A Methodology for Intelligent Path Planning

Suman Chakravorty, and John L. Junkins,
Department of Aerospace Engineering
Texas A&M University, College Station

***Abstract*— In this work we present a methodology for intelligent path planning in an uncertain environment. Examples would include a mobile robot exploring an unknown terrain or a UAV navigating enemy territory while avoiding radar detection. We show that the problem of path planning in an uncertain environment, under certain assumptions, can be posed as the adaptive optimal control of an uncertain Markov decision process. The strategy for path planning then reduces to computing the control policy based on the current estimate of the environment, also known as the "certainty equivalence" principle in the adaptive control literature. Further we show that the path planning problem, as formulated in this paper, possesses special structure which can be used to significantly reduce the dimensionality of the problem. Finally we apply this methodology to the problem of path planning of a mobile rover in an unknown, unstructured terrain.**

I. Introduction

In this work we present a methodology for "intelligent path planning" in an uncertain environment. The state space of any path planning problem can be expressed as the ordered pair (s, q) where s represents the system state and q represents the state of the environment at the state s. For example, in the case of a robot exploring an unknown terrain, s corresponds to the (x, y) co-ordinates of the robot and q corresponds to the height of the terrain $z(x, y)$ at the point (x, y). This methodology is not limited to only robotic path planning and is equally applicable to UAV navigation and multi-spacecraft imaging problems [1], [2]. The goal of the path planning strategy is to use all available information about the environment, until the current time instant, in order to plan the "best possible" path. We show that the planning problem can be modeled as a Markov decision process whose transition probabilities are uncertain. Then, we show that the "intelligent path planning" paradigm reduces to the adaptive optimal control of a Markov decision process.

There has been substantial research in the adaptive control of controlled Markov Chains, or Markov Decision Processes, in the past two decades and it is usual to distinguish the different methods into "direct" and "indirect" adaptive control. In indirect adaptive control, the transition probabilities of the underlying Markov chain are estimated and the control is applied based on the most recent estimate of the transition probabilities [3], [4], [5]. This is known as the so-called "certainty equivalence principle". The "direct" approach to stochastic adaptive control falls under the category of "reinforcement learning" methodologies wherein the optimal control is calculated directly without resorting to estimating the transition probabilities of the underlying Markov chain [6], [7], [8]. These methods can be further distinguished into "Q-learning" and "adaptive-critics", please refer to [6], [7] for more details. Underlying all these methods is Bellman's "principle of optimality" or Dynamic Programming, which is a methodology for sequential decision-making under uncertainty [9]. In this work, we show that the path planning problem can be reduced to the adaptive optimal control of a Markov decision process and thus the above methodologies can be applied to the same. In fact, we take the indirect approach to adaptive control since we are also interested in mapping the environment. Moreover, we show that the path planning problem has special structure which can be exploited to significantly reduce the dimensionality of the problem.

After three decades of research and development, robot motion planning is a mature area of research. Various different methodologies have been devised for the design of collision-free path planning of mobile robots in known environments [10]. In the past decade, there has been an increasing interest in the case when the environment in which the robots are operating is partially unknown. The uncertainty in the environment is treated on a deterministic worst case [11], [12] or in a probabilistic average case basis [19]. In "probabilistic robotics", there has been substantial research in the localization of a mobile robot while simultaneously mapping the environment [13], [14]. The expectation maximization algorithm and extended Kalman filters have been used to accomplish this objective and these methods are distinguished into the concurrent mapping and localization (CML) [15] and simultaneous localization and mapping (SLAM) [17], [18] algorithms respectively. However, the primary focus of this work [13]-[18] has been the estimation of the robot position while simultaneously mapping the operating environment and consequently, not much attention has been paid to the control aspect of this problem. In [19], [20], [21], a game-theoretic framework is proposed for robotic motion planning wherein the uncertainties inherent in the robotic system and the environment are modeled probabilistically. Then, the authors resort to Bellman's principle of optimality [9] in order to tackle the motion-planning problem. However, even a probabilistic model of a robotic system and its environment suffers from uncertainties, i.e., the transition probabilities of the underlying markov processes are uncertain. Thus, the problem of control of such a robotic system needs to take this uncertainty into account and as such can be posed as a stochastic adaptive control problem. We make this connection in the current work. We further note that the methodology presented in this paper remains valid for path planning problems in which the environment can be adequately modeled through a stationary random process. Moreover, we show that the intelligent path planning problem has very special structure which can be exploited to significantly reduce the dimensionality of the problem. In [22], the problem of planning the motion of a robot in the presence of randomly appearing obstacles, in a manufacturing setting, is studied. In the above mentioned work, there is a probabilistic component to the cost function,

dependent on the random obstacles. Assuming certain a priori distributions for the random obstacles, the expected value of the probabilistic cost component is sequentially estimated using a Bayesian estimator and the control is applied based on the Principle of optimality. This work is the closest in spirit to the stochastic adaptive control approach as presented in this paper, since the control is applied based upon the most current estimate of the environment. However, the same problem can be transformed into a stochastic adaptive control problem by introducing an environment state and thus absorbing the uncertainty in cost due to the environment into uncertainty regarding the transition probabilities of the underlying Markov chain.

The original contributions of the current work are as follows. We identify that the problem of "intelligent path planning" can be reduced to the adaptive optimal control of an uncertain Markov decision process. Then, we use the "certainty equivalence principle" [3] in order to adaptively/intelligently change the control policies for path planning in an unknown environment. This is equivalent to applying the control which is optimal with respect to the recent most estimate of the transition probabilities of the underlying Markov chain, i.e., the most current knowledge of the environment. Further, we show that the problem has a very special structure which can be exploited to significantly reduce the dimensionality of the problem. The rest of the paper is organized as follows. In section 2, we model the path planning problem and show that under certain assumptions it is equivalent to the adaptive control of a Markov decision process. In section 3, we outline the salient properties of the path planning problem. In section 4, we apply the methodology to the problem of a mobile rover negotiating an unknown unstructured terrain. In section 5, we draw conclusions on the current work and discuss avenues for further research.

II. Modeling of the Path Planning Problem

In this section, we model the path planning problem. We show that the process underlying path planning in an unknown environment can be modeled as a controlled Markov chain or a Markov decision process (MDP). In the case of an unknown environment, the transition probabilities of the underlying process are not known and have to be estimated while planning paths in the environment in an "intelligent" fashion. This corresponds to adaptive control of an MDP. In the following because of paucity of space, we shall omit the proofs of the results.

The state space of the path planning problem is defined by the ordered pair (s, q) where s represents the system state and q denotes the environment state, the environment variable being dependent on the system state. Throughout this development, we shall assume that the system state s is known perfectly while the environment state, q, might be noise corrupted. For example:

- In the case of robotic exploration of unknown terrain, s corresponds to the (x, y) co-ordinates of the robot and q corresponds to the height of the terrain $z(x, y)$ at the point (x, y).
- In the case of a UAV navigating enemy territory while avoiding radar detection, the s variable corresponds to the ground tack (x, y) of the UAV while q corresponds to the binary valued variable indicating the presence or lack of radar coverage at the point (x, y).
- In the case of a robot navigating on a shop floor, s represents the position of the robot on the shop floor and q is a binary valued variable which represents the presence or absence of an obstacle at the point s.
- In the case of multi-spacecraft imaging, the s variable represents the positions of the (u, v) coverage of the component spacecrafts in the imaging system while q represents the frequency content of the image at the point s [2].

In fact, we note that if the environment at any system state can be modeled as a stationary random process, the methods developed in this paper can be applied to such a system. For example, if the presence of obstacles at a certain point on the manufacturing shop floor can be modeled as a stationary random process, then the method is applicable.

For simplicity of treatment, we shall assume that the state space is finite. Let the number of system states be N and let the number of environment states be D. Thus, the total number of states in the state space is ND. Let the set of control actions be denoted by U and let the number of control actions possible be denoted by M. We shall denote any particular control action by u. We make the following Markovian assumption about the system.

$\mathcal{A}$ **2.1:** Let $\mathcal{F}^t = \{(s_0, q_0), u_0, \cdots, (s_{t-1}, q_{t-1}), u_{t-1}\}$ represent the history of the process till time t. Then

$$p((s_t, q_t)/\mathcal{F}^t) = p((s_t, q_t)/(s_{t-1}, q_{t-1}), u_{t-1}). \quad (1)$$

Then, we can represent the dynamics of the system probabilistically through the probability density function (pdf) $p((s_t, q_t)/(s_{t-1}, q_{t-1}), u_{t-1})$, where the function represents the probability of the system making a transition to state (s_t, q_t) given that the current state is (s_{t-1}, q_{t-1}) and the current control is u_{t-1}.

We make the following assumptions about the system:

$\mathcal{A}$ **2.2:** We assume that the current system state, s_t, is independent of the past environment state, q_{t-1}, i.e.,

$$p(s_t/(s_{t-1}, q_{t-1}), u_{t-1}) = p(s_t/s_{t-1}, u_{t-1}). \quad (2)$$

The above assumption is satisfied if the system state is "controllable". Though this assumption is strictly not true for any choice of control, nevetheless it can be satisfied through an appropriate choice of control.

$\mathcal{A}$ **2.3:** The current environment state, q_t is dependent only on the current system state, s_t, i.e.,

$$p(q_t/s_t, (s_{t-1}, q_{t-1}), u_{t-1}) = p(q_t/s_t). \quad (3)$$

The above assumption corresponds to the fact that the environment is "uncontrollable" through the system.

Then, we have the following result for the system transition pdf:

Proposition 2.1: Under assumptions $\mathcal{A}$2.2-$\mathcal{A}$2.3, the transition pdf $p((s_t, q_t)/(s_{t-1}, q_{t-1}), u)$ can be written as:

$$p((s_t, q_t)/(s_{t-1}, q_{t-1}), u_{t-1}) = p(s_t/s_{t-1}, u_{t-1})p(q_t/s_t). \quad (4)$$

The pdf $p(s_t/s_{t-1}, u)$ quantifies the localization and the control uncertainties inherent in the system and is assumed to be known beforehand. This is a reasonable assumption if we have good knowledge of the sensors and the system. The

terrain/ environmental uncertainty $p(q_t/s_t)$ is unknown and successive estimates $p_t(q/s)$ are made of the uncertainty as the planning proceeds to completion. The question is then, "how to use this increasing information of the terrain in order to plan better paths?". To answer this question, we make the following assumption about the system:

$\mathcal{A}$ **2.4:** The goal of path planning can be framed as an infinite horizon discounted stochastic optimization problem, i.e., given the initial state (s_0, q_0), the control policy $\pi^*(s_0, q_0) = \{\pi_1, \pi_2, ...\}$ corresponding to the optimal path is given by

$$\pi^*(s_0, q_0) = argmin_\pi E(\sum_{t=1}^{\infty} \beta^t c((s_t, q_t), (s_{t-1}, q_{t-1}), \pi_{t-1})/(s_0, q_0)), \quad (5)$$

where $c((s_t, q_t), (s_{t-1}, q_{t-1}), u_{t-1})$ is a positive pre-defined cost that the system incurs in making the transition from state (s_{t-1}, q_{t-1}) to (s_t, q_t) under the control action u_{t-1}, $E(.)$ denotes the expectation operator and $\beta < 1$ is a given discount factor.

Note that the above assumption also defines the goal of path planning in an uncertain terrain, namely that the average value of the total discounted cost incurred by the system needs to be minimized. Then, using the Bellman principle of optimality [6], it can be shown that the optimal policy is stationary, i.e., the optimal control is independent of time, and that the optimal control at a state (s, q), $u^*(s, q)$, is given by the optimality equation [6]:

$$u^*(s,q) = argmin_u \sum_{(r,p)} p(r/s,u)p(p/r)[c((r,p),(s,q),u) + \beta J^*(r,p)], \quad (6)$$

where $J^*(r, p)$ is the optimal cost-to-go from the state (r, p). Moreover, J^* satisfies the following fixed point equation:

$$J^*(s,q) = \min_u \sum_{(r,p)} p(r/s,u)p(p/r)[c((r,p),(s,q),u) + \beta J^*(r,p)]. \quad (7)$$

Adaptive Control involves controlling an uncertain system while simultaneously obtaining better estimates of the system parameters. We envisage the problem of path planning as one of adaptive control of an uncertain Markov decision process, (i.e., the transition probabilities of the Markov decision process are not known). In such a scenario, the strategy of adaptive control is to use the policy that is optimal with respect to the current estimate of the system, since it corresponds to the current knowledge of the system that is being controlled and is referred to as the "certainty equivalence principle" in adaptive control [3].

Let $T = \{t_1, t_2, \cdots, t_k, \cdots\}$ denote the set of all times at which the control policy is updated during the path planning. Let the updated control policy at time instant t_k be denoted by $u_k(s, q)$. Let $p_t(q/s)$ denote the estimated environmental uncertainty at the time t. Then, the control update at time $t_k \in T$, $u_k(.)$, using the principle of optimality and the "certainty equivalence principle", is given by

$$u_k(s,q) = argmin_u \sum_{(r,p)} p(r/s,u)p_t(p/r)[c((r,p),(s,q),u) + \beta J_k(r,p)], \quad (8)$$

where

$$J_k(s,q) = \min_u \sum_{(r,p)} p(r/s,u)p_t(p/r)[c((r,p),(s,q),u) + \beta J_k(r,p)]. \quad (9)$$

Let $T_s = \{t_{s_1}, \cdots, t_{s_k}, \cdots\}$ denote the times at which observations at the system state s are made. Let $1_{(s,q)}$ denote the indicator function of the event that the environment state q is observed at the system state s. We estimate the transition probabilities $p_t(q/s)$ using the Monte-Carlo method:

$$p_t(q/s) = \begin{cases} \frac{1}{N}\sum_{k=1}^{N} 1_{(s,q)}^{t_{s_k}}, & if \quad t \in T_s \\ p_{t-1}(q/s), & otherwise \end{cases} \quad (10)$$

In the above equation, N is such that $t = t_{s_N}$.

In the development till this point, we have outlined the model for path planning in an unknown environment. We have shown that the problem can be modeled as the adaptive control of a Markov Decision Process. Then, using the principle of optimality and the "certainty equivalence principle", we have outlined a methodology for updating the path planning policies. Moreover, we have also shown that the transition probabilities of the underlying MDP have a special structure and outlined a Monte-Carlo scheme to progressively estimate the same. In the next section, we shall list a few salient properties of the path planning problem.

III. Properties of the Path Planning Problem

In this section, we shall enumerate a few salient properties of the path planning problem. In particular, we shall show that under the Monte-Carlo estimation scheme defined in eq. 10, the sequence of control policies converges to the true optimal policy, i.e., the optimal policy with respect to the true transition probabilities. Also, we shall show that the special structure of the problem results in significant reduction of the dimensionality of the problem.

Recall the estimation scheme enumerated in eq. 10. Let the sequence of control policies be denoted by $\{u_1, u_2, \cdots, u_n, \cdots\}$. We make the following assumption about the system under the sequence of policies $\{u_1, u_2, \cdots, u_n, \cdots\}$.

$\mathcal{A}$ **3.1:** Every system state s is visited infinitely often under the sequence of control policies $\{u_1, u_2, \cdots, u_n, \cdots\}$.

The above assumption is the equivalent of the "persistent excitation" condition in adaptive control which seeks to tackle the "exploration-exploitation" trade off inherent in every adaptive control problem [3].

Now, we state the following results which establish that the cost-to-go function associated with the sequence of control policies $\{u_1, u_2, \cdots, u_n, \cdots\}$, converges to the optimal cost-to-go function. Note that $p(q/s)$ refers to the true environmental uncertainty and $p_t(q/s)$ refers to an estimate of the same at time t. In the following development it shall be implicit that

all the convergences are w.p. 1 and we shall drop the explicit reference to convergence with probability 1. Now we can state the following result which establishes the convergence of the cost-to-go functions to the optimal cost-to-go function.

Proposition 3.1: Under assumptions $\mathcal{A}$2.2- $\mathcal{A}$3.1, the cost-to-go functions $J_k(s,q) \rightarrow J^*(s,q)$ as $k \rightarrow \infty$.

The above proposition establishes the result that the sequence of cost-to-go functions converges to the optimal cost-to-go function. Next, we shall simplify the optimality equations based on the special structure of the path planning problem, which allows us to significantly reduce the dimensionality of the problem.

Consider the optimality equation:

$$J(s,q) = \min_u \sum_{(r,p)} p((r,p)/(s,q),u)[c((r,p),u,(s,q)) + \beta J(r,p)]. \quad (11)$$

Let

$$\sum_{(r,p)} p((r,p)/(s,q),u)c((r,p),u,(s,q)) = \bar{c}(s,u,q). \quad (12)$$

Noting that

$$p((r,p)/(s,q),u) = p(r/s,u)p(p/r), \quad (13)$$

it follows that

$$J(s,q) = \min_u [\bar{c}(s,u,q) + \beta \sum_r p(r/s,u)\bar{J}(r)], \quad (14)$$

where

$$\bar{J}(s) = \sum_q p(q/s)J(s,q). \quad (15)$$

Note that

$$u^*(s,q) = arg\min_u [\bar{c}(s,u,q) + \beta \sum_r p(r/s,u)\bar{J}(r)]. \quad (16)$$

Thus, we can conclude from the above development that we need only have an average value of the cost-to-go function $J(s,q)$, at the system state s, namely $\bar{J}(s) = \sum_q P(q/s)J(s,q)$, in order to be able to evaluate the optimal control at any state (s,q). This significantly reduces the dimensionality of the problem. However, there still remains the problem of estimating the average cost-to-go vector $\bar{J}(s)$. In order to answer this question, note that

$$\bar{J}(s) = \sum_q p(q/s) \min_u \{\bar{c}(s,u,q) + \beta \sum_r p(r/s,u)\bar{J}(r)\}. \quad (17)$$

Hence, $\bar{J}$ is the fixed point of the "average" dynamic programming operator $\bar{T} : \Re^N \rightarrow \Re^N$, defined by the following equation:

$$\bar{T}\bar{J}(s) = \sum_q p(q/s) \min_u \{\bar{c}(s,u,q) + \beta \sum_r p(r/s,u)\bar{J}(r)\}, \forall\, s. \quad (18)$$

The following proposition states that the "average DP operator", $\bar{T}$, is a contraction mapping and thus the average cost-to-go function can be evaluated using the method of successive approximations.

Proposition 3.2: The average DP operator, $\bar{T}$, is a contraction mapping in $\Re^N$ under the ∞ norm.

Recall that N denotes the number of system states, D denotes the number of environment states and M denotes the number of control actions. Consider the following fixed point equations:

$$\bar{J}(s) = \sum_q p(q/s)\{\min_u (\bar{c}(s,u,q) + \beta \sum_r p(r/s,u)\bar{J}(r))\}, \quad (19)$$

$$J(s,q) = \min_u [\bar{c}(s,u,q) + \beta \sum_{(r,p)} p((r,p)/(s,q),u)J(r,p)]. \quad (20)$$

The "average DP" iteration involves N^2MD operations every iteration of the successive approximation algorithm, while the full DP operator involves N^2MD^2 operations. Thus, exploiting the structure of the path planning problem results in reduction of the number of operations by a factor of D. Moreover, the "average DP" approach allows the cost-to-go vector to be stored in N elements while the full DP approach would require ND elements. This is a saving of a factor of D. This is not surprising since the structure of the problem dictates that the environment state q is "uncontrollable" and thus, in order to evaluate the optimal control at any given exploration state, only an average value of the cost-to-go at the system state, (averaged over all possible environment states), is required.

In the next section, we present a simulation study where the methodology developed so far is used for the path planning of a mobile rover navigating an unknown terrain.

IV. Simulation Study

In this section, we apply the methodology developed so far in this work to the problem of path planning of a mobile robotic rover exploring an unknown terrain. The terrain is unstructured and is estimated as the maneuvers proceed to completion. Also, it is assumed that the terrain does not change throughout the duration of the path planning maneuvers.

In this example the system state, s, is the (x,y) grid points and the environment variable q is the height of the terrain, z, at the grid point. We discretized the (x,y) plane into a 20x20 grid, i.e., the number of system states, N, is equal to 400. The height of the terrain was discretized into 10 equispaced intervals, i.e., in this case the number of environment states, D, is equal to 10. There were four allowable control actions, $N/E/S/W$, which corresponded to the rover going North/ East/ South/ West to the adjacent grid point. Note that the above choice of control actions makes the system state "controllable". Thus, in this example, the number of controls, M, was equal to 4.

We assumed that there was no uncertainty in control or localization, i.e., $p(r/s,u)$ corresponds to a point measure. This is an idealized situation, however, the method remains valid even for situations when there is uncertainty $p(r/s,u)$.

We assumed that we had no prior knowledge of the terrain that was to be negotiated. Thus, the initial environment uncertainty, $p_0(q/s)$, was modeled as a uniform distribution. We assumed that the observations, q, at s, were corrupted by Gaussian noise whose variance was proportional to the actual height, z, of the terrain at the grid point $s = (x_i, y_j)$. The rover made observations (s,q) only when it was at the grid point s. However, note that one of the salient features of the terrain path planning problem is that in order to make observations of the terrain at the state, s, the rover does not need to be at that state. Currently, we are incorporating these

features into the methodology.

The incremental cost function $c((r,p),(s,q),u)$ penalized path length of the maneuver and also, the time-to-go to the goal point, which was assumed to be the (1,1) grid point. The respective penalties were in the ratio 10:1, i.e., the rover was penalized ten times as much for climbing steep terrain with respect to minimizing the incremental path length. The goal state was a cost-absorbing state, i.e., once the rover reached the goal state, it would stay there forever without incurring any cost.

The control policies were updated after every trip to the goal point. The initial state was randomly varied. This is necessary in order that the rover be able to fully explore the terrain, i.e., the "persistent excitation" assumption is satisfied. Another approach could be to have a suit of way points distributed uniformly across the terrain, in some order, and letting the rover find optimal routes from one way point to the next. This would also insure that the terrain is explored sufficiently. As a final note, this phenomenon is a manifestation of the "exploration-exploitation" trade off in adaptive control [3].

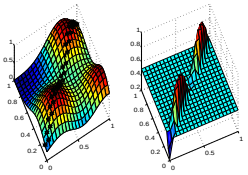

Fig. 1. Performance after 1 trial

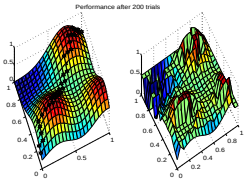

Fig. 2. Performance after 200 trials

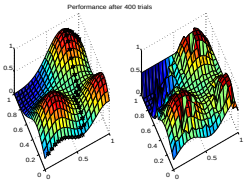

Fig. 3. Performance after 400 trials

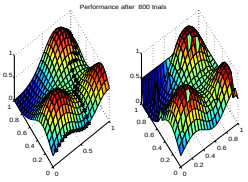

Fig. 4. Performance after 800 trials

The simulation results are shown in Fig.1- Fig.5. The left plot in each figure shows the actual terrain and the path of the rover along the terrain according to the current estimate of the terrain (which is shown in the right plot). As can be seen from the plots, the rover is able to finds a reasonably good route after around 400 trials. The optimal policies after that oscillate between a finite number of alternative policies. This is a facet of all reinforcement learning problems since there might be more than one control action which satisfies the principle of optimality at every state.

Thus, in this section, we have presented a simple example where the methodology developed in this paper was applied to the problem of a mobile robotic rover negotiating an unknown terrain. The results obtained from the simulation studies seem to validate the methodology for intelligent path planning as presented in this paper.

V. **Conclusions**

In this work, we have presented a methodology for intelligent path planning. We have shown that any path planning problem, under certain reasonable assumptions, can be reduced to the adaptive control of a Markov decision process and have used the "certainty-equivalence" principle in order to adaptively/intelligently update the control policies for path planning. We have also shown that the problem possesses special structure which can be exploited to significantly reduce the dimensionality of the problem. We have applied the methodology to the problem of path planning of a mobile robotic rover in an unknown terrain and have obtained satisfactory results indicating that this approach might be a viable strategy for intelligent path planning.

In the development, we have assumed that the path planning problem can be posed as a discounted stochastic optimization problem. Theoretically, it would be of interest to see if this condition can be relaxed to admit other kinds of optimization problems. Of more practical interest is the frequency of the control update. At one end of the spectrum, we could update the control policies at every time instant, while at the other end of the spectrum, we could wait until our estimates of the environment converged before we update the control policy. However, both these extremes are possibly not "optimal" and the best solution might be somewhere midway. We surmise that the policies need to be changed when the environment starts to look "significantly different" from the estimate according to which the current control has been planned. However, these are qualitative statements and need to be quantified precisely. This will be one of the directions of our future research. We have also assumed that we have perfect state information. To make the methodology more applicable to realistic examples, this assumption needs to be relaxed. This is another avenue of research that we are currently pursuing. Also, the grid based representation scheme used in the terrain rover example is not computationally efficient. We shall try to explore other function approximation methods to make the problem computationally more tractable. The intelligent path planning methodology presented here could be an intermediate planner for the exploration of unknown terrain. We surmise that there should be a global level planner for the exploration problem which would choose among a suite of such path planning problems at every stage of the exploration. This high-level controller could be automated or under human supervision. But there is a need to incorporate

intelligence into this global planner. Moreover, there is need to take into consideration the local planners, which would ensure that the paths that are planned are followed accurately by the system and incorporate them into the system architecture so that they fit seamlessly. This could be another avenue of future research.

REFERENCES

[1] I. K. Nikolos, K. P. Valavanis, N. C. Tsourvelodis, N. A. Kostaras, "Evolutionary Algorithm Based Offline/Online Path Planner for UAV Navigation," *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, vol. 33, no. 6, pp. 898-912, Dec. 2003

[2] S. Chakravorty, P. T. Kabamba and D. C. Hyland, "Modeling of Image Formation in Multi-Spacecraft Interferometric Imaging Systems," *AIAA paper 2004-5895*.

[3] P. R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification and AdaptiveControl*, Englewood Cliffs, NJ: Prentice Hall, 1986.

[4] V. Borkar and P. Varaiya, "Adaptive Control of Markov Chains, I: Finite parameter set," *IEEE Trans. Aut. Contr.*, vol. AC-24, pp953-958, 1979.

[5] P. Mandl, "Estimation and control in Markov chains," *Adv. Appl. Prob.*, vol. 6, pp. 40-60, 1974.

[6] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*, Belmont, MA: Athena, 1996.

[7] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, Cambridge, MA: MIT Press, 1998.

[8] R. S. Sutton, A. G. Barto and R. J. Williams, "Reinforcement Learning is Direct Adaptive Optimal Control," *IEEE Control Systems Magazine*, vol. 12, no. 2, Apr. 1992, pp. 19-22.

[9] R. E. Bellman, *Dynamic Programming*, Princeton University Press, Princeton, NJ, 1957.

[10] J. C. Latombe, "Robot Motion Planning," Kluwer, Boston, MA, 1991.

[11] J. C. Latombe, A. Lazanas and S. Shekhar, "Robot Motion Planning with Uncertainty in Control and Sensing", *Artificial Intelligence*, 52:1-47, 1991.

[12] M. T. Mason, "Automatic Planning of Fine Motions: correctness and Completeness", *Proc. of IEEE Conf. Robotics. Automat.*, pp 484-489, 1989.

[13] S. Thrun, "A Probabilistic On-Line Mapping Algorithm for Teams of Mobile Robots," *The International Journal of Robotic Research*, vol. 20, no. 5, May 2001, pp. 335-363.

[14] S. Thrun, "Probabilistic Algorithms in Robotics," *AI Magazine*, vol. 21, no. 4, 2002, pp. 93-109

[15] W. Burgard, D. Fox, H. Jans, C. Matenar and S. Thrun, "Sonar-based Mapping of large-scale mobile robot environments using EM,"' *Proceedings of the International Conference on Machine Learning*, Bled., Slov., 1999.

[16] J. J. Leonard and H. J. S. Feder, "A computationally efficient method for large-scale concurrent mapping and localization," *Proceedings of the Ninth International Symposium on Robotics Research*, Salt Lake City, UT, 1999.

[17] J. A. Castellanos, J. M. M Montiel, J. Neira and J. D. Tardos, " The SP map: A Probabilistic Framework for Simultaneous Localization and Mapping," *IEEE Transactions on Robotics and Automation*, vol. 15, 1999, pp. 948-953.

[18] G. Dissanayake, H. Durant-White and T. Bailey, " A Computationally Efficient Solution to the Simultaneous Localization and Mapping Problem," *ICRA'2000 Workshop W4: Mobile Robot Navigation and Mapping*, April 2000.

[19] S. M. Lavalle, "Robot Motion Planning: A Game-Theoretic Foundation", *Algorithmica*, vol. 26, pp. 430-465, 2000.

[20] S. M LaValle and S. A. Hutchinson, " An Objective-based Framework for Motion Planning Under Sensing and Control Uncertainties," *Internat. J. Robotics Res.*, 17(1):19-42, January, 1998

[21] S. M LaValle and R. Sharma, "On Motion Planning in Changing, Pertially-Predictable Environments"', *Internat. J. Robotics Res.*, 16(6):775-805, December, 1997.

[22] H. Hu and M. Brady, "Dynamic Global Path Planning with Uncertainty for Mobile Robots in Manufacturing," *IEEE Transactions on Robotics and Automation*, vol. 13, no. 5, pp. 760-767, October 1997.

[23] B. V. Gnedenko, *The Theory of Probability*, New York, NY: Chelsea, 1968