Decentralized Cooperative Search by Networked UAVs in an Uncertain Environment

Yanli Yang, Ali A. Minai and Marios M. Polycarpou

Abstract— **This paper addresses the problem of cooperative search in a given environment by a team of Unmanned Aerial Vehicles (UAVs). We present a decentralized control model for cooperative search and develop a real-time approach for on-line cooperation among vehicles, which is based on treating the possible paths of other vehicles as "soft obstacles" to be avoided. Using the approach of "rivaling force" between vehicles to enhance cooperation, each UAV takes into account the possible actions of other UAVs such that the overall information about the environment is increased. The simulation results illustrate the effectiveness of the proposed strategy.**

I. INTRODUCTION

Control of networked multi-vehicle systems that are intended to perform a coordinated task is currently an important and challenging field of research [1], [2], [3], [4], [5]. This is due to the fact that collaborative teams of aerial and ground vehicles can perform a number of highly beneficial tasks in military and civilian applications. However, a major obstacle to the realization of such systems still remains the design of coordination and decision algorithms to achieve complex, adaptable, and flexible system behavior.

This paper focuses on the multi-vehicle cooperative search problem where a team of UAVs seeks to find targets in a dynamic and risky environment. In this problem, the vehicles treat all uncertain areas as possible destinations in order to identify as many targets as possible. However, due to the UAVs' energy limitations and the various uncertainties in complex scenarios, such as imperfect sensor accuracy and "pop-up" threats, the UAVs cannot use the exhaustive coverage path planning methods (e.g., Zamboni search [6]) to explicitly pass over all points in the search area. Thus the vehicles need highly autonomous path planning capabilities. Our research focuses on this problem. We have previously proposed a decentralized control framework for emergent coordinations among vehicles and developed several heuristic cooperative path planning algorithms [7], [8], [9]. Some other related works on the cooperative search problem include [10], [11]. The UAV cooperative search problem is also related to the multi-robot mapping and exploration problem [12], [13] in the robotics area.

In this paper, we extend our previous method by explicitly incorporating threats in the control model and in the cooperative path planning strategies, which makes the search problem model fit more closely to real battlefield situations. We formulate the updated search problem as a finite horizon optimal control problem, develop a coordination method based on the *rivaling force* approach [7], and evaluate the proposed scheme through simulation.

The remainder of the paper is organized as follows. Section II presents the decentralized control model for the multi-vehicle cooperative search problem. Section III describes the proposed cooperative path planning strategy. The rivaling force based coordination algorithm is given in Section IV. Some simulation results and discussion are presented in Section V. Section VI concludes the paper with some final observations.

II. PROBLEM DEFINITION

We consider a team of UAVs engaged in searching for targets in an environment of known size with the objective to identify as many targets as possible and minimize the loss or damage of the UAVs during the mission.

A. The Environment

The *environment* is a bounded $L_x \times L_y$ cellular area, where each position is termed a *cell*. The environment is populated by stationary non-threatening targets and threats. The number and locations of the targets are initially unknown. We assume that there is at most one target in each cell. There are also M stationary threats, $\gamma_i,\ i = 1, \ldots, M$, which have anti-craft capabilities, such as surface-to-air missiles (SAMs). The threat γ_i is located at (x_i^γ, y_i^γ) and it has *a priori* known attack region ϕ_i which is the range over which the threat is capable of destroying the UAVs with *a priori* known kill probability $p_{kill}^i \in [0, 1]$.

B. The UAV Dynamics Model

The team consists of N identical UAVs moving synchronously in discrete time, searching the given environment for targets. Each UAV is equipped with a sensor (with imperfect detection accuracy) and communication capabilities. At each time step, the UAV can move from one cell to another neighboring cell, subject to some maneuverability constraints. In order to simplify the problem we start with the assumption of perfect communication among UAVs, which means that, at each time step, a UAV can receive the sensing information and state information from other UAVs instantaneously through communications. This assumption will be relaxed later.

The state of UAV i at time t is denoted by $v_i(t)$, which is comprised of three components: $v_i(t) = [\lambda_i(t), o_i(t), \delta_i(t)]$. The first component

This research was sponsored by the Defense Advanced Research Project Agency (DARPA) under Contract F33615-01- C3151 issued by the AFRL/VAK.

Y. Yang, A. Minai and M. M. Polycarpou are with the Dept. of Electrical and Computer Engineering and Computer Science, University of Cincinnati, Cincinnati, OH 45221-0030, USA. (Email: polycarpou@uc.edu)

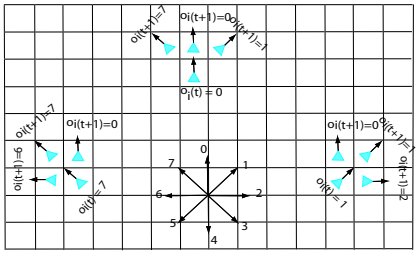

Fig. 1. Possible transition choices for agents in all 8 orientations.

$\lambda_i(t) = (x_i(t), y_i(t)) \in \{1, 2, \ldots, L_x\} \times \{1, 2, \ldots, L_y\}$ is the i-th vehicle's position in the environment at time t. The second element $o_i(t) \in \{0, 1, \ldots, 8\}$ is the vehicle's orientation defined by $\{0$ (north), 1 (northeast), 2 (east), 3 (southeast), 4 (south), 5 (southwest), 6 (west), and 7 (northwest)$\}$. The third element $\delta_i(t) \in \{0, 1\}$ of the state $v_i(t)$ is a flag indicating whether the UAV i has been destroyed at time t, where $\delta_i(t) = 1$ means that the UAV i is alive at time t.

The UAV's dynamics is subject to its physical curvature radius constraints, reflected in the fact that it can only change its orientation by at most one step, that is, $o_i(t+1) \in \{o_i(t) - 1, o_i(t), o_i(t) + 1\}\ mod\ 8$. This essentially means that the UAV's maximum turning capability is $45°$. Thus each UAV has three possible positions for the next time step, i.e. turn left, turn right or go straight, which is designated by $\{l$ (left), f (front), r (right)$\}$. Figure 1 shows this graphically for various orientations.

The control decision for UAV i is its path selection at each time step t, denoted by $u_i(t) \in \{l, f, r\}$. The UAV's state can also be changed by threats. The UAV i will be destroyed with probability p^j_{kill} in threat j's attack region, thus causing the vehicle's state flag $\delta_i(t) = 1$ to change into $\delta_i(t + 1) = 0$. The threat actions in the environment, denoted by $\omega(t) = [\omega_1(t), \omega_2(t), \ldots, \omega_M(t)]$, determine the transition of $\delta_i(t)$, which is a stochastic event. In summary, a vehicle's transition function can be expressed as:

$$v_i(t + 1) = f_v(v_i(t), u_i(t), \omega(t)) \tag{1}$$

In our model, UAVs use the q-steps-ahead path planning method [7], that is, each UAV plans its path q steps ahead of its current location, adding a new move at each time-step. For simplicity, in this paper we use $q = 1$, but the extension to $q > 1$ is straightforward. Thus, at time-step t, the UAV i makes its path decision $u_i(t + 1)$. At time t, the UAV executes an *action* comprising the following three steps:

1) It makes decision $u_i(t + 1)$ to choose a new orientation, $o_i(t + 2)$.
2) It then find its position $\lambda_i(t + 2)$ as the neighbor of $\lambda_i(t + 1)$ facing orientation $o_i(t + 2)$.
3) Finally, it executes its decision $u_i(t)$ and updates its state $v_i(t+1) = [\lambda_i(t+1), o_i(t+1), \delta_i(t+1)]$ by going to grid location $\lambda_i(t+1)$ with orientation $o_i(t+1)$ and changing $\delta_i(t + 1)$ according to threat actions $\omega(t)$.

C. The UAV Information Base

UAVs use three cognitive maps, the *target probability map*, $P(t)$, the *threat probability map*, $K(t)$, and the *certainty map* $\mathcal{X}(t)$ as its knowledge base for the mission. In the target probability map $P(t)$, each cell (x, y) has a value $p(x, y, t) \in [0, 1]$ representing the probability of a target being present in cell (x, y), termed the *target probability*:

$$p(x, y, t) = P(\text{target present at } (x, y)) \tag{2}$$

The threat probability map $K(t)$ stores the *threat probability* of each cell (x, y) denoted as $k(x, y, t) \in [0, 1]$ which represents the probability that the UAV will be destroyed at cell (x, y) by any threat. We have

$$\begin{aligned} k(x, y, t) &= P(UAV \text{ destroyed at } (x, y) \text{ by threats}) \\ &= 1 - \prod_{j=1}^{n} (1 - p^j_{kill}(x, y)) \end{aligned} \tag{3}$$

where n is the number of threats whose attack regions cover position (x, y).

The certainty map $\mathcal{X}(t)$ stores the *certainty* value of each cell (x, y), denoted as $\chi(x, y, t) \in [0, 1]$, which corresponds to the degree to which the cell has been searched. If $\chi(x, y, t) = 0$ then the cell has not been searched until time t. On the other hand, if $\chi(x, y, t) = 1$ then the cell has been fully searched. This factor is used to drive the UAVs to explore the un-searched regions.

In the decentralized search model, each UAV i carries its own cognitive maps, $P^i(t)$, $K^i(t)$ and $\mathcal{X}^i(t)$. The initial values of the cognitive maps, $P^i(0)$, $K^i(0)$ and $\mathcal{X}^i(0)$ are used to reflect the *a priori* knowledge about environment. For example, if all targets are land-based, the locations corresponding to a lake may begin with $p^i(x, y, 0) = 0$ and $\chi^i(x, y, 0) = 1$ (i.e., the location is free of targets). The threat probability map $K^i(0)$ is initialized according to the locations and types of the known threats. The maps, $P^i(t)$, $K^i(t)$ and $\mathcal{X}^i(t)$ are updated on-line using the new information obtained by UAV i's sensor scan and by communication with other UAVs. The assumption of perfect communication, in fact, makes the information base for all UAVs the same, and we can use global maps $P(t)$, $K(t)$ and $\mathcal{X}(t)$.

In the certainty map $\mathcal{X}(t)$, most cells typically begin with a certainty of zero. Each time a UAV visits cell (x, y) and makes a scan, the certainty changes according to the rule

$$\chi(x, y, t + 1) = \chi(x, y, t) + 0.5(1 - \chi(x, y, t)) \tag{4}$$

This is a simple way to track the number of useful "looks" each cell has had and captures the notion of diminishing returns with each look. For the threat map, we assume that the threats are all stationary and known *a priori*. So the threat probability map is time-invariant, that is $K(t) = K(0)$.

Next we discuss the update of the target probability map $P(t)$. As UAV i visits a cell (x, y) at time t, it makes a sensor scan to detect targets. The resulting observation is denoted by $b_i(x, y, t) \in \{0, 1\}$, where $b_i(x, y, t) = 1$

indicates a target detection and $b_i(x,y,t)=0$ indicates no target detected. The sensor's detection accuracy is characterized by two parameters, the sensor detection rate p_c and the false alarm rate p_f which are defined as:

$$p_c = P(b_i(x,y)=1 \mid A) \quad (5)$$

$$p_f = P(b_i(x,y)=1 \mid \bar{A}) \quad (6)$$

where A denotes the event that a target is actually located in cell (x,y).

The update rule for the target probability, which is derived based on Bayesian inference, is given by:

$$p(x,y,t) = b_i(x,y,t)\Lambda_1 + (1-b_i(x,y,t))\Lambda_2 \quad (7)$$

where

$$\Lambda_1 = \frac{p_c p(x,y,t-1)}{p_c p(x,y,t-1) + p_f(1-p(x,y,t-1))}$$

$$\Lambda_2 = \frac{(1-p_c)p(x,y,t-1)}{(1-p_f)(1-p(x,y,t-1)) + (1-p_c)p(x,y,t-1)}$$

It can be shown that by using the above update equations, $p(x,y,t+1) > p(x,y,t)$ for $b_i(x,y,t)=1$ and $p(x,y,t+1) < p(x,y,t)$ for $b_i(x,y,t)=0$ when $p_c > p_f$. Throughout this paper, we assume $p_c > 0.5 > p_f$, i.e., the sensors are informative.

Finally, we use a binary variable $\zeta(x,y,t)$ to indicate whether a target has been confirmed or not in cell (x,y). Initially, all cells except those with known targets have $\zeta(x,y,0)=0$. The condition for updating is:

$$\zeta(x,y,t) = \begin{cases} 1 & \text{if } p(x,y,t) \geq \theta \\ 0 & \text{else} \end{cases} \quad (8)$$

where θ is a pre-defined threshold close to 1.

D. Optimal Control Problem Formulation

To accomplish the search task efficiently, each UAV needs to find an optimal path to follow based on its knowledge of the environment. In our model, the UAVs' decision process is decentralized in the sense that each UAV makes decisions independently. This decentralized decision making problem can be formulated as an optimal control problem as follows.

As defined, the i-th UAV's decision at time t is to select its move for time $t+1$, $u_i(t+1)$, leading to its position at time $t+2$. The decision is based on the environment state $x^i(t)$, which is composed of the UAV's cognitive maps and its knowledge of all the vehicles' states and decisions. We define $v^i(t) = [v_1^i(t), v_2^i(t), \ldots, v_N^i(t)]$, where $v_j^i(t) = [\lambda_j^i(t), o_j^i(t), \delta_j^i(t)]$ denotes vehicle i's knowledge on vehicle j's state at time t; $u^i(t) = [u_1^i(t), u_2^i(t), \ldots, u_N^i(t)]$ where $u_j^i(t) \in \{l, f, r\}$ denotes the vehicle i's knowledge on vehicle j's decision at time t. Note that in practice, $v_j^i(t)$, $u_j^i(t)$ might not be the same as $v_j^k(t)$, $u_j^k(t)$ respectively because of communication limits under some scenarios. However, due to the perfect communication assumption, they are the same in this paper, which means that all the UAVs share the same environment state represented by

$$x(t) = \{P(t), K(t), \mathcal{X}(t), v(t), u(t)\}$$

where $v(t) = [v_1(t), v_2(t), \ldots, v_N(t)]$ and $u(t) = [u_1(t), u_2(t), \ldots, u_N(t)]$. Thus, for UAV i, its decision $u_i(t+1)$ is a function of the current environment state

$$u_i(t+1) = h_i(x(t)) \quad (9)$$

As the UAVs execute their decisions, $u(t)$, their sensors return the scan readings $b(t) = [b_1(t), b_2(t), \ldots, b_N(t)]$ and the threats take actions $\omega(t) = [\omega_1(t), \omega_2(t), \ldots, \omega_M(t)]$. Both $b(t)$ and $\omega(t)$ are stochastic quantities and they, together with the vehicles' actions, determine the new environment state through a stochastic transition function, f_s:

$$x(t+1) = f_s(x(t), u(t), b(t), \omega(t)) \quad (10)$$

Equations (9) and (10) define the dynamics of the system, with functions h and f_s depending on the specific cooperative control strategy used. Note that the UAVs' decisions cause the environment state transitions which, in turn, affect the decisions of the UAVs. The dynamics are stochastic due to the stochasticity of $b(t)$ and $\omega(t)$.

The objective of the search mission, which will last T_f units of time, is to locate as many targets as possible while minimizing UAV losses. This can be achieved by cooperative path planning among the multiple UAVs such that the following payoff function is maximized:

$$E\{G(x(T_f)) - \sum_{t=0}^{T_f-1} J(x(t), u(t))\} \quad (11)$$

where the terminal payoff function $G(\cdot)$ and the cost function $J(\cdot,\cdot)$ are defined as

$$G(x(T_f)) := \sum_{i=1}^{N} \pi_v \delta_i(T_f) + \sum_{(x,y)\in E} \pi_t \zeta(x,y,T_f) \quad (12)$$

$$J(x(t), u(t)) := \sum_{i=1}^{M} c(u_i(t)) \quad (13)$$

The positive constants π_v and π_t represent the weight allocated to the importance of UAV safety versus target discovery, while c is a positive-valued function that represents the cost of moving UAV i as designated by $u_i(t)$. Because we use identical UAVs and we assume the same cost for moves in every direction, c is a constant function. The first term in (12) represents the total number of surviving UAVs, while the second term represents the number of identified targets. Thus, the objective of the search problem can be described as maximizing the terminal payoff function $G(x(T_f))$.

Dynamic programming [14] is one possible approach for this optimal path selection problem. However, it is computationally prohibitive because of the large dimensionality of the state space in this problem. Instead, we develop an approximate dynamic programming method using a multi-objective cost function, where the cooperation among UAVs is achieved using the "rivaling force" approach [7].

III. Cooperative Path Planning Method

In this section, we describe a path selection decision function $h(x(t))$, which the UAVs can use for cooperative path planning based on their current information. The decision function is based on the expected rewards associated with each of the three possible paths for the next time step. The reward definition takes into account the following four subgoals: 1) Target Confirmation, 2) Environment Exploration, 3) Threat Avoidance, 4) Cooperation.

The reward obtainable at the next step is called *immediate reward*. However, a UAV should not select a path only with the best immediate reward but a path that will bring more rewards over the long term. Therefore, UAVs use a limited look-ahead policy to select their paths in the proposed path planning method, that is, they also consider longer term rewards in their path selections. Next, we discuss the heuristic estimation of the immediate reward and the long-term reward, respectively.

A. Immediate Reward Estimation

The expected immediate reward for a UAV searching cell (x, y) at time $t+1$, denoted as $\rho(x, y, t+1)$, is the payoff for target confirmation and UAV survival. It is represented as a multi-objective cost function which is a linear combination of four types of rewards corresponding to the four subgoals:

$$\begin{aligned}\rho(x,y,t+1) &= \omega_1\rho_f(x,y,t+1) + \omega_2\rho_e(x,y,t+1) \\ &\quad +\omega_3\rho_t(x,y,t+1)+\omega_4\rho_c(x,y,t+1) \quad (14)\end{aligned}$$

where $\rho_f(x, y, t+1)$ is the target confirmation reward, $\rho_e(x, y, t+1)$ is the environment exploration reward, $\rho_t(x, y, t+1)$ is the threat avoidance reward and $\rho_c(x, y, t+1)$ is the cooperation reward. The definitions for these rewards are given below. By changing $\omega_i, i \in \{1, 2, 3, 4\}$, the relative importance of the four rewards can be scaled.

1) Target Confirmation Reward: To achieve the search objective, the UAVs need to maximize the number of confirmed targets. A UAV will get a reward π_t in one cell if it can confirm a new target there. So the expected target confirmation reward in cell (x, y) at time $t+1$ is defined as:

$$\begin{aligned}&\rho_f(x,y,t+1)\\ &= P(\text{new target confirmation in cell } (x,y) \text{ at time } t+1)\cdot\pi_t\\ &\quad + P(\text{non-target confirmation in cell } (x,y) \text{ at time } t+1)\cdot 0\\ &= P(\zeta(t)=0 \cap \zeta(t+1)=1)\cdot\pi_t\\ &= P(\zeta(t)=0 \cap b_i(x,y,t+1)=1 \cap p(x,y,t)\geq\beta)\cdot\pi_t\\ &= P(\zeta(t)=0)\cdot P(b_i(x,y,t+1)=1)\cdot P(p(x,y,t)\geq\beta)\cdot\pi_t \quad (15)\end{aligned}$$

where β is a constant indicating the minimum probability that $p(x, y, t)$ can take such that the $p(x, y, t+1)$ (generated using update Equation (7)) will be greater than the threshold value θ. The value of β can be obtained using (7) and the specific values of p_c, p_f and θ.

Let A denote the event that a target is actually located in cell (x, y). Using the total probability theorem and $p(x, y, t) = P(A)$, $p_c = P(b_i(x, y, t+1) = 1 \mid A)$ and $p_f = P(b_i(x, y, t+1) = 1 \mid \bar{A})$, we obtain:

$$\begin{aligned}&P(b_i(x,y,t+1)=1)\\ &= P(b_i(x,y,t+1)=1\,|\,A)P(A)+P(b_i(x,y,t+1)=1\,|\,\bar{A})P(\bar{A})\\ &= (p_c-p_f)p(x,y,t)+p_f \quad (16)\end{aligned}$$

Therefore, we get:

$$\rho_f(x,y,t+1) = \begin{cases} [(p_c-p_f)p(x,y,t)+p_f]\cdot\pi_t \\ \quad \text{if } \zeta(x,y,t)=0 \text{ and } p(x,y,t)\geq\beta \\ 0 \quad \text{else}\end{cases} \quad (17)$$

The above equation indicates that each UAV should select a path consisting of cells with high target probabilities.

2) Environment Exploration Reward: Since targets are usually relatively sparse in the practical situations, it is important for the UAVs to explore the environment to obtain new information on potential targets. As discussed before, the certainty value $\chi(x, y, t)$ can be used to drive the UAVs to explore un-searched regions. The environment exploration reward, ρ_e, is defined as the expected certainty increase caused by a UAV's visit to cell (x, y):

$$\begin{aligned}\rho_e(x,y,t+1) &= E[\chi(x,y,t+1)-\chi(x,y,t)]\\ &= 0.5(1-\chi(x,y,t)) \quad (18)\end{aligned}$$

We can see that the environment exploration reward $\rho_e(x, y, t)$ is a decreasing function of $\chi(x, y, t)$. Hence, for exploration purposes, it is better for the UAVs to visit cells with lower certainty values $\chi(x, y, t)$.

It is easy to notice that the target confirmation reward and environment exploration reward are not always mutually compatible. These two imperatives can be viewed as the classic exploration vs. exploitation tradeoff in game theory.

3) Threat Avoidance Reward : Due to the presence of threats, UAVs can be destroyed, resulting in a reduction in the terminal payoff function $G(\cdot)$. The threat avoidance reward ρ_t is defined as the avoided loss in the terminal payoff function if a UAV is not destroyed in cell (x, y) at time $t+1$:

$$\rho_t(x,y,t+1) = (1-k(x,y,t+1))(\pi_v+\bar{n}(t+1)\pi_t) \quad (19)$$

where $\bar{n}(t+1)$ denotes the estimated average number of targets which could be identified by the UAV from time $t+2$ until time T_f. Therefore, $\bar{n}(t+1)$ should become smaller as t increases. Note that $k(x, y, t+1)$ is known because the threat map is time-invariant. To gain threat avoidance rewards, a UAV needs to avoid cells with high threat probabilities.

4) Cooperation Reward : Since UAVs plan their paths independently, it is natural that two or more UAVs may choose the same paths because they all want to obtain the associated high rewards. This will be more pronounced if the UAVs happen to be very close and have overlaps in their candidate positions for the next time step. These possible overlaps in the search paths will waste the team's search effort and cause a reduction in the global payoff function. We include a cost function that penalizes vehicles being close to each other and heading in the same direction so as to reduce the possible overlaps. In this paper, we utilize the "rivaling force" based method to generate the cooperation cost function. The cooperation reward, ρ_c, is defined as the negative of the "rivaling force", F_i:

$$\rho_c(x,y,t+1) = -F_i(x,y,t+1) \quad (20)$$

where $F_i(x, y, t+1)$ is a function of other vehicles' positions $\lambda_j(t+1)$ and orientations $o_j(t+1), j \in \{1, 2, \ldots, N\}$, $j \neq i$. Detailed information regarding the generation of the rivaling force function F defined by (27) is given in Section IV.

B. Long-Term Reward Estimation

The long-term expected reward function is used to steer UAVs away from decisions that may yield good immediate payoffs but reduced benefits in the long-run. The long-term expected reward, denoted by $\phi(x, y, t+1)$, is defined as the maximum reward accumulated by following a path starting at cell (x, y) over step $t+1$ to $t+T$, $T \geq 1$. Since each UAV has three candidate positions to go in the next time step, it has 3^{T-1} possible paths over time $t+1$ to $t+T$. The expected reward for path $j \in [1, 2, \ldots, 3^{T-1}]$ is denoted as $\phi^j(x, y, t+1)$ and is the sum of the accumulated rewards in the path and the expected future reward after $t+T$. Let (x_m^j, y_m^j) denote a cell in path j which the UAV will visit at time $t+m$, where $1 \leq m \leq T$. The reward obtained by the UAV in that cell at time $t+m$ is given by:

$$\begin{aligned}&\rho_l(x_m^j, y_m^j, t+m)\\ &= \left[\prod_{s=1}^{m-1}[1-k(x_s^j, y_s^j, t+s)]\right][\omega_1\rho_f(x_m^j, y_m^j, t+m)\\ &\quad +\omega_2\rho_e(x_m^j, y_m^j, t+m)+\omega_3\rho_c(x_m^j, y_m^j, t+m)] \quad (21)\end{aligned}$$

where $\rho_f(x_m^j, y_m^j, t+m)$ is the target confirmation reward in cell (x_m^j, y_m^j) at time $t+m$, $\rho_e(x_m^j, y_m^j, t+m)$ is the environment exploration reward and $\rho_c(x_m^j, y_m^j, t+m)$ is the cooperation reward. The product $\prod_{s=1}^{m-1}[1-k(x_s^j, y_s^j, t+s)]$ gives the probability that the UAV will still be alive to reach that cell and obtain the reward. Due to the fact that $\rho_l(x_m^j, y_m^j, t+m)$ depends on $p(x_m^j, y_m^j, t+m-1)$ and the positions and orientations of UAVs at time $t+m$, the expected reward cannot be known at decision time step t when $m > 1$. Hence, we need to find a way to estimate these values. Here we simply use the corresponding known values at time t to substitute the unknown values over time $t+1$ to $t+T-1$ in (21). Thus we get the heuristic estimate of ρ_l as:

$$\begin{aligned}&\bar{\rho}_l(x_m^j, y_m^j, t+m)\\ &= \prod_{s=1}^{m-1}[1-k(x_s^j, y_s^j, t+s)][\omega_1\rho_f(x_m^j, y_m^j, t+1)\\ &\quad +\omega_2\rho_e(x_m^j, y_m^j, t+1)+\omega_3\rho_c(x_m^j, y_m^j, t+1)] \quad (22)\end{aligned}$$

Meanwhile, the future reward after time $t+T$ following path j can be denoted as:

$$\rho_a^j = \prod_{s=1}^{T}[1-k(x_s^j, y_s^j, t+s)](\pi_v + \bar{n}(t+T)\pi_t) \quad (23)$$

So, the total expected reward for path j can be denoted as

$$\phi_j(x, y, t+1) = \sum_{m=1}^{T}\bar{\rho}_l(x_m^j, y_m^j, t+m) + \omega_4\rho_a^j \quad (24)$$

where ω_4 is the same weight as in (14). And the expected long-term reward $\phi(x, y, t+1)$ at cell (x, y) is

$$\phi(x, y, t+1) = \max_{j\in[1,2,\ldots,3^{T-1}]}\phi_j(x, y, t+1) \quad (25)$$

It is obvious that, when $T = 1$, the expected long-term reward at cell (x, y) is the same as the immediate expected reward given by (14). This definition of long-term reward derives a way to allow the UAVs to look ahead in making their decisions.

IV. Coordination Method

As discussed before, cooperation between vehicles does not arise naturally in the proposed decentralized scheme since every vehicle tries to optimize its own behavior. Therefore, we develop a real-time approach to realize cooperative search using the concept of rivaling force developed in [7]. The main idea is to avoid simultaneously searching a cell by more than one UAVs. This is accomplished by treating the areas around a UAV as "soft obstacles" to be avoided in other vehicles' path selections. The rivaling force exerted by a UAV on a neighboring UAV is obtained by a type of *artificial potential field* method [15]. At each step, a UAV considers in its path decision the overall rivaling force which it is exerted upon it by other vehicles.

If UAV i plans to visit cell (x, y) at time t, it will receive a rivaling force exerted by other vehicles. The magnitude of the rivaling force coming from vehicle j depends on the vehicle j's position and orientation at time t. We can obtain the minimum number of steps that vehicle j would need to reach cell (x, y), denoted as $l_j(x, y, t)$, using vehicle j's position $\lambda_j(t)$ and orientation $o_j(t)$. This variable reflects the cost for vehicle j to search (x, y) in the near future. Because the UAVs cooperate to achieve a group objective, a higher cost for vehicle j searching (x, y) makes it more appropriate for vehicle i to search the cell. In this case, the rivaling force exerted by vehicle j upon vehicle i will be small. Using an approach similar to the artificial potential field method, we define the force exerted by vehicle j on vehicle i in cell (x, y) at time t as:

$$F_{ij}(x, y, t) = \frac{1}{l_j(x, y, t)} \quad (26)$$

The total rivaling force received by UAV i for moving to a cell (x, y) at time t can be given as

$$F_i(x, y, t) = \sum_{j=1, j\neq i}^{N} F_{ij}(x, y, t) \quad (27)$$

Note that different UAVs may receive different rivaling forces when located in the same cell because of differences in the positions/orientations of neighboring UAVs. This force is a penalty for UAV i for entering a cell that is a suitable selection for other vehicles.

V. Simulation Results

To assess the performance of the approach described above, we simulated a team of five UAVs searching a 20×20 cellular environment with 20 targets and 5 threats. There is no a priori topographical information and no other sources of information on target distribution. Thus $\chi(x, y, 0) = 0$ and $p(x, y, 0) = 0.5$ for each cell (x, y) in the environment. For all the simulation runs in this paper, the homogeneous targets and threats are randomly assigned to the environment while the UAVs' initial locations and orientations are held constant. The threats' attack regions are set to $\phi = 2$, the threat probability $p_{kill} = 0.2$, the UAVs' sensor detection rate $p_c = 0.8$ and the sensors' false alarm rate $p_f = 0.1$. All simulations were run for 250 time steps.

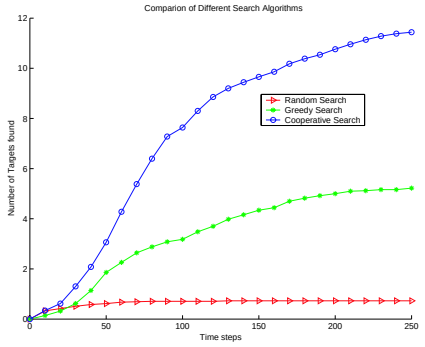

Fig. 2. Number of targets found as a function of time: Comparison with greedy and random search algorithms. All data is averaged over 50 runs.

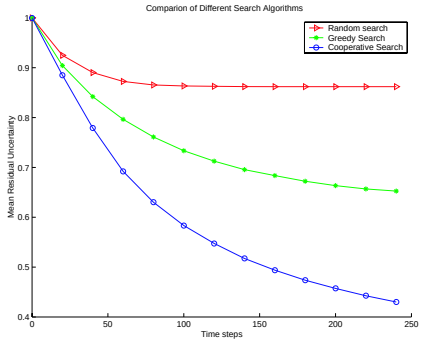

Fig. 3. Residual uncertainty as a function of time: Comparison with greedy and random search algorithms.

We used two measures of performance:

- Number of targets found up to the current time-step t, i,e., the number of cells with $\zeta(x, y, t) = 1$.
- The residual uncertainty left in the environment.

$$U(t) = \frac{\sum_{(x,y)\in E}(1 - \chi(x, y, t))}{\sum_{(x,y)\in E}(1 - \chi(x, y, 0))} \tag{28}$$

The performance of the proposed cooperative search algorithm was compared to that of a random search and a greedy search algorithm. In the random search strategy, the vehicles do not use any available information about the target and threat distributions but simply move in a random direction within the search region. In the greedy search strategy, the vehicles move at each step to the candidate cells with highest reward for target confirmation and to avoid damage to the UAV. The reward can be obtained by using the immediate reward definition (14) in section III-A with $\omega_4 = 0$. In this strategy, the UAVs can share information about where other UAVs have searched so that all the UAVs share the same maps but they perform little distributed path selection in order to coordinate their actions. In the proposed cooperative search method, the vehicles will move to the cells with the highest reward defined as (25), where we set $T = 3$.

Figure 2 shows the number of targets found by each algorithm as a function of time. Figure 3 shows how the mean residual uncertainty in the environment declines with time for different search algorithms. This value gives a good indication of the coverage of the search region. It is obvious that the cooperative search method provides a significant improvement on both performance measurements.

VI. Conclusion

In this paper, we have presented a formulation for the cooperative search problem. The objective of the search mission is to find and confirm as many targets as possible while minimizing UAV losses. A key issue for this cooperative control problem is the design of a cooperative scheme such that the team of vehicles perform path planning cooperatively based on the information they get. We develop a cooperative path planning algorithm based on a heuristic multi-objective cost function method, which can overcome the computational complexity of looking for an optimal dynamic programming solution. The cooperation among UAVs is achieved using a rivaling force approach. The simulation results demonstrate that the heuristic approach is an intuitive and computationally efficient method to tackle the cooperative search problem.

References

[1] K. Nygard, P. Chandler, and M. Pachter. Dynamic network optimization models for air vehicle resource allocation. In *Proc. of the ACC*, pages 1853–1856, June 2001.

[2] P. Chandler, M. Pachter, and S. Rasmussen. UAV cooperative control. In *Proc. of the ACC*, pages 50–55, June 2001.

[3] T. McLain, P. Chandler, S. Rasmussen, and M. Pachter. Cooperative control of UAV rendezvous. In *Proc. of the ACC*, pages 2309–2314, June 2001.

[4] J. S. Bellingham, M. Tillerson, M. Alighanbari, and J. P. How. Cooperative path planning for multiple UAVs in dynamic and uncertain environment. In *Proceedings of the 41th IEEE Conference on Decision and Control*, pages 2816–2822, December 2002.

[5] G. Arslan, J. D. Wolfe, J. Shamma, and J. L. Speyer. Optimal planning for autonomous air vehicle battle management. In *Proceedings of the 41th IEEE Conference on Decision and Control*, pages 3782–3787, December 2002.

[6] V. Ablavsky and M. Snorrason. Optimal search for a moving target: a geometric approach. In *AIAA Guidance, Navigation, and Control Conference and Exhibit*, Denver, CO, August 2000.

[7] M. Polycarpou, Y. Yang, and K. Passino. A cooperative search framework for distributed agents. In *Proceedings of the 2001 IEEE International Symposium on Intelligent Control*, pages 1–6, 2001.

[8] Y. Yang, M. Polycarpou, and A. Minai. Opportunistically cooperative neural learning in mobile agents. In *Proceedings of the 2002 World Congress on Computational Intelligence*, pages 2638–2643, May 2002.

[9] Y. Yang, A. Minai, and M. Polycarpou. Decentralized cooperative search in UAV's using opportunistic learning. In *AIAA Guidance, Navigation, and Control Conference and Exhibit*, August 2002.

[10] M. Flint, M. Polycarpou, and E. Fernandez. Cooperative control for multiple autonomous UAV's seaching for targets. In *Proceedings of the 41th IEEE Conference on Decision and Control*, pages 2823–2828, December 2002.

[11] M. L. Baum and K. M. Passino. A search-theoretic approach to cooperative control for uninhabited air vehicles. In *AIAA Guidance, Navigation, and Control Conference and Exhibit*, August 2002.

[12] M. Moors, R. Simmons, W. Burgard, D. Fox and S. Thrun. Collaborative multi-robot exploration. In *Proc. Intl. Conf. on Robotics and Automation*, May 2000.

[13] E. Milios, I. Rekleitis and G. Dudek. Accurate mapping of an unknown world and online landmark positioning. In *Proc. of Vision Interface*, pages 455–461, 1998.

[14] D. P. Bertsekas. *Dynamic Programming and Optimal Control: Vol. 1*. Athena Scientific, MA, 1995.

[15] O. Khatib. Real-time obstacle avoidance for manipulators and mobile robots. In *International Conference on Robotics and Automation*, pages 500–505, St. Louis, March 1985.