Title: SEMI-OPTIMAL PATH FINDING IN A WHOLLY UNKNOWN ENVIRONMENT

Abstract: A mobile robot, which is guided by a new learning algorithm is presented. The first task for the robot is searching a path like a human being. That is, finding a semi-optimal path to a given goal in a wholly unknown, unpredictable and partly, dynamic large-scale environment. Therefore, it demands high technique of environment modeling through learning. Because the robot should realize its missions within the framework of the task knowledge, there appears a new phenomenon and a Dead Corner theory: genetic algorithm and reinforcement learning algorithm are the basis for the presented theory. Powerful methods are described in the mathematics tool of MDP models. The essential method of the sub-goal method is similar with the Sub-goal method of Options over MDP theory and has been successfully proved. The presented theory is a refined theory of A Learning Classifier System theory. Most of the theory has been proved in a visualization platform of an animation simulator.
DESCRIPTION

SEMI-OPTIMAL PATH FINDING IN A WHOLLY UNKNOWN ENVIRONMENT

TECHNIQUE FIELD

The first task for our autonomous learning robot is searching a path like a human being in a wholly unknown environment. In other words, it is finding a semi-optimal path to a given goal in a wholly unknown environment. The task description for the robot is to follow goal's direction and approach to goal. Thus, there is only one problem, which is collision avoidance in a wholly unknown environment.

Therefore, it demands a high technique of learning for environment modeling. Being similar with the theory of A Learning Classifier System, which is a MIT press in 1998, our theory also has two basis algorithms: genetic learning algorithm and reinforcement learning algorithm.

Learning algorithm is an essential topic in artificial intelligence. Utilizing the theory of Markov Decision Process, which is a discrete time process, is a powerful approach for analyzing human decision problems in artificial intelligence. How to avoid obstacles is a kind of decision process. Moreover, the motion of a robot can be considered as a discrete time process. Therefore, we utilize the MDP theory to describe and analyze the basis of our HiMBS theory.

Temporal abstraction is the essential topic in artificial intelligence area for analyzing human decision process. Options over MDP is proposed by R.S. Sutton basing on MDP theory and semi-MDP theory, which is a theory of temporal abstraction within the framework of reinforcement learning. We also utilize the Option concept to describe our powerful novel environment modeling method: Sub-goal task method.

A great important note is that, although our HiMBS theory is also an approach of temporal abstraction in artificial intelligence, our approach is not an *Action Replay Process (ARP)* or a *Synchronous Value Iteration (SVI)* process as the theory of Options over MDP and Q-learning. Therefore, the most important things for understanding the MDP part of our HiMBS theory are the understandings of several fundamental MDP concepts: MDP, MDP tree, State, Action, State value, Policy and Options.

BACKGROUND ART

Our algorithm of HiMBS is a new learning algorithm for a mobile robot. It is designed for making an authoritative autonomous learning mobile robot. A conventional opinion on the role of the autonomous
learning robot is that it should be able to interact with a wholly unknown natural environment, which is unpredictable in both time and space and perhaps, contains dynamic obstacles sometimes. This is one of the most difficult problems for most man-made machines all over the world.

Most conventional learning algorithms on mobile robot demand much human empirical knowledge to enable a robot to explore and exploit a wholly unknown environment. Mostly, the algorithm's designer not only gives the task question to a machine but also spends tremendous time to implant non-task method answer keys to this machine, that is, these methods cannot be realized as task description does but greatly depend on human's other non-task method empirical knowledge to realize task. Strictly speaking, the environment in most of these conventional learning theories is not exactly wholly unknown. Some representative algorithms are discussed latterly.

Once more, human's other empirical knowledge we mention here means that the answer keys of these kinds of knowledge do not utilize knowledge of task to realize task. For instance, to the general task of path finding, that is, the task of following goal's direction from a start point and approaching to this given goal, method of pre-giving an environment map to a robot and method of wall following are not considered as task method answer keys. They are the human's other non-task empirical knowledge answer keys to realize the task.

In our HiMBS algorithm, we utilize task knowledge, which is the task description of following goal's direction and approach to goal, to realize the task of finding a semi-optimal path from a start point to an arbitrary given goal.

**Our Background of Some Conventional Learning Algorithm Theories and Path Finding Theories**

The supervising algorithm is a conventional famous learning algorithm. However, it often wastes much time on desiring a rarely happening architecture in an actual experiment. It cannot deal with a wholly unknown environment exactly.

A learning classifier system affords a very good algorithm framework. It consists of reinforcement learning algorithm and genetic algorithm. However, it is impossible to make a robot to avoid arbitrary obstacles by just constructing a blank algorithm structure to show avoiding several simple obstacles. We are not sure that this algorithm is possible to deal with a wholly unknown environment and make robot show more complex behaviors.

Jong Hwan Lim and Dong Woo Cho in Robotica 1998 propose the method of employing the concept of Quadtree to solve paths finding task. "Robust Sub-goal Planning and Motion Execution for Robots in
Fuzzy Environments”, which is proposed by Jianwei Zhang and J. Raczkowsky in IROS, is also for realizing the same task as ours. Both of them also utilize the same Sub-goal method as ours to solve the environment-modeling problem.

The environment in path finding can be described as a tree structure: a particular Quadtree or a general MDP tree for a whole environment (See Fig. 0). In the general MDP tree, each node represents a possible passing point for a robot. Every group of nodes that starts from the starting point to the final point of task goal forms a path. An optimal path can be derived from all of the found paths. Therefore, how to find the nodes of a path is the first essential topic for every path finding theory.

The method of employing a Quadtree to find nodes is not a task method, that is, it is other additional human empirical knowledge from the task knowledge. This method assumes the robot has a Quadtree knowledge and can decompose its workspace into several sub-nodes. These nodes are chosen as sub-goals to be reached successively. However, this decomposition method can not solve the obstacle avoidance problem, in other words, this method can not model environment. In order to solve this problem, the designer implants another human empirical knowledge into robot, which is following the boundaries of objects, that means robot has the empirical knowledge of wall following knowledge to pass obstacles. In a word, this theory employs much human non-task empirical knowledge to enable a robot to successfully find a path in an unknown environment.

The method of robust sub-goal planning theory also does not utilize task method to find nodes. This method assumes that obstacles are polygons and these polygons are pre-known. Thus, it is possible for this method to utilize the method of searching a Tangent-graph to find nodes. Obviously, environment in this method is not an actually unknown environment.

In a word, although these two methods also treat the nodes of a tree as sub-goals to find a path, they utilize too much human non-task empirical knowledge to find the nodes of an environment’s Markov tree. However, our method of finding these nodes is a novel method. These nodes are found by a robot’s own experience of learning and interacting with its environment but not by any pre-assumption of our human being. The robot is guided by its task knowledge: following goal’s direction and approaching to goal.

Therefore, our invention here focuses on how our task method can find the nodes of a semi-optimal path by robot itself. That is, how can our HiMBS robot realize the problem of environment modeling through learning? An optimal path can be derived after all semi-optimal paths are found in our HiMBS theory. Our invention here firstly considers how our robot can find a semi-optimal path.
Our Background of Options over MDP

The theory of Options over MDP is newer than Dyna-Q learning, which is also developed by R.S. Sutton. It introduces many methods on exploring and exploiting an environment and how to calculate an optimal policy. Many methods are such as: 1. The reinforcement learning (MDP) framework. 2. SMDP (option-to-option) methods. 3. Sub-goals for learning options. Etc. They have proposed many mathematics methods to describe and solve human decision process as a Markov Decision Process. Collision avoidance is a kind of human decision process and a discrete time process. Therefore, we utilize MDP to describe our method.

However, as mentioned by R.S. Sutton himself. The theory of Options over MDP has offers a lot, but the most interesting cases are beyond it. We consider it means that the details of evaluating actions’ credits, actual methods of how to interact with a large-scale wholly unknown environment and one of the most difficult actual problems: robot’s position navigation system problem. They also say that many key issues remain incompletely understood. For example: the source of sub-goals (and of course the method of realizing them) and integration with state abstraction, etc. Although our approach is different from these ARP processes, a similar sub-goal method, which is the only essential method for solving our problem of environment modeling through learning, is proposed by us to successfully realize the task of environment modeling.

Benchmark Mobile Learning Robot Problems:


The benchmark mobile learning robot problem is how to make an authoritative autonomous learning robot. That is, to judge whether a mobile robot is this kind of robot or not is according to its interacting environment’s fuzzy degree. A mobile robot can be said an autonomous learning robot if its interacting environment is wholly unknown, as complex and arbitrary as an actual space will be, which has every kind of the possibilities of containing dynamic obstacles at every corner of its. However, being capable of interacting with a wholly unknown environment is not the only judging criterion of an autonomous learning robot. This environment capability has been a pursuing final goal for the long past human’s history. That is, in order to enable a robot to interact with a really wholly unknown environment, we, as designers, should implant a high certain learning capability to a robot. What kind of a learning capability we should implant into a robot is a great problem.
Moreover, there is one of the most difficult actual problems for a mobile robot. Setting up an accurate and robust position-localizing system is always arguably one of the essential necessities to meet the objectives of an autonomous guided mobile robot in these environments. There are two representative existing position navigation systems: Global Position System (GPS) and Inertial Navigation System (INS). Realizing these localization systems in a small range area may be accurate and cost reasonable. On the contrary, the accuracy of these navigation systems in a large range arbitrary natural environment will greatly decrease, cost expensive and destroy the correctness of robot’s interactions.

DESCRIPTION OF THE INVENTION

1. Our robot’s conditions, improvements and the structure of this description of the invention.

Perhaps, there exist many other methods to realize the same task as our system: the task of environment modeling through learning. However, our methods are realized under a robot’s worst conditions: the robot only has one kind of external sensor, which is only capable of detecting obstacles one step around robot. Here, one step refers to the length between robot’s two footmarks. The robot does not have any outer position navigation system. Furthermore, the robot should realize its task within the framework of the task knowledge, which is the least knowledge that we should give to a robot when it is put into an environment. However, the robot has to interact with a wholly unknown environment.

Therefore, what is the task description for our robot? We give an example in Fig. 1. In such an empty and lightless environment except for the Goal, the task is that robot should approach to goal and reach goal by following the goal’s direction. One criterion for judging reaching the goal is that when the robot is stopped, in other words, the robot can judge that there are two same states, we can inform robot that it has reach its task goal. Corresponding to the theory of Options over MDP [R.S. Sutton], the task can be described in such a MDP Policy model:

\[ P \in S \times A, \text{ goal } \in \{0,1\}, \text{ method } = \theta \]

---Task MDP policy model (1)

Here \( \theta \) means following the goal’s direction and approaching to the goal. We will describe more details of this model latter. The theory of Options over MDP is a framework of temporal abstraction in reinforcement learning. It is as general as a mathematics tool on analyzing methods of exploring an environment and exploiting it by calculating policy-values, action-values and state-values for a robot.

We have made some important improvements from these two theories: 1. we touch on some most
important actual issues: how to enable robot to interact with a large-scale wholly unknown environment. We propose a particular method of the framework of the task knowledge to realize environment modeling through sub-goals learning. 2. we propose a Zhouran Zen Inertial Navigation System with only one inertial sensor of encoder to guide a robot and evaluate its actions' credits. 3. we propose the essential issue of sub-goal method to enable robot to overcome the complexity of an environment, to find a semi-optimal path quickly. 4. Our sub-goal method can absorb the actual great mobile robot's problem: motor slipping (Because the slipping occurring state will be treated as a deadlock state) and the accumulative dead reckoning errors (Because robot moves according on orientation).

Our method is quite novel because we try to use the task knowledge to finish task and this method is as natural as human's daily intelligent activities. We compare our HiMBS robot with our human-being under the same task of finding a semi-optimal path to a given goal to show our robot's naturalness. (See table 1.0)

<table>
<thead>
<tr>
<th></th>
<th>A human</th>
<th>Our HiMBS robot</th>
</tr>
</thead>
<tbody>
<tr>
<td>0. Can interact with a wholly unknown environment</td>
<td>0. The same as a human</td>
<td></td>
</tr>
<tr>
<td>1. No need of any position navigation system</td>
<td>1. The same as a human</td>
<td></td>
</tr>
<tr>
<td>2. Only need to know the goal's direction, eigenvalue or goal's some literature features.</td>
<td>2. The same as a human, better to have an estimated position of goal</td>
<td></td>
</tr>
<tr>
<td>3. Has a strong memory's problem</td>
<td>3. Has a strong memory.</td>
<td></td>
</tr>
<tr>
<td>4. Has experience knowledge about environment and methods: e.g. walk along a wall to pass.</td>
<td>4. No knowledge on environment.</td>
<td></td>
</tr>
<tr>
<td>5. Has knowledge of its front, back, etc...</td>
<td>5. Not have such knowledge.</td>
<td></td>
</tr>
<tr>
<td>6. Can see further than one footstep around</td>
<td>6. Can only detect obstacle a step around it.</td>
<td></td>
</tr>
</tbody>
</table>

The structure of this description of our invention is: 1. Brief introduction of our robot’s conditions, improvements and the structure of this description of the invention. 2. Our HiMBS theory, its program procedure and an essential sub-goal theorem. 3. Experiment results and an error function. 4. Conclusions and our future work.

2. **Our HiMBS theory, its program procedure and an essential sub-goal theorem.**
An autonomous learning system should be able to endow robot with human-level intelligence to interact with a natural environment. Mostly, the natural environment is wholly unknown, unpredictable at both time and space. Sometimes, it also contains dynamic objects, for instance, an obstacle, an active predator or a locomotive goal. Of course, it consists of arbitrary obstacles, which have uncertain size and shape. Conventionally, if an autonomous robot can not successfully realize this mission in such a natural environment, it is not considered as an authoritative autonomous learning robot. Since an autonomous learning robot interacts with an environment as uncertain as a natural creature does. Therefore, its interaction is a process of from sub-optimization to optimization, a process of from trials to planning, a process from errors to correctness. The genetic algorithm is quite suitable to such a process. We utilize it as the control central of robot's actions. However, the environment is so far complex. There are various different stimuli. Whereas, there is only one standard of credit for genetic algorithm to run rules (actions, etc.) process. The reinforcement learning algorithm is suitable for translating various stimuli into single credit. It is a good evaluating tool. Thus, we use the reinforcement algorithm as a cooperative algorithm of GA.

The architecture of our HiMBS system is shown as Fig. 3. Since the environment is wholly unknown to our robot, genetic algorithm make our robot be sure to explore and exploit the environment thoroughly. Genetic algorithm has four modules: reproduction, selection, crossover and mutation. Reproduction produce some fundamental strategies and of course, primary actions at each footmarks of robot. For instance, strategy of the framework of the task method knowledge is executed in this module and the primary strategy of sub-goal method. Selection acts as an auction selector to select every action and every strategy to command the actuator or other component systems to produce actions or rules. The roles of both crossover and mutation are only realized at finding other paths in order to find an optimal path, which are not the central topic in this paper. They both change their genes of a path to find other paths. Gene of a path means the important points of a path but not all the points of a path. The roles of reinforcement learning algorithm are: evaluating credits of each actions and each control rules in order to let the Selection module works, evaluating stimuli of a robot, for example the stimulus of encountering an obstacle, the stimulus of reaching task goal, etc.

We use the sub-goal's MDP model to describe the most fundamental strategy, which is the framework of the task method knowledge. For instance, as we show in the figure, when a robot is stopped, genetic algorithm transfer the current goal and the termination condition of the stopped process to MDP sub-goal model, this model then produces sub-goals as the following current goal to reinforcement learning
algorithm. Reinforcement learning module use the current goal to continue to evaluate robot's every action. The MDP sub-goal model judges the termination condition of the sub-goal process.

The procedures of our whole system's programming are:

* In the 1st whole loop: robot utilizes the task method knowledge to explore the environment, then it can find one quite short path and can find the optimal path at once if there is no any obstacle because of our task method method. Our robot utilizes an inertial navigation method to build a two dimensions coordinates plane for its environment. Moreover, in the first whole loop, after robot arrives at its task goal, it can correct the estimated coordinates of task goal's position.

* During the 2nd whole loop: robot will try to optimize the found path by calling the mutation module and calculate the path's policy value according to its learning experience on environment's model.

We have mentioned that our HiMBS theory is a refined theory of Options over MDP. One of the biggest improvements is the most fundament problem of the credit model. Together with the Q-learning theory, the Dyna-Q learning theory, their credit models are only in a real number space, that is, they only consider the award of a credit, which is \( C_i = r_i \), \( i = 0,1,2,3, \ldots \). However, according to the newest reinforcement learning theory, a credit does consist of not only an award but also a punishment. Therefore, we upgrade the most primary credit model into a complex number space, which is \( C_i = r_i + \gamma \omega_i \), \( i = 0,1,2,3, \ldots \). Here, \( \omega_i \) represents the punishment part of a credit. Thus, from our credit model, we can have more accurate information on robot's interaction and can enable us to separate two paths that have the same policy-value as Q-learning but has two different environments. We can find the optimal path then. Therefore, all of the values calculations of policy-values (in mobile robot case, simplified speaking, path-values), state-values and action-values are changed. For example, the calculation of state-values is changed from formulation (2) to (3).

\[
\begin{align*}
\lambda_{i+1}^{\pi}(s_i) \\
= \sum_{s \in A} \pi(s, a)[r_i + \gamma \sum_{s'} P_{ss'}^{aa} \lambda_{i+1}^{\pi}(s')]
\end{align*}
\]

(2)

However, in the whole loop of exploring an environment, we do not utilize the SVI process (Synchronous Value Iteration process). In other words, we do not start with an arbitrary approximation calculation to find paths and then make paths planning. Because exploring every points of an environment by starting with an arbitrary approximation calculation demands numerous memories and takes plenty of time.
\[ V_{\pi}(s_i) = \{ \sum_{a \in A} \pi(s, a)[r_a + \gamma \sum_{s'} p_{ss'}^a \text{Re}(V_{\pi}(s'))] \} + \gamma \sum_{s'} p_{ss'}^a \text{Im}(V_{\pi}(s')) \]  
---(3)

Our method of utilizing the framework of the task method knowledge has an immediate approximation. In addition, we utilize the cumulative calculation formulations, such as formulation (3), at the second whole loop, that is, soon after a robot has finished its task. Therefore, our calculations of state-values in the first whole loop of exploring an environment is simplified as formulation of (4). Here, \( \pi(s, a) \) means the probability of taking action \( a \) in the state of \( s \) under the policy of \( \pi \) (task MDPP model (1)). \( r_a \) means a step forward reward of taking action \( a \) in the state of \( s \).

\[ V_{\pi}(s_i) = \{ \sum_{a \in A} \pi(s, a)r_a \} + \gamma \sum_{s'} \pi(s, a) \omega_{s's_i} \]  
----(4)

Thus, how to evaluate actions' credit and how to realize our three objectives are described as below:

To realize the first objective of environment modeling, there is a new dead-corner theory that appears in the first whole loop.

As we mentioned before, the method condition of our robot is described in an empty and lightless environment. In the task MDPP model (1), \( P \gamma \) means a policy. \( S \) means the state space of a robot's environment. \( A \) means the actions unit of a robot. \([0,1]\) means the probability of a policy is between 0 to 1. \( \theta \) represents to a certain method to realize the policy. A policy means the mapping from states to the probabilities of taking each available primitive action. A briefly and more accurate explanation of policy is that policy is a possible sequence of actions-states pairs. In paths finding theory, a policy can be simply called as a path.

Therefore, what does the model (1) means exactly? When a robot is inputted into a wholly unknown environment for the goal, it receives the only environment knowledge: the goal's direction and the estimated coordinates of the goal. (Strictly speaking, coordinates of the goal are possible to be gotten from the goal's direction.) The task for the robot, even in an empty and lightless environment (except for the robot and the goal), is described as: Following the goal's direction. If robot is not stopped, robot keeps on approaching. The initiating conditions are written as / . If the goal is treated as an obstacle, robot will fall into an oscillating state, which is called a deadlock state. Thus, the deadlock state conditions can be a criterion to judge task goal. The deadlock termination condition of a global task process is represented by
\[ \beta. \]

Thus, in order not to use an outer navigation system and to evaluate the actions credits, we propose a Zhinan Zen Inertial Navigation System and Evaluation method (abbreviated as ZZINS_EF, see Fig. 3).

Our robot has eight moving freedoms. We use the ZZINS_EF to evaluate credits of these eight possible actions according to whether the direction of an action makes robot go farther or nearer to the task goal. The credit of an action has the highest credit among the eight actions if the direction of an action is the same as the current direction of task goal. Because of our ZZINS_EF method, our robot has an immediate efficiency of finding the shortest path at once in an empty and lightless environment and does not need any outer position navigation system. Our robot is initiated because of the direction of the current goal but not the coordinates of robot. Therefore, our robot can absorb a quite high level of motor’s slipping errors.

Thus, eight possible moving directions have their own corresponding credit.

Therefore, what do the initiating conditions \( I \) and the deadlock termination conditions \( \beta \) mean exactly in this empty and lightless environment? The most simplified meaning of \( I \) is that if the initiating conditions are not equivalent with the deadlock termination conditions \( \beta \) at the goal, the robot can keep on going. We define the state of robot’s \( n \)th footmark as \( s_j, j = 0,1,2,3,4, \ldots \). Therefore, its next state is \( s_{j+1} \). The deadlock termination conditions \( \beta \) means that there exist at least two same footmarks’ states at the same position of around goal in this empty and lightless environment. Mostly, a wholly unknown environment is quite complex. Plenty of particular literature characteristics of an objective (obstacle) can represent one unique state element to a human, however is difficult for a computer. Whereas, coordinates of a position are not difficult for a computer, we consider the coordinates of a footmark as a state element of the state at that footmark.

Thus, the value of a state \( s_j \) at the \( n \)th footmark is written as \( I^\beta_j (s_j^1), j = 0,1 \). \( I^\beta_j (s_j^1) \) is the same as the formulation (4). It summarizes all of the possible actions values at state \( s_j \).

\( I^\beta_j (s_j^1) = (x_j, y_j) \) is the coordinates particular state element of coordinates of state \( s_j \).

Therefore, corresponding to the theory of Options over MDP, the Options-MDP (abbreviated as OMDP) model of our framework of the task method knowledge can be written as:

Here, the task can be described by the option of \( O \). It consists of the initiating conditions of \( I \) at the start point and the deadlock termination conditions of \( \beta \) at the position of task goal. CG means the
current goal. TG means the task goal. \( P_y \) means the MDP policy model of (11).

\[
O(I, P_y; \beta)
\]

\[
P_y: S \times A | CG = TG, method = \theta
\]

\[
I: V_i^{P_y}(s_i^') \neq V_i^{P_y}(s_i^{'+k}) | CG = TG
\]

\[
\beta: V_i^{P_y}(s_i^') = V_i^{P_y}(s_i^{'+k}) | CG = TG
\]

\[
\tau = 0, 1, 2, 3, 4,...; k = 2, 3, 4, 5, 6,...
\]

\[
j = 0.1.
\]

---------- Task OMDP model (TOMDP) (5) 

We assume the state now is \( s_{i+k} \). Because the method condition of our OMDP model framework, there appears a Dead-Corner problems. However, in order to try to utilize the task knowledge to solve the Dead-Corners, to realize the given task, we find that it is still possible to use the OMDP model to finish the task. In our following paper, we will introduce our dead corner theory.

To solve these dead-corners problems, in other words, the dead corners environment modeling, we propose a sub-goal theorem to realize it, which is mentioned as an essential issue by R.S. Sutton.

In order to construct a sub-goal OMDP model, we just change the task OMDP model little. We assume that the state of \( s_i \) is the state at the position of judging a deadlock state. When a robot is stopped by a dead corner at the point of \((x_1, y_1)\), it will set up sub-goals near the directions of its possible moving directions at a changeable distance of \( len \). Normally, a 80 times length of a step of \( len \) is enough for passing most arbitrary dead corner and obstacle if the environment does not contain a particular big obstacle. We give an example of sub-goal in the Fig. 4.2. The following (6) is our Sub-goal's OMDP model.

\[
O(\beta, P_y; I)
\]

\[
P_y: S \times A | CG = SG \Leftrightarrow TG, method = \theta
\]

\[
I: V_i^{P_y}(s_i^') \neq V_i^{P_y}(s_i^{'+k}) | CG = SG
\]

\[
\beta: V_i^{P_y}(s_i^') = V_i^{P_y}(s_i^{'+k}) | CG = TG
\]

\[
\tau = 0, 1, 2, 3, 4,...; k = dynamic const \in [2, 3, 4, 5, 6, 7,...]
\]

\[
j = 0.1.
\]

We only switch the current goal between \( TG \) (task goal) and \( SG \), switch the position of \( \beta \) and \( I \) and limit dynamic changeable constant of \( k \) but not arbitrary variable of k in the task OMDP model (5). We do not change the essential part of task OMDP model, which is \( \theta \) method.

An important note is why \( k \) is dynamic changeable. There are several reasons: one is that the
environment is wholly unknown. One is that the robot has a non-zero size. One is that there exist arbitrary size dead corners. One is that sub-goals may be set up inside the body of an obstacle. That means the sub-goals are unreachable. The final reason is that we want to enable robot to learn an accurate dead-corner model.

We summarize a sub-goal theorem as below:

In order to use the natural sub-goal method and try to use the task knowledge to solve the Environment Modeling problem, We set up a Sub-goal OMDP model (SOMDP) from the task OMDP model by changing little of it.

When a robot is following into a deadlock state, some sub-goals will be set up near to the possible moving directions at a certain distance. We assume the state at this footmark of judging a deadlock state is $s_i$. The SOMDP model will begin to be set up according to the termination condition of TOMDP’s $\beta$. That means the current goal will be changed to sub-goal (SGi) and then robot will begin to move for $k$ steps after $s_i$ under the sub-goal policy $P_{yi}$. When $k$ steps are over and the state-value of $V^{(b)}_{i,s_{1+1}}$ satisfies the termination condition $1$, the current goal will be switched back to task goal $\ominus(TG)$. That is, the option $O$ of SOMDP is finished. However, the option $O$ will be set up again at the $2k$ steps after $s_i$ if the state value of $V^{(b,o)}_{i,s_{1+2k}}$ satisfies the initial condition $\beta$ under the same task goal policy $P_{y0}$. If, at the $2k$ step footmark, $\beta$ is incorrect, the SOMDP will be over, that means the robot has passed the dead corner and this dead corner’s environment modeling has been finished. If $\beta$ is correct, $k$ will increase to be $k + 1$. The same SOMDP is set up again and the sub-goal process will begin as the former sub-goal process.

Therefore, there occurs an interesting learning feature of TRIAL. The robot usually tries many times to make an accurate dead-corner model, which is shown by our experiment.

Thus, the first objective of our environment modeling through learning by robot itself is realized. A model of a dead corner consists of three points normally. Please see an example of Fig. 4.

We treat the model points of a dead corner as genes of a path. Therefore, we activate the Mutation module of Genetic Algorithm to set every gene of a path to be current goal for some time in orders during the robot’s approaching to the task goal. The different orders of the genes treated as current goal, the different paths will be found. Additionally, our sub-goal method in paths finding is also guaranteed to exploring the environment thoroughly and finds an optimal path. More details of these topics will be presented in our following papers.
3. Experiment results and an error function.

In order to show the correctness of most issues of our sub-goal learning theory, we present a dynamic experiment result here. In this experiment, robot encounters a dynamic obstacle firstly, a theoretical dead corner secondly and a particular physical dead corner finally. (See Fig. 6)

Because the interacting environment is wholly unknown to robot and our robot has to learn to get a correct model of its environment, robot cannot get a correct model of this dynamic obstacle without trying. Fig. 5 is the TRIAL graphic of this dynamic experiment.

In the table 5.0. DO means the dynamic obstacle. The red coordinates of (97,440) means the 1st sub-goal, which is chosen firstly, is unreachable. It is in the body of Obstacle 1. DC1 is the first Dead Corner. DL means the state of Dead Lock. Under DL, S means the start step of Dead Lock, E is the last step of this Dead Lock. TRIALS 1(3) means the steps of the first TRIAL is 3 steps. SG1 means the 1st Sub-goal of this Dead Corner. SP means the position of the Dead Lock. MP means the last point of this Dead Corner's model.

<table>
<thead>
<tr>
<th>Contents</th>
<th>DL</th>
<th>TRIALS</th>
<th>SG0 (x,y)</th>
<th>SG1 (x,y)</th>
<th>SP (x,y)</th>
<th>MP (x,y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obstacle</td>
<td>S</td>
<td>E</td>
<td>1(3)</td>
<td>2(4)</td>
<td>Max</td>
<td></td>
</tr>
<tr>
<td>DO</td>
<td>88</td>
<td>93</td>
<td>94</td>
<td>102</td>
<td>112(5)</td>
<td>(</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(488.49)</td>
<td>(374,326)</td>
</tr>
<tr>
<td>DC1</td>
<td>26</td>
<td>26</td>
<td>267</td>
<td>275</td>
<td>(517.73 2)</td>
<td>(908,341)</td>
</tr>
</tbody>
</table>

From this experiment, we can get a TRIAL function on Environment Modeling through learning as Formulation (5), which, according to the sense of modeling task, is an error function. Here. \( S_n \) means the steps of robot from initial position to the position when robot encounters a Dead Corner. \( t \) means time. A process is considered as a TRIAL during which the current goal is sub-goal but not the task goal.

A TRIAL consists of several steps. \( S(t) = S_{n0} + 2 \cdot S_{mn} \cdot \int T(t) + P_{t0} \cdot \int T(t)[1 + T(t)] \)
\( P \) means the minimal moving steps during one **TRIAL**. \( T(t) \) means the number of trying times at time \( t \), that is the number of **TRIALs**. \( S_{\text{min}} \) means the minimal step-size of a dead corner, it can be any integer number. In this dynamic experiment, we let \( S_{\text{min}} = \text{3 steps} \) to fast the process.

According to this error function, we can get the size of one side of a Dead Corner from the corresponding error function, which is

\[
\text{LengthDC} = (S_{\text{min}} + P_n) \times T(t) \tag{6}
\]

We compare steps in the first loop of exploring and exploiting the environment with the steps in the second loop of original paths optimizing in Fig. 5.1 and Fig. 5.2. Results show the correctness of the learned dynamic obstacle's model and the righteousness of our sub-goal OMDP theorem.

4. **Conclusions and our future work.**

Our HiMBS algorithm is a novel learning approach of Environment Modeling in a wholly unknown environment. It enables a robot to independently and actually autonomously learn the obstacles' model. It thoroughly makes a mobile robot being free from the restriction of its environment.

We make an animation simulator simulate what an actual wholly unknown environment will be and successfully test our HiMBS algorithm. We try to pre-design all of the possible problems in an actual mobile robot experiment in its two dimensions plane. For instance, robot has a non-zero size; it has an arbitrary size and shape. Obstacle also has an arbitrary size and shape: furthermore, sometimes, obstacle is in the state of dynamic moving. Moreover, there is one of the most difficult problems: a difficult dead reckoning problem, etc. These problems have been successfully overcome by our HiMBS algorithm. Therefore, although there may be still some defects of our HiMBS, we believe HiMBS algorithm is a powerful and successful algorithm for mobile robot to find a semi-optimal path to a given goal in a wholly unknown environment. This animation test platform has also testified our Zhinan Zen INS navigation theorem and evaluation theory. In a word, the industrial program experience enables us to find the essential actual issue of sub-goal method and the new Environment Modeling learning method of **TRYING** in mobile robot.

A three-dimension simulating environment will make our algorithm behaviors be more interesting and complex. Our algorithm may fail more times at the first loop. However, it will quickly correct its errors.
and learn successful modes. Additionally, in order to make an authoritative autonomous learning robot, it is necessary for us to test our algorithm in other tasks. Otherwise, we are not sure to make a general algorithm.

How can our sub-goal method in path planning enable a robot to find other paths except the first found path? Is our sub-goal method in path planning guaranteed to explore and exploit a wholly unknown environment thoroughly? What are the details of our path optimizing theory? Etc. These topics will be made clear in our following papers.

Finally, many essential concepts in robotics are left for us to continuously develop, such as, the intelligence definition and the least knowledge for a learning algorithm, the state concept in a computer, how to judge whether an algorithm is better or not. Etc. This paper of ours is a beginning of the research approach of discussing the least knowledge for a learning algorithm and the state concept in a computer. Our this invention will be start of a new artificial intelligence approach.

**BRIEF DESCRIPTION OF THE DRAWINGS**

Fig. 1 The Markov tree structure of every paths finding environment.

Fig. 2 The task description in an empty and lightless environment.

Fig. 3 The architecture of our HiMBS.

Fig. 4 Zhan Zen Inertial Navigation System and Evaluation model (abbreviated as ZZINS_EF)

Fig. 5 The model of a dead corner.

Fig. 6 Learn to model a dynamic obstacle

Fig. 7 A successful dynamic experiment result

**BEST MODE FOR CARRYING OUT THE INVENTION**

The least hardware conditions of our robot are: one step-around-robot extern sensor or better one for detecting obstacles, motors and wheels or other instruments for moving, an inertial encoder sensor but no needs of Gyro sensor, accelerator sensor or tachometer, a controller for running our algorithm. The spirit software condition of the robot is our HiMBS executable file. Etc.
INDUSTRIAL APPLICABILITY

If a one-step-around-robot extern sensor or a better one for detecting obstacles is available, our algorithm can be utilized in industrial applications. For example, a waiter robot in restaurants, a mine searching machine, a deep-marine-object searching submarine, etc.

References


Claims

Claim 1. We hope to be protected in any industrial path finding field. Any program that utilizes our sub-goal task method should obey the invention law. Any program, which has a trying behavior to avoid collision and use our sub-goal method in a wholly unknown environment, is considered as utilizing our invention.

Claim 2. Any program should obey the laws to utilize any protected idea in this invention of ours.
Drawings

Starting Point

S(0)......S(p)

S(i)

S(n)

S(k)

Goal

Description of any space's: Markov tree

Figure 1

Robot

Path1

Goal

(R_G, \theta_G) \rightarrow (X_G, Y_G)

Task method in an empty and lightless environment

Figure 2
Figure 3

Figure 4

Example of Zhinan Zen Inertial Navigation System and Evaluating Function method.
Notes:
MPi: ith model point; SGi: ith subgoal
DCP: the dead corner point

Figure 5
**INTERNATIONAL SEARCH REPORT**

**A. CLASSIFICATION OF SUBJECT MATTER**

| IPC 7 | B25J/16 |

According to International Patent Classification (IPC) or to both national classification and IPC.

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

| IPC 7 | B25J |

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**EPO-Internal**

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>ZHANG J., ET AL: &quot;ROBUST SUBGOAL PLANNING AND MOTION EXECUTION FOR ROBOTS IN FUZZY ENVIRONMENTS&quot; PROCEEDINGS OF &quot;INTELLIGENT ROBOTS AND SYSTEMS '94&quot;, IROS'94., 1994, pages 447-453, XP002160783 cited in the application the whole document</td>
<td>1, 2</td>
</tr>
</tbody>
</table>

X Further documents are listed in the continuation of box C. Patent family members are listed in annex.

<table>
<thead>
<tr>
<th>Special categories of cited documents:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;A&quot; document defining the general state of the art which is not considered to be of particular relevance</td>
<td></td>
</tr>
<tr>
<td>&quot;E&quot; earlier document published on or after the international filing date</td>
<td></td>
</tr>
<tr>
<td>&quot;L&quot; document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</td>
<td></td>
</tr>
<tr>
<td>&quot;O&quot; document referring to an oral disclosure, use, exhibition or other means</td>
<td></td>
</tr>
<tr>
<td>&quot;P&quot; document published prior to the international filing date but later than the priority date claimed</td>
<td></td>
</tr>
<tr>
<td>&quot;X&quot; later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</td>
<td></td>
</tr>
<tr>
<td>&quot;Y&quot; document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</td>
<td></td>
</tr>
<tr>
<td>&quot;Z&quot; document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</td>
<td></td>
</tr>
<tr>
<td>&quot;F&quot; document member of the same patent family</td>
<td></td>
</tr>
</tbody>
</table>

Date of the actual completion of the international search: 22 February 2001

Date of mailing of the international search report: 14/03/2001

Name and mailing address of the ISA

European Patent Office, P. B. 5818 Patentlaan 2 NL-2280 HV Ridderik, Tel. (+31-70) 340-2040, Tx 31 651 epo nl, Fax: (+31-70) 340-3016

Authorized officer: Hurtado-Albir, J
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>DEL R ET AL: &quot;Reinforcement learning of goal-directed obstacle-avoiding reaction strategies in an autonomous mobile robot&quot; ROBOTICS AND AUTONOMOUS SYSTEMS, NL, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, vol. 15, no. 4, 1 October 1995 (1995-10-01), pages 275-299, XP004001929 ISSN: 0921-8890 the whole document ___</td>
<td>1,2</td>
</tr>
<tr>
<td>A</td>
<td>SAMEJIMA K ET AL: &quot;Adaptive internal state space construction method for reinforcement learning of a real-world agent&quot; NEURAL NETWORKS, ELSEVIER SCIENCE PUBLISHERS, BARKING, GB, vol. 12, no. 7-8, October 1999 (1999-10), pages 1143-1155, XP004180502 ISSN: 0893-6080 the whole document ___</td>
<td>1,2</td>
</tr>
<tr>
<td>A</td>
<td>BARSHAN, B. ET AL: &quot;Inertial navigation systems for mobile robots&quot; IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION, vol. 11, no. 3, June 1995 (1995-06), pages 328-342, XP002160801 page 328 - page 331 ___</td>
<td>1,2</td>
</tr>
<tr>
<td>Category</td>
<td>Citation of document, with indication where appropriate of the relevant passages</td>
<td>Relevant to claim No.</td>
</tr>
<tr>
<td>----------</td>
<td>---------------------------------------------------------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>A</td>
<td>SMITH R E ET AL: &quot;Combined biological paradigms: A neural, genetics-based autonomous systems strategy&quot; ROBOTICS AND AUTONOMOUS SYSTEMS, NL, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, vol. 22, no. 1, 10 November 1997 (1997-11-10), pages 65-74, XP004100314 ISSN: 0921-8890 the whole document</td>
<td>1,2</td>
</tr>
</tbody>
</table>