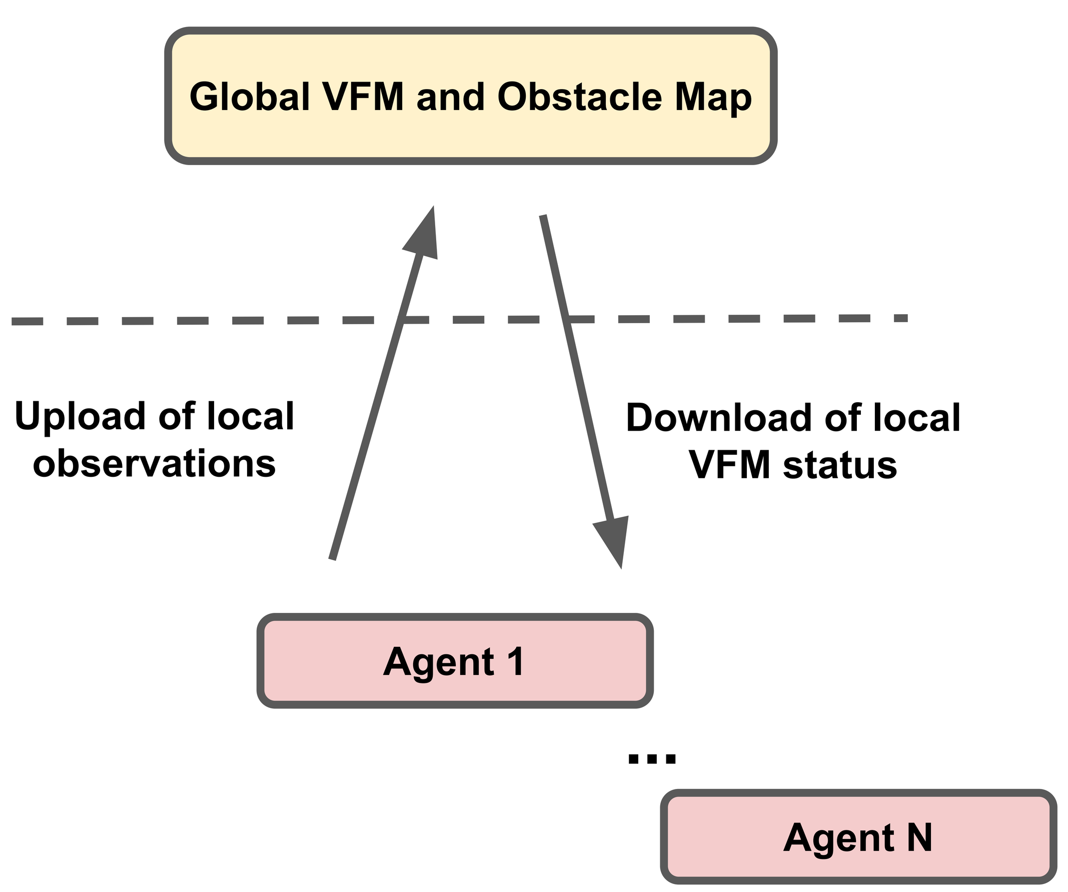


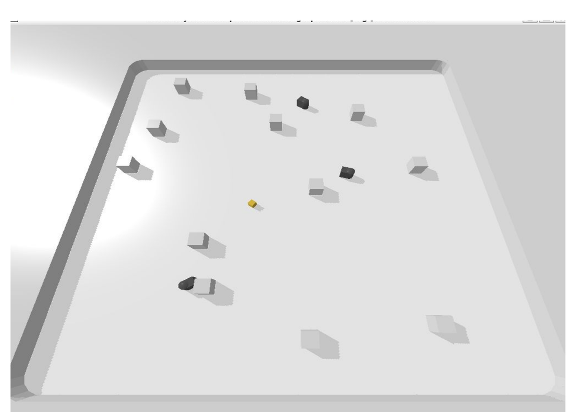
Algorithm 1 Multi-agent VFM-based exploration algorithm

```
procedure MULTI-AGENT VFM
    f_init(S)                          ▷ Initialize empty global Map
    while target not found do
        for i ← 1 to n do              ▷ Updating n agents
            p ← F_p(robot_i)           ▷ Get robot pose
            o ← F_obs(robot_i, rgbd)   ▷ Get RGB-D obs
            if target found then
                break
            end if
            S ← F_update(p, o, S)      ▷ Update global map
        end for
        for i ← 1 to n do
            p ← F_p(robot_i)
            S_local ← F_crop(S, p)     ▷ Get local map
            a_t ← π(S_local)           ▷ Get command from policy
            F_move(a_t)                ▷ Execute command
        end for
    end while
end procedure
```

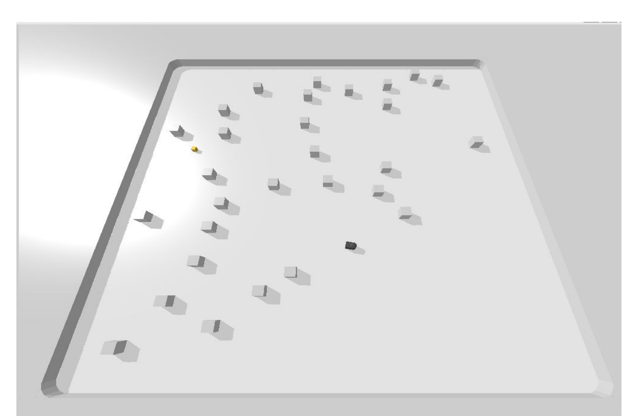
Experiments

Simulated Bandwidth Consumption for deployment on a remote exploration task

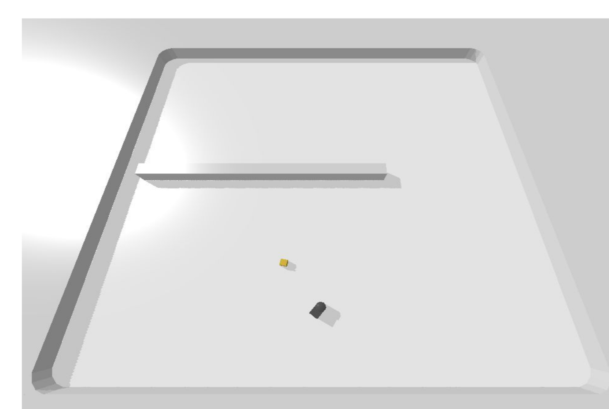




1x Size Arena



2x Size Arena



Large Divider Arena

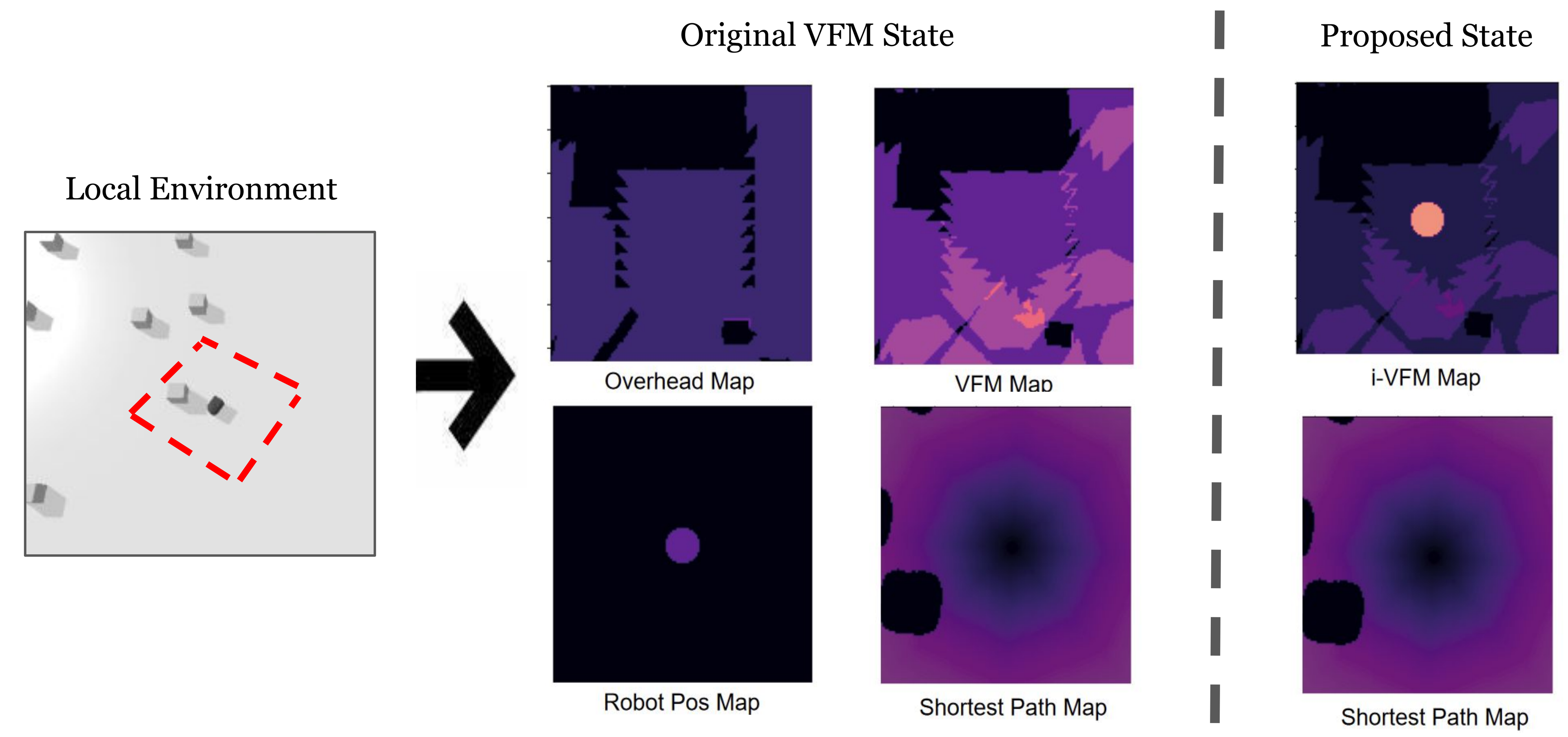
Key Metrics

- **Bandwidth** measures the total transmission rate over all agents in megabytes
- **Path Efficiency (PE)** measures the average new exploration per unit path
- **Repetitive Exploration Rate (RER)** measures the ratio of observed area that overlaps a previously observed area

EVALUATION PERFORMANCE IN DIVIDER AND 2X SIZED ARENAS

Trial	Policy	Agents	RER ↓	PE ↑	Steps ↓	Overlap ↓	Bandwidth ↓	Coverage ↑	Not Found ↓
Divider	VFM	One	0.492 ± 0.517	5914 ± 2250	31.8 ± 34.3	N/A	4.3 ± 4.6	0.480 ± 0.308	6 / 200
		Two	0.477 ± 0.473	5601 ± 2353	32.3 ± 33.0	0.1 ± 0.1	4.4 ± 4.4	0.466 ± 0.316	0 / 200
		Four	0.502 ± 0.496	5574 ± 2626	35.5 ± 35.6	0.2 ± 0.1	4.8 ± 4.7	0.473 ± 0.319	0 / 200
	i-VFM	One	0.584 ± 0.424	5524 ± 2305	33.6 ± 30.0	N/A	2.4 ± 2.1	0.460 ± 0.302	5 / 200
		Two	0.577 ± 0.529	5529 ± 2376	32.7 ± 30.8	0.1 ± 0.1	2.4 ± 2.2	0.443 ± 0.293	1 / 200
		Four	0.602 ± 0.651	5664 ± 2519	39.3 ± 41.8	0.2 ± 0.2	2.8 ± 3.0	0.472 ± 0.289	0 / 200
2x Arena	VFM	One	0.484 ± 0.471	6136 ± 1961	59.3 ± 62.9	N/A	8.1 ± 8.5	0.477 ± 0.315	1 / 200
		Two	0.507 ± 0.522	6206 ± 1909	62.4 ± 67.9	0.1 ± 0.1	8.5 ± 9.1	0.473 ± 0.319	1 / 200
		Three	0.422 ± 0.401	5995 ± 2223	54.4 ± 54.2	0.1 ± 0.1	7.4 ± 7.3	0.444 ± 0.297	1 / 200
		Four	0.419 ± 0.343	6230 ± 2078	54.8 ± 49.6	0.1 ± 0.1	7.5 ± 6.7	0.453 ± 0.306	1 / 200
		Five	0.375 ± 0.303	6867 ± 1804	49.9 ± 44.9	0.1 ± 0.1	6.8 ± 6.1	0.440 ± 0.290	3 / 200
	i-VFM	One	0.575 ± 0.427	5957 ± 1696	66.5 ± 59.5	N/A	4.8 ± 4.3	0.485 ± 0.287	1 / 200
		Two	0.550 ± 0.409	5735 ± 2084	66.8 ± 58.6	0.1 ± 0.1	4.8 ± 4.2	0.477 ± 0.310	1 / 200
		Three	0.494 ± 0.370	6072 ± 2062	61.8 ± 52.8	0.1 ± 0.1	4.5 ± 3.8	0.452 ± 0.293	0 / 200
		Four	0.491 ± 0.366	6083 ± 1992	61.9 ± 51.1	0.1 ± 0.1	4.5 ± 3.7	0.461 ± 0.290	3 / 200
		Five	0.420 ± 0.354	6421 ± 2188	56.1 ± 51.5	0.1 ± 0.1	4.1 ± 3.7	0.406 ± 0.292	3 / 200

State Representation



Conclusion

1. We present a multiagent exploration approach with a bandwidth usage that scales **linearly** with the explored area
2. We show that our agents generalize to the multi-agent scenario despite being trained alone and having no central decision making mechanism.
3. Our **i-VFM** state formulation performs comparably to the original VFM, while requiring half the bandwidth capacity.