1 Abstract

We present a simple, active-vision inspired scenario to test general approaches to problems in a Markovian world where there is a choice to be made about the observation. Three approaches – minimising entropy, maximising information gained and maximising expected future reward – were tested. These were compared to each other and to two choiceless approaches. Results show a slight preference for the maximised expected future reward, although all the considered approaches performed comparably.

2 Introduction

What to observe is a choice that almost all living beings make constantly. Most creatures have evolved significant flexibility in sensory apparatus which brings with it a great degree of freedom. Alongside this has developed an innate ability to effortlessly and subconsciously decide how best to focus these resources. Reproducing this choice would be convenient in a number of situations from focusing computational effort on a small window of data, choosing a test or sensor to devote resources to or simply swivelling a robotic head. In this WRITING we present a simple scenario fabricated to compare several approaches to solving this general problem.

3 Related Work

There is a significant body of work addressing related problems, but very little that moves beyond relatively *ad-hoc* approaches tailored for specific goals or environments. We are looking at something related to Bacjsy's definition of Active Sensing, although hoping to sidestep the foundation in image processing. [1]

Following on from Bajcsy there is a large body of work regarding autonomous robotic heads, or at least mobile cameras. This kind of a system seems ideal, as it provides clear choices to shape the stream of incoming data. Despite this very little work addresses the question of where to look in an abstract way, instead tending to code in desired behaviours by hand. Often this results in fixating on features particularly for simultaneous localisation and mapping. [4, 3] Other examples involve panning and zooming cameras [7]

A related concept is the idea of *salience* which arises in the context of image processing and neuroscience. Primarily founded on the eye movements of primates, salience seeks to provide some measure of which parts of a scene are 'interesting' or important to the observer in some way. Often a neural focus is taken, mimicking the mechanism of much primate visual processing. A classic example is the system presented by Itti and Koch which calculates several feature maps at a number of scales and combines them to form a 'salience map' which incodes the interestingness on a per-pixel basis. [6] These kinds of models do seem to produce results similar to that of humans and have found numerous applications and refinements. [8] .

Something that arises from the salience literature is a split between 'bottom up' and 'top down' salience. The bottom-up components are the raw features which stand out from the rest (essentially the aforementioned salience map) while the top down component tries to capture how specific goals change the relevance of certain features. [9]

4 Experimental Setup

The goal of the experiments was to construct a simulation which captured some of the more general issues. For simplicity everything was chosen to be discrete. It was necessary to have a situation where the choice different choices of observation could lead to significantly different observations. The resulting scenario is localisation of an autonomous agent on a known map. The world is small enough that beliefs can be handled directly, which removes the need for approximation techniques.

The world state is simply the agents' two dimensional coordinates on an $n \times n$ grid, giving n^2 possible states. Each cell is either black or white with a single green cell acting as the target, which the agents attempt to reach. The agents are free to move over any square – the colour only affects the observations.

1. Select sensor action a from set A
2. Submit a to world, receive observation $y \in Y$
3. Update beliefs according to sensor model $P(y|a, x)$
4. Choose manipulatory action $b \in B$
5. Advance state according to transition model $P(x_{t+1}|b, x_t)$
6. Advance beliefs according to transition model $P(x_{t+1}|b, x_t)$

Figure 1: Sequence of events within each time-step of the simulation

At each time step agents are able to make a single observation of one of the adjacent squares in either of the four cardinal directions: north, south east or west. These correspond to up, down, right or left respectively. Observations are drawn from the set of possible observations Y which simply contains representatives for black, white and green. There is some probability ρ_y (a parameter of the simulation) that the correct observation will be made, with the two possible incorrect observations having a corresponding $\frac{(1-\rho_y)}{2}$ chance of being returned. If the agent attempts to look off the edge of the board, it is treated as looking at a black cell.

After this the agents choose an action which manipulates the state, again one of the four cardinal directions. The result of this action is always to move the agent in the specified direction, except when they are attempting to move off the board in which case they simply remain in place. A summary of the steps that occur every time-step is given in figure 1

The desired behaviour was brought about by the predetermined value function $U \colon X \to \mathbb{R}$ (X is the space of possible world states). This function takes a set of coordinates $x = (x_1, x_2)$ and returns the negative Manhattan distance to the target –

$$U(x) = -(|x_1 - t_1| + |x_2 - t_2|)$$

where (t_1, t_2) is the position of the target square. This is designed to simulate the sort of value function that might be learned by an agent in a world with a negative reward for every square except the target, which should encourage locating the target as fast as possible.

4.1 Maps

Several different maps were used to test the different methods of choosing a sensor action. The first, seen in figure 2 is divided into 4 quarters each with a different pattern. The goal is always in the centre and the agents start in a randomly chosen corner. In order to reach the target, the agent must determine which corner it is in. This map should then favour methods which focus on localising quickly because as soon as the agent is localised sufficiently it can proceed to take the shortest route to the centre.

To test in a more general setting with similar pressures, a large number of random maps were also generated during testing. This was done by looping through every possible square and deciding with a certain probability whether it should be black or white. The target was placed afterwards. Adjusting the balance of black and white squares was done by changing the probability, which trades of the usefulness of a non-majority coloured square in localising with the chances of actually seeing one. Again the agents were placed at random corners of a 15 by 15 map with a target in the middle, figure 3 shows some examples of these maps.

5 Approaches

The following details the different approaches for choosing a sensor action that were tested. All agents use precisely the same mechanism for choosing movement actions – they all maintain a belief distribution and choose greedily the action with the highest expected utility according to their beliefs. The difference comes in the actions used to shape those beliefs over time and this has a substantial impact on the performance of the agents.

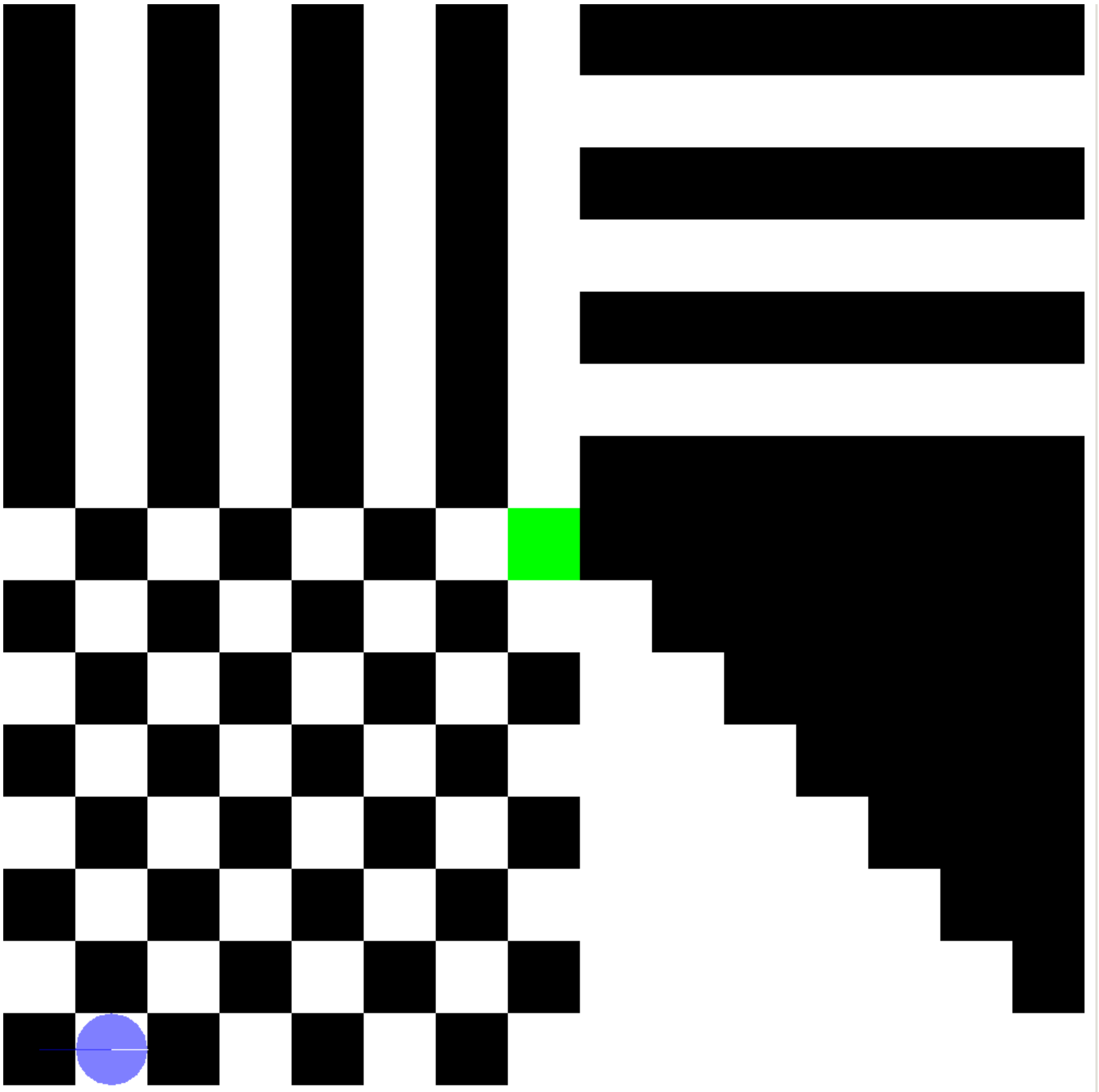
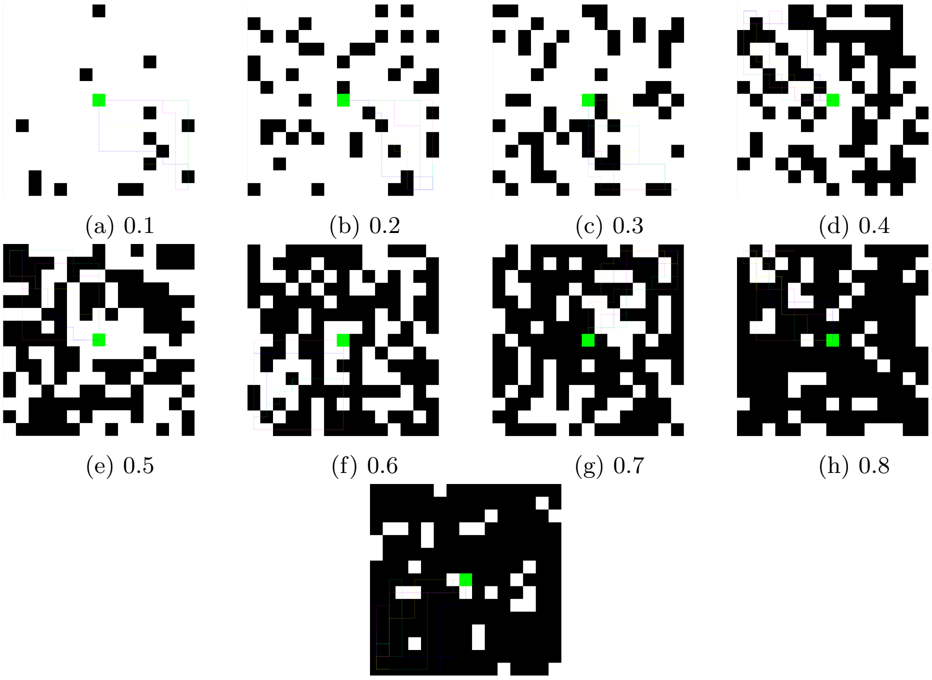


Figure 2: Map used for testing effect of changing sensor model. Agents start in a corner of the map, chosen randomly at the same trial. Trials were run concurrently with the group of agents all starting in the same corner. This map should favour those that favour localisation as the best action is unclear until the agent is aware which corner it is in.


(a) 0.1

(b) 0.2

(c) 0.3

(d) 0.4

(e) 0.5

(f) 0.6

(g) 0.7

(h) 0.8

(i) 0.9

Figure 3: Examples of random maps with differing probabilities

5.1 Basic Approach

The simplest approach tested was to always look straight ahead. This can be seen as simulating a robot with a limited, fixed vision system – the direction of the sensor is always the same as the direction of the agents previous motion.

5.2 Random

A second simple approach was also tested alongside the more considered options: choosing a random direction to look at every step.

5.3 Minimise Entropy

This is the obvious approach if the goal is to localise the agent. In the case the goal of the sensor action is to provide new beliefs with the lowest entropy possible. Thus for each possible observation, we have to calculate the expected entropy over the resultant belief states. This takes the form of a sum over all possible observations weighted by the probability of receiving that observation. As the probability of an observation depends on the state of the world, we have to sum over our current beliefs in order to determine the observation probability. This yields:

$$a^* \leftarrow \arg\min_{a \in A} \sum_{y \in Y} \sum_{x \in X} P(y|a, x) \, \mathrm{Bel}(x) H(X')$$

where $H(X')$ is the entropy (using a base 2 logarithm) of the belief distribution over X as it would be after having taken action a and received observation y.

5.4 Bayesian Surprise

The motivation behind this approach is to maximise the amount of information gained by an observation. This is achieved by maximising the relative entropy between the prior and the posterior beliefs after having made an observation. [5, 2]

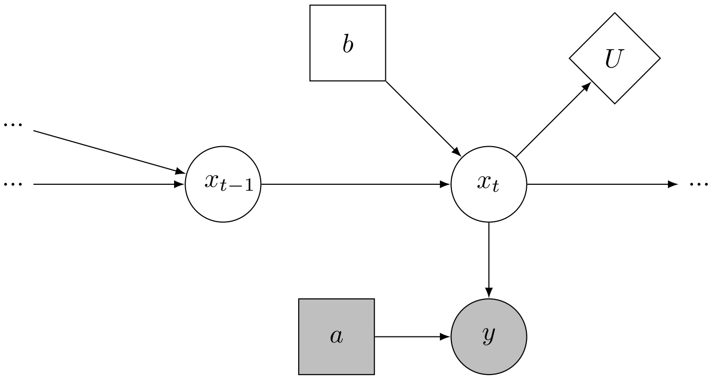


Figure 4: A model for the decision making process. Square nodes are choices while the diamond node is the reward, dependent on the state of the world. First a is chosen, then y is observed then b is chosen then a reward is gathered and then the state x advances, dependent on b and the current state of x.

In order to choose the best sensor action we again compare the expected values across each possible action giving

$$a^* \leftarrow \arg\max_{a \in A} \sum_{y \in Y} \sum_{x \in X} P(y|a,x) \operatorname{Bel}(x) K(\operatorname{Bel}(X), \operatorname{Bel}(X|a,y)).$$

$K(\operatorname{Bel}(X), \operatorname{Bel}(X|a,y))$ is the relative entropy, or Kullback-Leibler divergence, of the prior beliefs about X and where they would stand after an update, defined as

$$K(\operatorname{Bel}(X), \operatorname{Bel}(X|a,y) = \sum_{x \in X} \operatorname{Bel}(x) \log \frac{\operatorname{Bel}(x)}{\operatorname{Bel}(x|a,y)}.$$

using a base e logarithm;

There are distinct similarities between the surprise approach and minimising entropy. In particular the choice is governed entirely by the agents beliefs and should always work to localise the agent quickly. Minimising entropy does this explicitly, but maximising surprise has similar behaviour in this world as the only actions that are going to provide much surprise will be those with a reasonable chance of spotting a landmark and thus helping the agent localise, reducing entropy of its beliefs.

5.5 Goal-oriented

This technique extends the approach used to choose the best movement action into the realm of the sensor action. In order to derive this, it helps to look at figure 4 which shows a graphical model of the process. The goal of the sensor action is then to maximise the reward obtained by the next choice. Thus we want to choose a^* such that it maximises the expected utility of the next action:

$$a^* \leftarrow \arg\max_{a \in A} \operatorname{E}_y \left[U_{b^*|a,y} \right]$$

$U_{b*|a,y}$ is the utility of behaving optimally (taking action b^*) after having taken sensor action a and receving observation y. With discrete observations, as we have here, the expected value is calculated as $\operatorname{E}_y \left[U_{b^*|a,y} \right] = \sum_{y \in Y} \operatorname{Bel}(y|a) U_{b^*|a,y}$. To determine $\operatorname{Bel}(y|a)$ we have to look at the sensor model and our beliefs in the state of the world: $\operatorname{Bel}(y|a) = \sum_{x_t \in X} P(y|a,x_t) \operatorname{Bel}(x_t)$.

The utility of some action b given a sensor action a and an observation y, denoted $U_{b|a,y}$, is the expected value (across beliefs in the world state) of the reward gathered after having taken action b. This then depends on the beliefs in the world state, the state transition model and the reward/value function $U(x)$. Specifically, the calculation is given as:

$$U_{b|a,t} = \sum_{x_t \in X} \sum_{x_{t+1} \in X} \operatorname{Bel}(x_t|a,y) P(x_{t+1}|x_t,b) U(x_{t+1}).$$

The earlier equation calls for b^*, the optimal choice of b. It is optimal to choose the value which maximises the above quantity, hence we have a final expression for the choice of sensor action

$$a^* \leftarrow \arg\max_{a \in A} \sum_{y \in Y} \sum_{x \in X} P(y|a,x) \operatorname{Bel}(x) \max_{b \in B} \sum_{x_t \in X} \sum_{x_{t+1} \in X} \operatorname{Bel}(x_t|a,y) P(x_{t+1}|x_t,b) U(x_{t+1}).$$

This is a computationally complex sum to calculate, especially in worlds with a large number of possible states. Fortunately some savings can be made – during each time stepb^* is dependent solely on the sensor action a and the observation y. As both are discrete we can precompute a table containing b^* and $U_{b^*|a,y}$ for each (a, y). This brings the complexity of the calculation more in line with the surprise approach.[1]

6 Results

The first set of trials were conducted on the map in figure 2. The average path length was measured across 2500 trials for ten different values of ρ_y, the probability of a correct observation which were evenly spaced between zero and one. Results can be seen in figure 5a for the 15 by 15 map, which has a shortest possible path of 14. As the uncertainty of the observation increases, the simulation breaks down. When there is almost nothing to be learned from the observations ($\rho_y = 0.3$) none of the approaches was able to complete a single trial within 300 steps. Up until $\rho_y = 0.5$ the goal-driven agent leads the pack. It is quite close but this approach seems to outperform several others as the uncertainty around the observation starts to increase.

This might make sense. As the observations become less useful, it becomes much harder for the agents to reduce the entropy of their beliefs or glean much information from the observation and as such the choices of the approaches based on these ideas is not much better than choosing randomly. The goal-oriented agent is not as worried about its beliefs as a whole, it will favour choices which might let it believe the goal is imminently attainable. Intuitively this should lead to more sensible decision making in situations where localisation is difficult and to better performance in circumstances where precise localisation is not strictly necessary to achieve the goal.

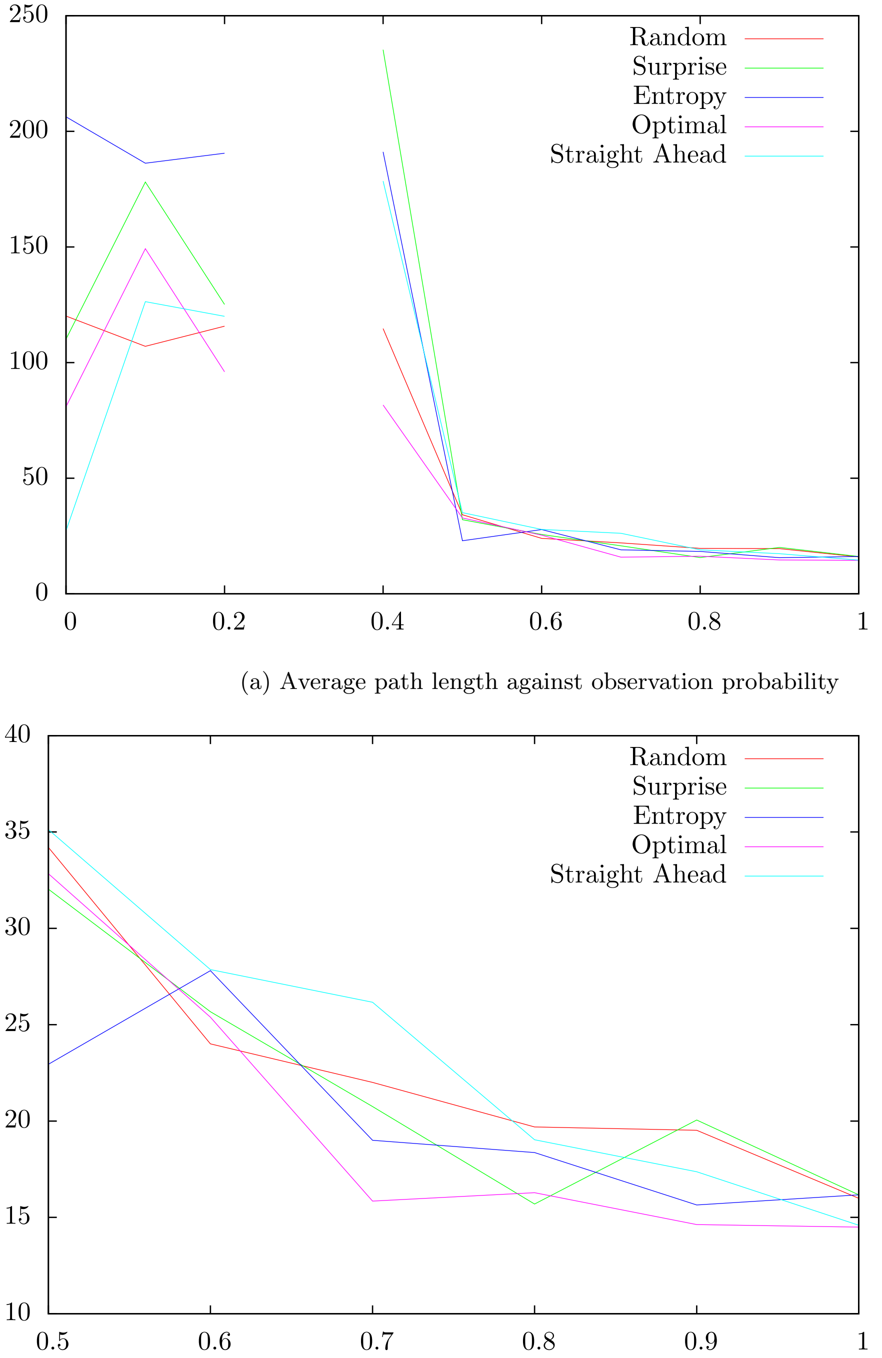
Subsequent trials took place on randomised maps and were highly inconclusive, even when results were taken over 10000 trials (on a smaller map due to time constraints). Figure 6 shows the average lengths plotted agains the probability of an individula square in the map being black. The individual approaches are inseparable and incredibly noisy. There is a clear upwards trend as the likelihood of very many white squares decreases. This is evident on the larger map as well in figures 7 and 8. This is to be expected given that any time an agent looks off the end of the map it is treated as looking at a black square. This is like having an implicit border of black squares, which means that when the majority of the map is white, being in a corner is a very good position in which to localise. Conversely when most of the squares are black there is very little assistance anywhere, so we would expect the agents to perform worse, which they all appear to.

7 Discussion

The results of the simulations show no clear winner. This seems remarkable because the goal-oriented approach should be behaving optimally at each time step. Given that it is often inseparable from a random choice, this does not appear to be the case. In fact this is not entirely unexpected as the fact that it behaves optimally given its current beliefs for a single time step does not mean it is behaving optimally across mutliple. The current goal-oriented algorithm for sensor choices only considers the impending manipulatory action. It is quite probable that were it to look further ahead performance would increase.

While a simple proposal, this is a difficult problem to solve especially given the computational effort required in computing a single step at a time in a complicated world. A possible avenue is to view the problem as a Partially Observable Markov Decision Process (POMDP) with the complication that there are two sets of actions with alternating availability. If this were feasible then there is a significant body of techniques for learning policies which would help to approximate truly optimal behaviour over time. Nevertheless this is still not a simple solution – POMDPs are very much an active research area and typical solutions tend to be rather complicated.

[1] Without precomputation, this approach is in $\Theta(aybx^3)$ where a, y, b, x are the number of sensor actions, observations, manipulatory actions and world states respectively. Precomputing pulls it into $\Theta(aybx^2)$, which is a potentially significant gain in a complex world. By contrast calculating the surprise is in $\Theta(ayx^2)$ as it is an argmax over A of a sum over Y of a sum over X of the KL divergence over X which is another sum over X. Minimising entropy is in precisely the same class of complexity as the only difference is taking the logarithm of the probability rather than the logarithm of a ratio of probabilities.



(a) Average path length against observation probability

(b) The same plot, focusing on probabilities 0.5 and greater

Figure 5: Results of 15 by 15 quarter map, averages across 2500 trials.

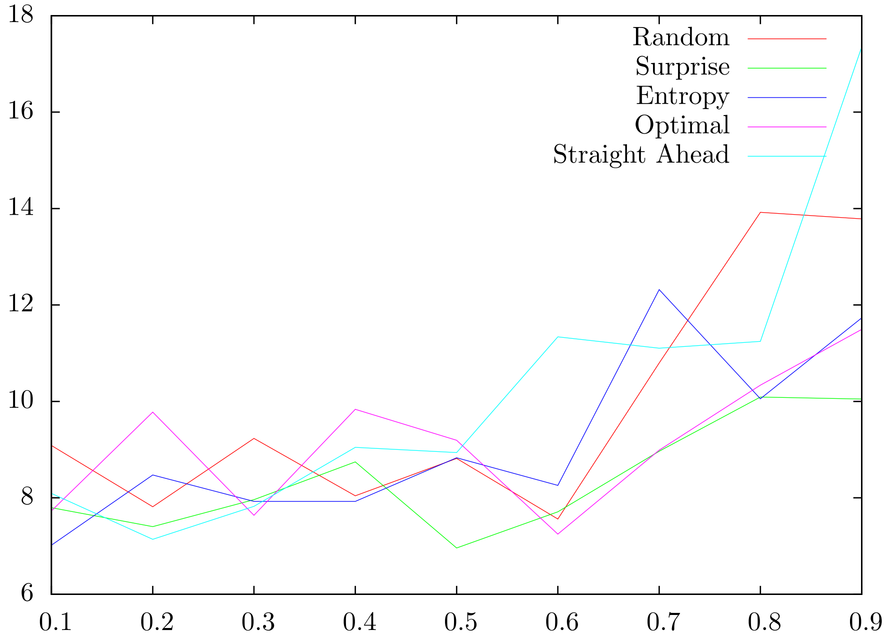


Figure 6: Average path length over 10000 trials on an 7 by 7 random map. Shortest possible path was 6 steps

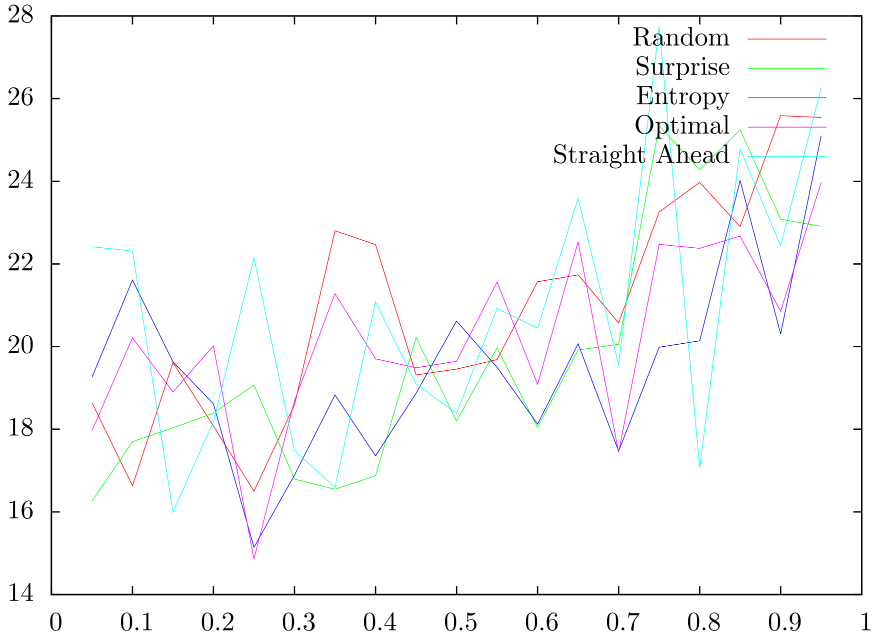


Figure 7: Average path length against map generation probability parameter. 15 by 15 map with target in the middle and agents starting on a corner, observation probability 0.8, 2500 trials at each point.

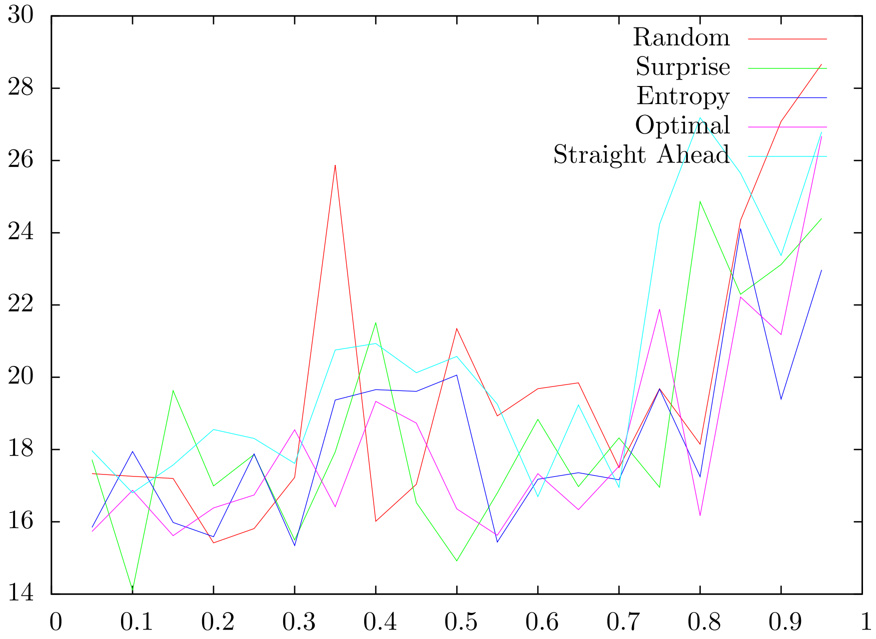


Figure 8: Average path length against map generation probability parameter. 15 by 15 map with target in the middle and agents starting on a corner, observation probability 0.9, 2500 trials at each point.

Interestingly, in most of the situations tested, the relatively naive approaches of trying to minimise entropy of beliefs or maximise information gained from the observation performed favourably. On a step-by-step basis these were comparable to the goal-oriented approach and on the trial that showed some separation between approaches both were competitive. This suggests that in the right circumstances (all the maps favoured rapid localisation) they might provide a reasonable approximation, despite ignoring the possible rewards entirely. Of course this has only been shown in one particularly contrived scenario, the goal-oriented approach may generalise better to diverse scenarios and also provides a better framework for extension into multi-step planning as it is constantly attempting to maximise the expected reward.

The two information-theoretic approaches also differ from the more decision-theoretic goal-oriented approach in that they are substantially easier to compute (although still not particularly simple). All are polynomial in the number of world states and with the pre-computed table their complexity is the same but the goal oriented approach loses as in order to populate the table it must iterate every pair of sensor actions and observations. This is a significant handicap and a good reason to favour one of the less complex approaches.

8 Conclusion

While the goal-oriented approach was a narrow winner on an ordered map, particularly under greater uncertainty, on a series of randomised maps there was no clear winner. This is likely due to the lack of ability to plan ahead which hamstrings all of the approaches outlined here. The fact that the optimal (in the decision theoretic sense) performed similarly to some fairly simple, more task specific approaches is encouraging in that it implies that the full optimal planning process might well able to be shortcut or estimated in some way.

References

[1] R. Bajcsy. “Active perception”. In: *Proceedings of the IEEE* 76.8 (1988), pp. 966–1005. ISSN: 0018-9219. DOI: 10.1109/5.5968.

[2] Pierre Baldi and Laurent Itti. “Of bits and wows: a Bayesian theory of surprise with applications to attention”. In: *Neural Networks* 23.5 (2010), pp. 649–666.

[3] A.J. Davison and N. Kita. “3D simultaneous localisation and map-building using active vision for a robot moving on undulating terrain”. In: *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on.* Vol. 1. 2001, I–384–I–391 vol.1. DOI: 10.1109/CVPR.2001.990501.

[4] A.J. Davison and D.W. Murray. “Simultaneous localization and map-building using active vision”. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.7 (2002), pp. 865–880. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2002.1017615.

[5] Laurent Itti and Pierre F Baldi. “Bayesian surprise attracts human attention”. In: *Advances in neural information processing systems.* 2005, pp. 547–554.

[6] Laurent Itti, Christof Koch, and Ernst Niebur. “A Model of Saliency-Based Visual Attention for Rapid Scene Analysis”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20.11 (1998), pp. 1254–1259. ISSN: 0162-8828. DOI: http://doi.ieeecomputersociety.org/10.1109/34.730558.

[7] Silviu Minut and Sridhar Mahadevan. “A Reinforcement Learning Model of Selective Visual Attention”. In: *Proceedings of the Fifth International Conference on Autonomous Agents.* AGENTS ’01. Montreal, Quebec, Canada: ACM, 2001, pp. 457–464. ISBN: 1-58113-326-X. DOI: 10.1145/375735.376414. URL: http://doi.acm.org/10.1145/375735.376414.

[8] Derrick Parkhurst, Klinton Law, and Ernst Niebur. “Modeling the role of salience in the allocation of overt visual attention”. In: *Vision Research* 42.1 (2002), pp. 107 –123. ISSN: 0042-6989. DOI: http://dx.doi.org/10.1016/S0042-6989(01)00250-4. URL: http://www.sciencedirect.com/science/article/pii/S0042698901002504.

[9] JeremyM. Wolfe. “Guided Search 2.0 A revised model of visual search”. English. In: *Psychonomic Bulletin & Review* 1.2 (1994), pp. 202–238. ISSN: 1069-9384. DOI: 10.3758/BF03200774. URL: http://dx.doi.org/10.3758/BF03200774.