University of Salford MANCHESTER

EPSRC Engineering and Physical Sciences Research Council

Artificially Intelligent Unmanned Aerial Vehicle

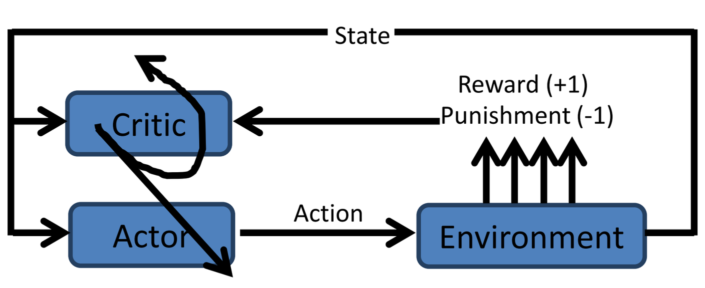
John Pinder, Prof Samia Nefti-Meziani, Dr Theo Theodoridis

Centre for Autonomous Systems and Advanced Robotics, University of Salford, Salford M5 4WT | J.Pinder@edu.salford.ac.uk

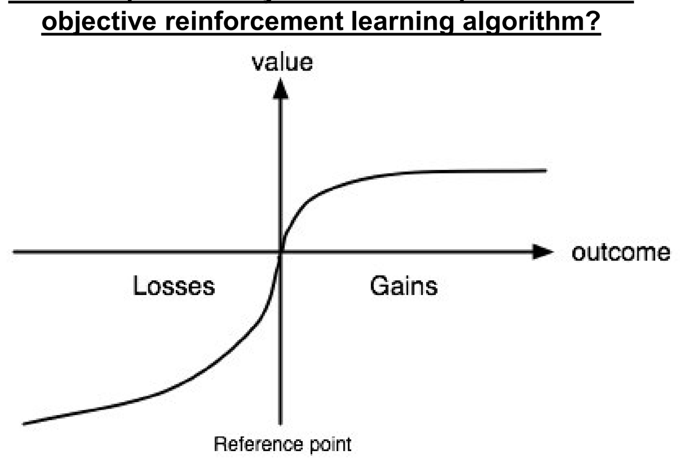
Introduction

Problem

GPFMORL Architecture



Can Prospect Theory be used to improve a multi-objective reinforcement learning algorithm?

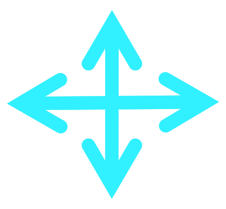


This research focuses on the problem of uncertainty handling during learning by multiple agents dealing in stochastic environments by means of Multi-Objective Reinforcement Learning (MORL). Reinforcement learning is a powerful mechanism for enabling agents to learn in an unknown environment, and most reinforcement learning algorithms aim to maximize some numerical value, which represents only one long-term objective. However, multiple long-term objectives are exhibited in many real-world decision and control problems; therefore, recently, there has been growing interest in solving multi-objective reinforcement learning (MORL) problems with multiple conflicting objectives.

Simulation Environment

The Reinforcement Learning method consists of 4 agents each responsible for moving in one of the for cardinal directions

- •North
- •South
- •East
- •West



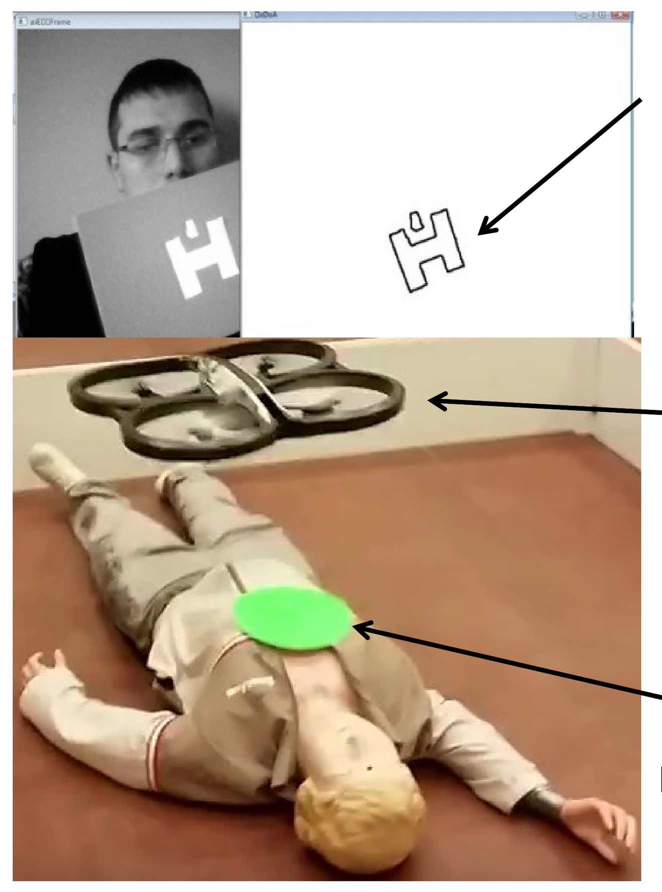
10x10 Windy Hills Grid World Simulation

1	2	3	4	5	6	7	8	9	10
11	**12**	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	**46**	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

The simulation environment is discretised into a 10x10 grid world where the aim is to learn the most efficient way to reach the goal (state46) from the start (state12) whilst satisfying the following 3objectives:

1. Reach Helipad
2. Avoid Boundaries
3. Avoid/Traverse Hills

Real World UAV Implementation

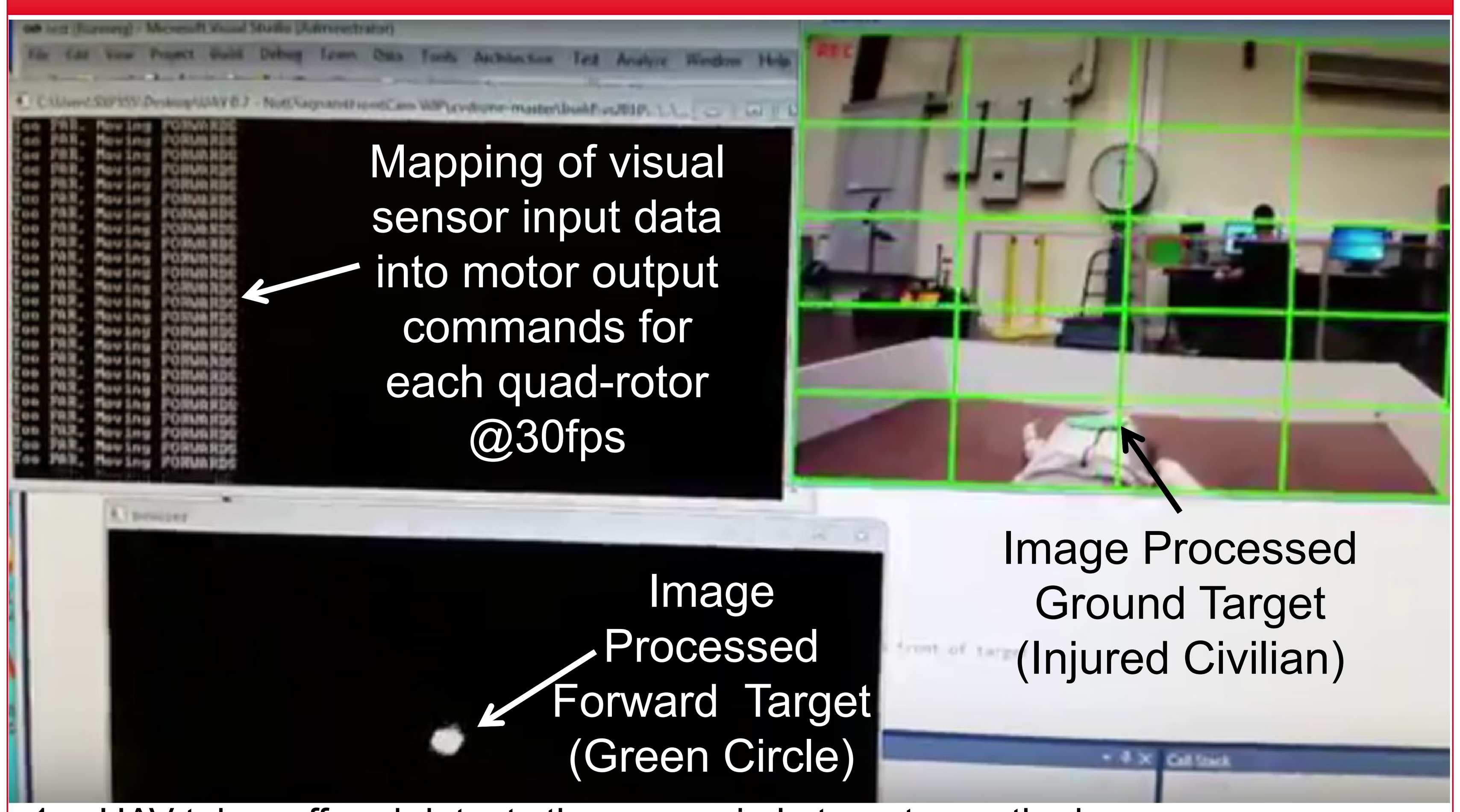


Helipad marker segmented from the background is used for determining altitude and Localisation

AR-Drone™ Wi-Fi Quad copter equipped with front and downward facing cameras, ultrasonic sensor, 6axis gyroscopes and accelerometers

System currently tracks a green circle to determine its current position however it could easily track peoples faces and use GPS in outdoor environments

UAV Search & Rescue Procedure



1. UAV takes off and detects the green circle target near the human.
2. UAV moves left, right, up or down to maintain a central position.
3. UAV begins to move forward until close enough to begin MORL.
4. UAV increases altitude to perceive a better field of view of the "H".
5. UAV tries actions when in different states to find optimal policy.
6. UAV now knows the best route to take from most states and lands.
7. UAV reports the GPS position of the human and brings crucial medical supplies

Methodologies

GPFMORL(PT)

Generalised Probabilistic Fuzzy Multi-Objective Reinforcement Learning is a hybrid learning algorithm that combines both the advantages of probabilistic theory with a fuzzy inference engine to handle multiple types of uncertainties. The single objective algorithm (GPFRL) has been improved to deal with multiple objectives by developing a Prospect Theory based scalarisation technique for the following three objectives in the proposed "Windy Hill Grid World" benchmark

1. Reach the goal (helipad) in as few steps as possible
2. Never become lost by stepping onto the boundary as few times as possible
3. Avoid the Hill obstacles unless its more efficient to traverse small hills

Multi-Modal Fuzzy Altitude Controller

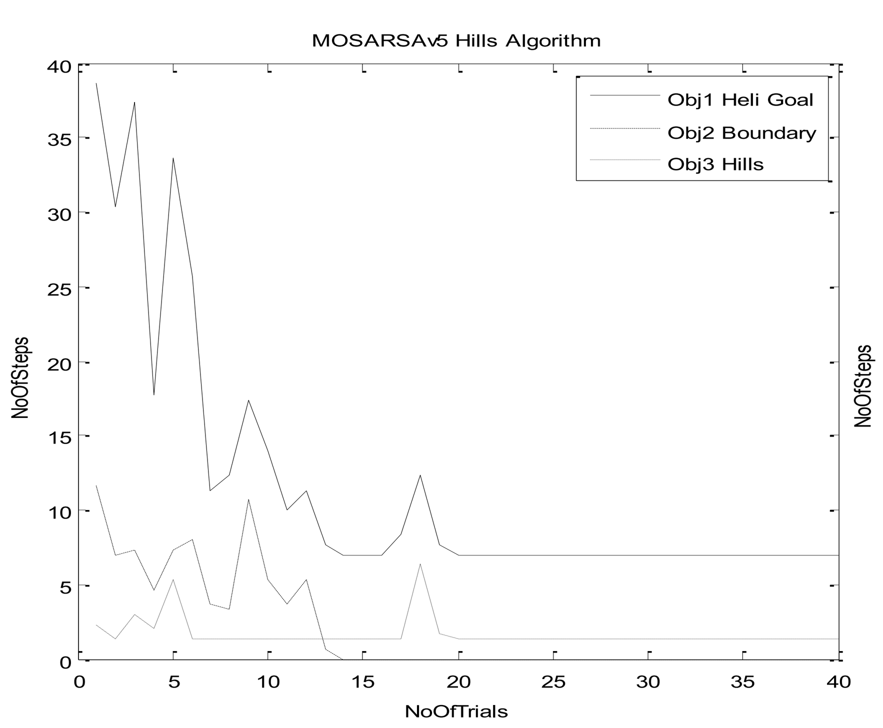
Using an Ultra-Sonic proximity sensor in conjunction with non linear visual area measurements of the helipad, a multi modal fuzzy controller is developed that enables smooth control of the drones altitude to ensure a consistent hovering height is always maintained, even in the event of traversing hills when necessary

Potential Field Heuristics (Reactive Behaviours)

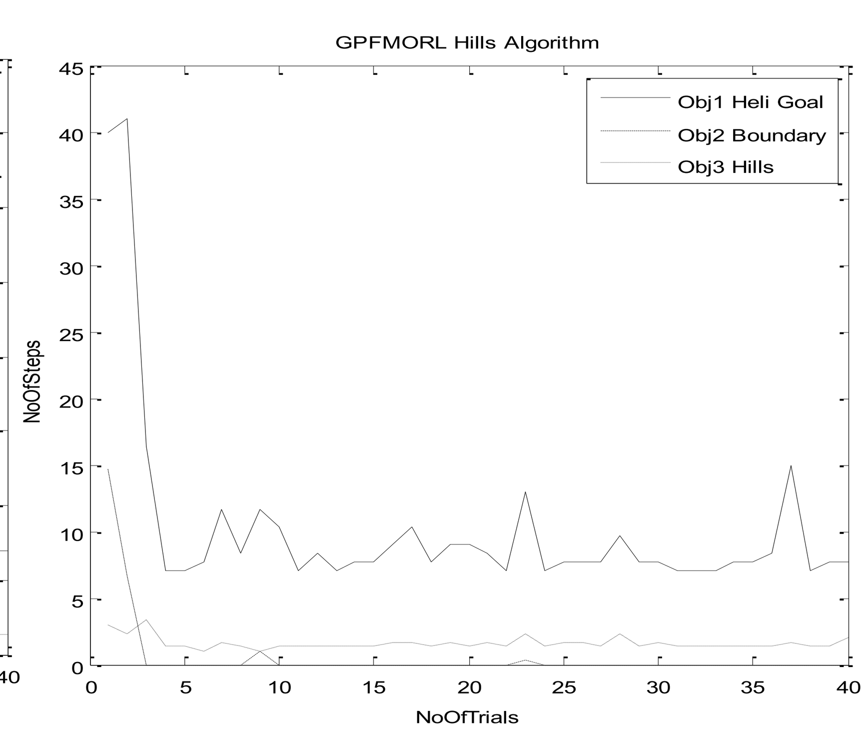
Due to the inherent field of view limitations when using a non-specialised lens for achieving localisation, further methods of control were used to ensure the helipad was always visible. Repulsive potential fields were added to the outer boundary sates of the perceived environment to ensure the drone never become lost. Attempts to use wide angle lens also improved performance slightly.

Algorithm Comparisons for 10x10 Grid World Deterministic Complex Environment

Multi Objecive Scalarisation Approach

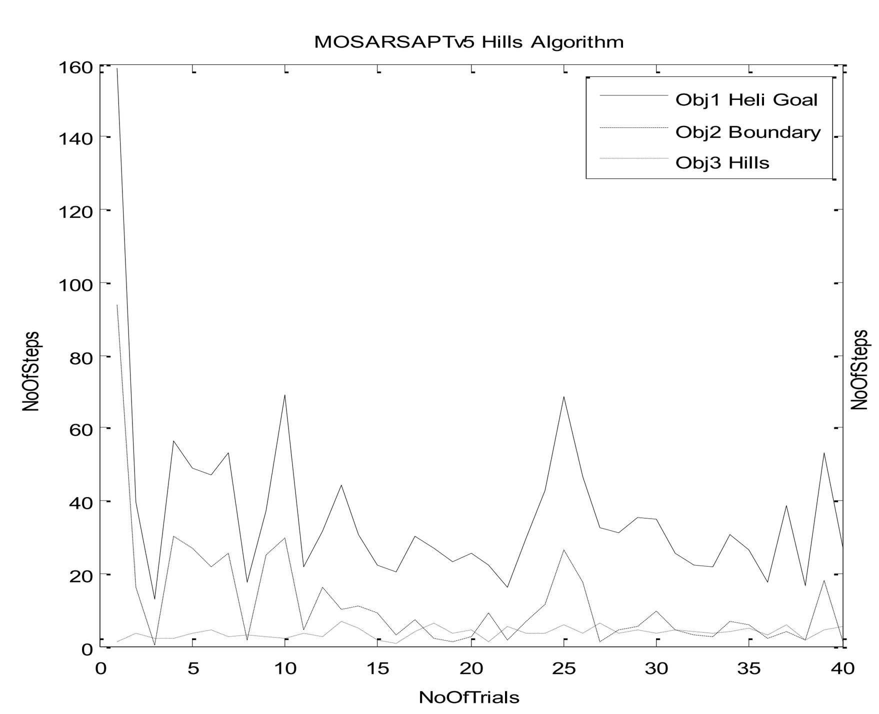


Av Bounds Count = 236
Av Hill Count = 58.6667

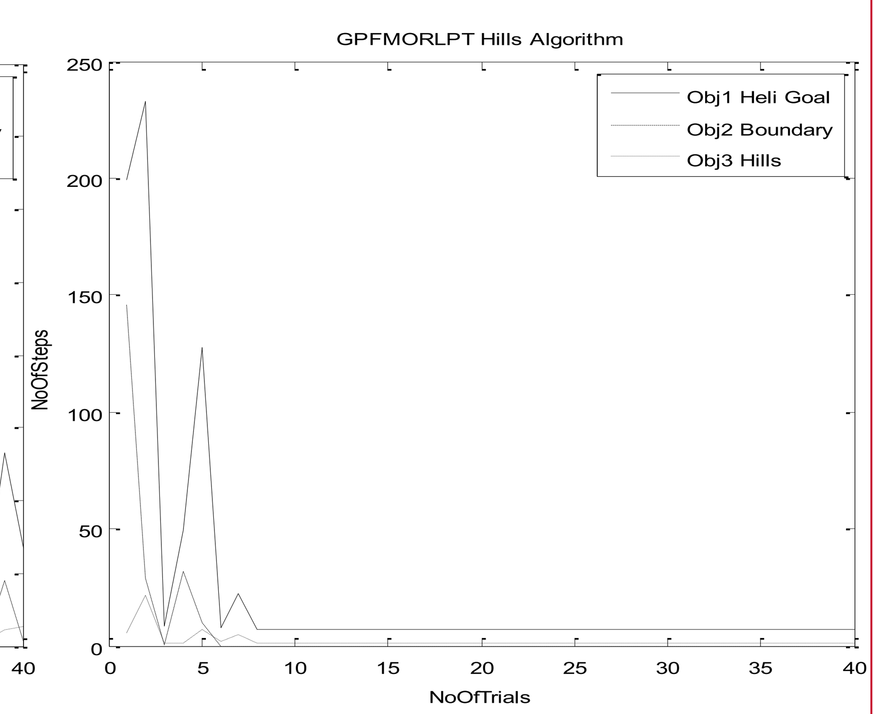


Av Bounds Count = 22.6667
Av Hill Count = 63

Multi Objective Prospect Theory Improvement



Av Bounds Count = 481.6667
Av Hill Count = 149.3333



Av Bounds Count = 217
Av Hill Count = 86.6667