Co-evolutionary Search Path Planning under Constrained Information-Sharing for a Cooperative Unmanned Aerial Vehicle Team

Jean Berger and Jens Happe

***Abstract*—Mobile cooperative sensor networks are increasingly used for surveillance and reconnaissance tasks to support domain picture compilation. However, efficient distributed information gathering such as target search by a team of autonomous unmanned aerial vehicles (UAVs) remains very challenging in constrained environment. In this paper, we propose a new approach to learn resource-bounded multi-agent coordination for a multi-UAV target search problem subject to stringent communication bandwidth constraints in a dynamic uncertain environment. It relies on a new information-theoretic co-evolutionary algorithm to solve cooperative search path planning over receding horizons, providing agents with mutually adaptive and self-organizing behavior. The anytime coordination algorithm is coupled to a divergence-based information-sharing policy to exchange high-value world-state information under limited communication bandwidth. Computational results show the value of the proposed approach in comparison to a well-known reported technique.**

I. INTRODUCTION

COST-effective construction of a recognized air picture (RAP) for military local area surveillance and reconnaissance missions is often critical to ensure and maintain situational awareness. In many cases, given the low cost and risk associated with resource utilization, the related RAP process increasingly relies on a team of mobile sensor agents or unmanned aerial vehicles (UAVs) (both terms are used interchangeably) to perform cooperative search, closing the gap between information need and information gathering. Early work on related search problems emerges from search theory [1], [2]. Proposing a mathematical framework and models leading to analytical solutions for simple static formulations, most efforts have progressively been devoted to algorithmic contributions to handle more complex dynamic problem settings [3]. However, reported work for these problems mainly focused on centralized search while assuming search effort to be infinitely divisible between cells, making it difficult to solve realistic path planning problems [3] in cases where cell target containment probability is sparsely distributed. Search theory solutions mostly relate to the effort allocation decision problem rather than path construction. Robot motion planning alternatively explored search path planning, primarily providing constrained shortest path type solutions for coverage problem instances [4]-[5]. In this setting, teammates must self-organize, autonomously manage their own resources, and coordinate their behavior to achieve a common global objective. Typical decision problem formulations and solutions are presented in [6]–[9]. Recent extensions to this work further address the critical information-sharing dimension of the cooperative search planning problem [10]–[13]. These problems capture to various degrees some simplified decentralized partially observable Markov decision process combining communication (informed-sharing) and control decisions (COM-DEC-POMDP) [14], [15] which has proven to be NEXP-complete [16]. Explicit solutions proposed for multiple-UAV cooperative search path planning are numerous. Some early approaches simply reduce computational complexity by relaxing some hard constraints to keep the problem tractable. Methods inspired from search theory propose procedures based on graph/tree search techniques (e.g. branch and bound, A*) but the determination of good heuristics to compute tight bounds for solution quality estimation remains very difficult [3] making them less appealing. Liao et al. [10] recently proposed a search path planning approach combining a cooperative path planning control solution coupled to a specific predetermined information-sharing policy. The proposed constrained solution and alternative approaches reported in [17]–[19] simply ignore planning action coordination through explicit communication of intents (agent path plans) anticipating insufficient bandwidth to support computer intensive problem-solving. The approach further assumes unbounded agent communication bandwidth over local neighborhood to mutually share entire sensor readings history with close neighbors. Current solutions proposed for UAV teams paid little attention to path planning coordination strategies designed to take advantage of limited communication bandwidth constraints more explicitly.

This paper presents a co-evolutionary search path planning approach under constrained information-sharing for a cooperative unmanned aerial vehicle team. It is primarily inspired from the information-theoretic framework reported in [17]–[19] and an extension to the information sharing policy presented in [10]. It extends previous work reported on multi-UAV target search by learning resource-bounded multi-agent coordination for a new problem, considering an

Jean Berger is with Defence R&D Canada – Valcartier, Quebec City, PQ, Canada (phone: 418-844-4000 x4645; fax: 418-844-4538; e-mail: jean.berger@drdc-rddc.gc.ca).

Jens Happe is with MacDonald, Dettwiler & Associates Ltd. Richmond, B.C., Canada (e-mail: jhappe@mdacorporation.com).

open-loop with feedback decision model with a rolling horizon, multiple objectives, heterogeneous agents, limited computational resources and communication bandwidth, as well as communication cost, in a time-constrained uncertain environment. It concurrently deals with multiple constraints, departs from predetermined control search plan policy based on implicit or passive plan coordination [19], and proposes a framework to construct joint path plans providing team flexibility and adaptation by co-evolving multiple agent behaviors simultaneously while mitigating communication needs and cost. The main contribution lies on a new information-theoretic co-evolutionary algorithm to solve cooperative search path planning over receding horizons, providing agents with mutually adaptive and self-organizing behavior. The anytime algorithm is coupled to an extended information-sharing policy to exchange world-state information and planned agent intents under constrained communication bandwidth.

The content of the paper is structured as follows. Section II first introduces problem definition, describing the main characteristics of a new cooperative information-gathering problem involving multiple UAVs to carry out target search. Then the main solution concept for the targeted problem is presented in Section III. A two-step decomposition scheme to sequentially achieve state estimation through information-sharing and cooperative path planning is outlined. Section IV reports and discusses computational results comparing the value of the proposed method to alternative techniques. Finally, a conclusion is given in Section V.

II. Problem

A. Description

A hierarchical multi-objective problem in which a team of heterogeneous UAVs cooperatively searches stationary targets in a bounded environment over a given time horizon is considered. The first objective is to maximize information gain or equivalently to minimize uncertainty or entropy [20] about target occupancy over the grid, the second consists in minimizing target discovery time, and the third aims at minimizing resource utilization, namely, energy consumption. The proposed hierarchical objective structure refers to lexicographic ordering, ranking solution quality along the described objectives, respectively in that order. The search environment defines a two-dimensional grid of N cells, populated by non-cooperative stationary targets with unknown locations and cardinality. Based on prior domain knowledge, individual cells are characterized by an initial probability of target occupancy, which defines an agent's cognitive map. Given an initial team configuration, n autonomous UAVs synchronously explore the environment, acting as stand-off imperfect sensors gathering observations while periodically exchanging state and plan information with one another through peer-to-peer (unicast) communication. UAVs perfectly and synchronously share information (observations, intents) with neighbor agents during each time step (visit). Information-sharing is subject to limited on-board computational resources and, range and bandwidth -constrained communication. Vehicles are assumed to fly at a constant velocity and at slightly different altitudes to avoid colliding with each other. Cooperative search consists in jointly constructing agent path plans to minimize team uncertainty (entropy) over cell target occupancy. As the system is distributed, each agent (vehicle) must continuously build and update its own cognitive map through local observations and information exchange with teammates while aligning its behavior toward reaching global team objective. A UAV's cognitive map refers to a knowledge base capturing local environment state representation, reflecting target occupancy belief distribution, positions and orientations of neighboring agents, its own direction and position, resource level, sensor observations and their sources (observing agent), as well as a past communication log with other agents. A typical agent cognitive map at a given time is illustrated in Figure 1.

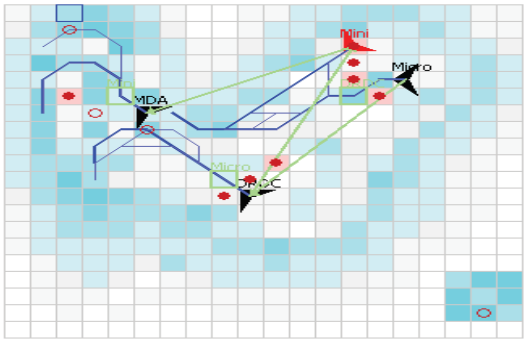

Figure 1. Agent uncertainty/cognitive map at time step t. Agent belief on target cell occupancy is displayed using color shades: the darker the cell the larger the belief. Projected agent plans are represented as possible paths and peer-to-peer communications as straight lines. Cells with filled (void) circles refer to discovered (unknown) targets.

In the current setting an agent has prior knowledge about its teammates (e.g. sensor observation models, maximum communication range and other properties). Bayesian filtering is used locally for data aggregation/fusion.

B. Communication

The agent team primarily behaves as a particular vehicular ad hoc network. Vehicles have self-localization capability, can recognize neighboring agents, and rely on a unicast or peer-to-peer communication scheme based on perfectly reliable communication channels. The model assumes range- limited communication restricting neighborhood interaction, and a finite bandwidth bounding the number of messages to be exchanged with a neighbor per time step. Agent communications with neighbors take place concurrently delivering/receiving messages on separate channels in parallel. Encoded as messages, communication decisions translate into observation streams, beliefs and/or intents to be shared. Based on the aforementioned small-world assumption, we also assume

instantaneous message-passing (negligible network latency). Energy cost supporting information exchange is quadratic in terms of the distance r connecting two agents ($\alpha r^2 + \beta$).

III. SOLUTION CONCEPT

The proposed solution to cooperative multi-UAV search path planning relies on a co-evolutionary information gathering approach. It extends previous work reported on multi-UAV target search [17]-[18], [10], [19], by learning resource-bounded multi-agent coordination considering an open-loop with feedback decision model over a receding horizon, multiple objectives, heterogeneous agents, limited computational and communication resources, as well as communication cost, in a time-constrained uncertain environment. The solution provides a framework to construct joint path plans providing team flexibility and adaptation by co-evolving multiple agent behaviors simultaneously while mitigating communication need and cost. The new information-theoretic co-evolutionary algorithm solves cooperative search path planning over receding horizons, providing agents with mutually adaptive and self-organizing behavior.

The main concept is based on an information-theoretic framework to address information-sharing and cooperative path planning as simultaneous problem-solving for communication and control decision variables is intractable. The agent information-sharing policy is based on maximum information/belief divergence in exchanging the most valuable information on state estimation with neighbors. It is based on the premise that maximizing belief consistency (common belief-sharing) among cooperative team mates at each time step should likely improve joint planning solution. Path planning coordination is then ensured through the co-evolution of population individuals expressed as a sequence of control actions (physical moves) over a given horizon, whose fitness is defined in terms of a weighted combination of expected information gain (differential entropy) capturing shared-rewards among neighboring agents current "best plan". The appeal of co-evolution lies in its natural ability to handle computational complexity, to generate multiple solutions, and to provide adaptability through 'anytime' behavior.

Each time step, an agent's decisions rely on a sequential two-stage decomposition process, namely, information–sharing (past observations) and planning. Decisions are made on what information (observations stream) to exchange at the beginning of the current episode, and the move (control action) to be executed during the next time interval. Moves planned over the next (receding) plan time horizon are also computed or revised. The time interval Δt_t is divided in two time subintervals, accounting for information-sharing and planning respectively, as shown in Figure 2.

High-value or latest observations are first exchanged between an agent and its close neighbors, the remaining time being devoted to cooperative planning, which in turn is subdivided into cycles, involving multiple planning iterations coupled to asynchronous periodic communication of intents. In order to reinforce agent synchronization,

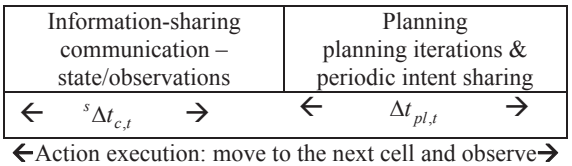

Figure 2. A two-stage decision process

we assume that each process stage has a constant predetermined duration, which imposes temporal constraints on information state communication and planning. Decision-making for the current episode is occurring concurrently with the execution of the action planned in the previous episode. Therefore, during time interval Δt_t, the vehicle executes a previously planned action moving to the next cell, and makes an observation. Combined temporal and communication bandwidth constraints along with message type sizes define a maximum number of messages to be exchanged in both directions during a single episode. Accordingly, a bound on the number of messages have been considered for the information-sharing stage of the decision process. The process can be summarized as follows:

t=1
For agent *i*= 1..*n*
  Initialize Population(*i*)
Repeat -- agent search path planning behavior
  Control action execution (planned at *t-1*)
  Information_Sharing ($t(^{s}\Delta t_{c,t})$) – Stage I
  Path_Planning ($t(\Delta t_{pl,t})$) – Stage II
  Observation and cognitive map update
  End of episode *t; t=t+1*
Until (end of search mission horizon: *t=L*)

Section III A and III B give further details on the related information-sharing and path planning processes.

A. Information-Sharing

A Given problem complexity, sequential communication and control actions decision processes relying on heuristic policies for observation and intents –sharing have been considered. Special emphasis is put on the information-sharing heuristic policy used to govern observation exchanges and update agent local beliefs in Stage I, as the description of a fairly simple intent-sharing policy is deferred to Section III B. Internal to the path planning process, intent-sharing is based on periodic "best plan" communication to barely support plan coordination.

Inspired from information theory, the proposed information-sharing (observation) policy is based on maximum belief divergence in exchanging the most valuable

information on state estimation with neighbors. The rationale is that cooperative autonomous agents maintaining belief consistency (some consensus) are likely to make decisions reflecting desirable behaviors in the best interest of the team. A simple measure of information divergence (relative entropy or Kullback-Leibler [20] could have been used alternatively) between agent i and j (div_{ij}) over the grid can be defined as follows:

$$div_{ij} = \frac{\sum_{c \in G: E_0(c) \neq 0} (p_i(c) - p_j(c))^2}{\sum_{c \in G E_0(c) \neq 0} 1} \qquad (1)$$

where $p_i(c)$ represents agent i's belief or probability that a target occupies cell c, and $E_0(c)$, the initial entropy of cell c. The larger the differential belief on cell occupancy, the higher the value of the related observations to be exchanged in order to minimize divergence. For instance, a "confirmed " target occupancy state which results from an agent belief reaching some threshold value or through message-passing, is more likely to eventually have its related observations be diffused to uninformed neighboring agents in priority due to a sudden belief change, further increasing divergence and therefore avoiding any future unnecessary visits. Expected gain from agents will generally be marginal for observations (possibly quite dissimilar) showing near similar cell target occupancy beliefs. However, it should be noticed that similar belief shared by two agents does not necessarily mean identical shared observations, as significantly different observation streams may still lead to the same belief. It is assumed that a maximum number of $M_{ij}(s)$ messages can be exchanged between two agents i and j during a given episode. The proposed divergence-based information-sharing policy for a member i of a team of n agents searching over a grid of N cells, can be summarized as follows:

Information-Sharing (t)

Update neighborhood membership of agent i: $Neigh_t(i)$

For all $j \in Neigh_t(i) / (Neigh_t(i) \cap Neigh_{t-1}(i))$ (new neighbor) do

If (j is NOT a recent neighbor: $j \notin \bigcup_{\tau=1}^{\tau'} Neigh_{t-\tau}(i)$) then

exchange beliefs between i and j on respective (possibly disjoint) regions of interest. (set of reachable cells generated by the projection of future moves over a given horizon – Fig. 3).

compute belief divergence over respective regions

sort belief divergence values in decreasing order

exchange up to $M_{ij}(s)$ observations for cells with largest belief divergence first

else

exchange latest observations with j

${}^iT_j = 1$

schedule next interaction: schedule(j)= $t + {}^iT_j$

For all $j \in Neigh_t(i) \cap Neigh_{t-1}(i)$ (old neighbor) do

If (t = schedule(j))

If (j was a neighbor of i longer than $2\sqrt{N/n\pi}$) then

i.e. $j \in \bigcap_{\tau=t-2\sqrt{N/n\pi}}^{t} Neigh_\tau(i)$

$Neigh_t(i) = Neigh_t(i) / \{j\}$

If (previous exchange is incomplete) then

based on regional beliefs exchanged:

proceed with high-value observations exchange over remaining region (if any)

set ${}^iT_j = 1$ until observations exchange is complete

else -- previous exchange completed

exchange ${}^iT_j < M_{ij}(s)$ latest observations with j:

$${}^iT_j = Max\left(1, \min\left\lfloor \left(\frac{\sqrt{E_0/n}}{2}, R_i, R_j, t\_div\_thr \right) \right\rfloor\right)$$

schedule next interaction: schedule(j)= $t + {}^iT_j$

else -- $t \neq$ schedule(j) [asynchronous communication]

If (i received a stream of k observations from j) then

i exchanges its k < $M_{ij}(s)$ latest observations with j

schedule next interaction: schedule(j)= $t + {}^iT_j$

Each time step, the agent updates its neighborhood membership processing concurrently new and old members. Then, agent communication transactions with new and old agent neighbors occur concurrently over the information-sharing period ${}^s\Delta t_{c,t}$. New members exchange respective cell target occupancy beliefs over a limited region, compute and sort belief divergence for these cells, and then gradually exchanges up to $M_{ij}(s)$ observations proceeding with cells having the largest belief divergence first. Should observation exchanges not be completed, remaining observations would gradually be communicated over the next few episodes. A recent former neighbor qualified as a new member would however exchange the latest observation only as belief divergence should be marginal. Neighbor type status is then modified to "old" before scheduling the next interaction to the next time step. It should be emphasized that belief exchange occurs over the recipient's region of interest, possibly overlapping with the sender's region of interest. An agent's region of interest is the set of reachable cells generated by the projection of future moves over a given horizon as shown in Figure 3.

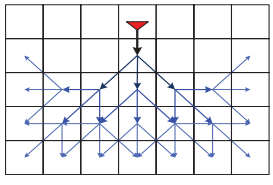

Figure 3 – Agent's region of interest displayed as forward move projection span (possible paths) over a 3–step time horizon.

Concurrently to the handling of new neighbors, old ones are processed according to previously scheduled interactions

determined either locally or, asynchronously through a neighbor agent. On the occurrence of interactions set asynchronously by neighbors, observations are mutually exchanged and the new planned interactions with those neighbors are rescheduled locally. When such a locally scheduled interaction with a neighbor occurs, some conditions are examined. Long-time (persistent) interacting agents beyond a certain time interval (delay) are first explored in order to share indirectly reported (but still unshared) observations from mutually dissimilar neighbors during the next cycle. In that case, a neighbor status is changed from 'old' to 'new'. Otherwise, any incomplete information-sharing process initiated during the latest episode is further continued, or alternatively, the ${}^{i}T_{j}$ latest observations are mutually shared with one another before scheduling a new rendezvous with a revised ${}^{i}T_{j}$. The latter is based on initial entropy E_0, communication ranges (R_i, R_j) and a constant period (t_div_thr).

B. Cooperative Search Path Planning

An open-loop agent solution to a multi-objective problem subject to limited computational resources and communication constraints is gradually constructed at each episode and progressively extended over a receding *T*-step horizon, while adjusting its path plan based on additional feedback. Episodic search path planning relies on co-evolution to learn agent coordination. Agents evolve their own population of individuals while sharing information about neighbor agent intents. Individuals are represented as chromosomes encoding for a given time step, a feasible path plan expressed as a sequence of intended control actions (physical moves a_{t+1} ... a_{t+T}) to be executed over a specific time horizon *T* (Figure 4).

a_{t+1}	a_{t+2}	a_{t+3}	…	a_{t+T-1}	a_{t+T}

Figure 4. Individual path plan representation at time step *t*.

Agents co-evolve their own path plan individuals through natural selection, recombination and mutation mechanisms over successive generations while periodically (T_{GA} or *max_gen* generations) exchanging their best individual with neighbor agent populations. An individual fitness is determined by combining(coalescing) its own local path plan along with the latest best-known neighbor agents' plans and evaluating the resulting joint plan. Here the fitness evaluation of a plan is based on the agent's interaction with other agents. The cycle is then repeated until the end of the planning period. As a result, the first action of the best computed path plan is executed at the next time step. The agent co-evolutionary anytime procedure can be summarized as follows:

```
Path_Planning(t)
  Adjust/update population individuals to reflect the latest
  agent's decision (current move)
  Repeat
    gen=0
    Repeat for each new generation
      For all new "best plan" message received
          adjust individual's fitness value
      Evolve Population - build a new generation
          generate λ new offspring using genetic
          operators (selection, recombination,
          mutation)
          evaluate fitness of new individuals
          eliminate the λ worst individuals of the
          expanded population
      gen=gen+1
    Until (end of a cycle (T_GA) or gen = max_gen)
    intent-sharing (send current 'best plan' to neighbors)
  Until (end planning period)
  Return (best computed path plan from Population)
```

The first step consists in adjusting the population members emerging from the last episode to account for current move execution, conserving feasible (consistent) individual path plans only, and generating randomly new individuals (path plans with *T* control actions) to maintain population size. Feasible path plans are simply shifted by one step to reflect the current time, and a new action for time step *T* (plan horizon) is drawn randomly from the three possible actions, with a probability proportional to its related information gain. The outer loop captures multi-agent coordination learning and intent-sharing between neighbor agents. The underlying evolutionary approach and its key components are mainly outlined in the inner loop. The steady-state evolutionary algorithm consists for each generation in expanding the population by λ offspring using genetic operations, and then removing the worst λ individuals to restore normal size. Recombination and mutation operators are sequentially applied with probability p_{Xover} and p_{Mut} ($p_{Xover} + p_{Mut} = 1$), respectively, until all λ new individuals are generated. Probability parameters are generally determined to balance search intensification and diversification. The process is repeated until a convergence criterion/condition is met (e.g. maximum number of generations or a threshold in solution quality improvement). Fitness and genetic operators are further defined below.

The initial population of path plan individuals is generated randomly, selecting control actions proportionally to the information gain shown over single-step horizons. Well-known heuristic methods could be further exploited to generate good individuals, allowing co-evolution to take advantage, and to at least match naturally the performance of the best reported procedures.

1) Fitness:

Fitness reflects an individual's propensity to reproduce. The individual's fitness should be determined and ranked

according to the hierarchical objective structure proposed in Section II A and based on lexicographic ordering: maximize information gain, minimize target discovery time and energy consumption. As dissimilar fitness values over population individuals are likely, we approximate the global fitness function along entropy minimization (or information gain maximization) to more easily rank individual scores. This approximation makes sense to the extent that target discovery rate is somewhat correlated to entropy and that resource utilization is mainly controlled by the predetermined information-sharing policy.

Fitness is defined in terms of a weighted combination of expected information gains (differential entropy contributions) capturing shared individual path plan and neighboring agents current "best plan" (neighbors' population best-known individual) respective expected rewards, while ignoring possible additional benefits that could have resulted from intermediate explicit outcome communication or exploitation. An agent's individual reward refers to the local information gain or entropy reduction expected by executing its path plan over the next plan horizon *T*. Subject to range and temporal constraints, communication of respective best plan (intents) take place periodically among neighboring agents after a certain number of generations, to update agent individual's fitness value. The nominal fitness function for an individual *ind* in agent population *i* characterized by a local path plan or sequence of actions $\{a_{it}\}$ (resulting in positions/expected information gain value pairs $\{y_{it}(a_{it}), {}^{i}g(y_{it})\}$) along with respective neighbors' best-known path plan and information gain ($\{y_{jt}, {}^{j}g(y_{jt})\}_{j\in Neigh(i)}$) over the next *T* time steps, is defined as follows:

$$fitness_{ind\in Pop(i)}(\{y_{jt}, {}^{j}g(y_{jt})\}_{j\in Neigh(i)\cup\{i\}}) = \sum_{j\in Neigh(i)\cup\{i\}} {}^{j}R$$
$$= \sum_{c\in G} \sum_{j\in Neigh(i)\cup\{i\}} \frac{{}^{j}g(c)}{\sum_{k\in Neigh(i)\cup\{i\}} {}^{k}g(c)} {}^{j}g(c) \quad (2)$$

$${}^{j}g(c) = {}^{j}E_0(c) - {}^{j}\overline{E}_T(c) = {}^{j}\Delta E(c) \quad (3)$$

$${}^{j}\overline{E}_T(c) = \sum_{\{z_{jt}(c)\}\in\{0,1\}^T} \left[\prod_{t=1}^{T} p(z_{jt}(c) \,|\, y_{jt}(a_{jt}), o_{jt}(a_{jt})) \quad {}^{j}\varepsilon_T \right] \quad (4)$$

$${}^{j}\varepsilon_T = {}^{j}E_T(p_T(X(c)=1 \,|\, \{z_{jt}(c)\}, \{y_{jt}, o_{jt}\})) \quad (5)$$

c: cell element of the grid: $c \in G = \{1,2,...,N\}$

T: time horizon

i: agent team member $i \in \{1,2,...,n\}$

a_{it} : action of agent *i* executed during time interval *t*
$a_{it} \in \{ahead, right, left\}$

$\{a_{it}\}_{1..t}$: path plan of agent *i* over history 1..*t*

y_{it} : position (cell element number) of agent *i* over time interval *t* as a result of action a_{it}

o_{it} : orientation of agent *i* over time interval *t* as a result of action a_{it}

$z_{it}(c)$: observation outcome of cell *c* by agent *i* at the end of time interval *t*, $z_{it}(c) \in \{0,1\}$

$\{z_{it}\}$: sequence of observations by agent *i* over history 1..*T*

$p(z_{it}(c))$: probability to observe outcome $z_{it}(c)$

$p_t(X(c)=1)$: belief of cell occupancy of cell *c* at the end of time interval *t*

${}^{i}E_T(p_T(X(c)=1 \,|\, \{z_{it}(c)\}, \{y_{it}, o_{it}\}))$: agent *i* entropy of cell *c* at *t*=*T* given the sequence of actions $\{a_{it}\}$ and observations $\{z_{it}\}$ over history 1..*T*

Given that the global team entropy is unknown by team members, it can be noticed from the fitness function that interfering agents (planning to visit the same cell) share information gain proportionally to their relative contribution to better account for the reduced team information gain. Two agents exploring the same cell with identical anticipated benefits will ultimately see their respective cell reward reduced by half. Agent's *j* expected information gain ($ {}^{j}g(c) $) reflects the local entropy reduction expected over cell *c* by executing the open-loop agent's path plans over all possible *T*-step scenarios. It can be expressed as the difference between current entropy ($ {}^{j}E_0(c) $) and expected entropy ($ {}^{j}\overline{E}_T(c) $) resulting from possible plan execution outcomes.

2) Selection: In order to balance and control selection pressure for breeding purposes, fitness values are sorted out and normalized using a linear ranking scheme [21], scaling respective values based on rank. Individual selection for mating purposes is then ensured following a fitness-proportional scheme [22].

3) Recombination:

This genetic operation recombines chromosomes from two selected parents in order to create a child. The proposed recombination operator X_path breeds two parent individuals sharing a compatible crossover point and generates an offspring by connecting together head and tail path segments inherited from both parents respectively, truncating control actions when the chromosome length exceeds the planning horizon *T* (see Figure 5) or appending missing control actions using a greedy method (selecting moves with maximum one-step information gain) to complete the solution when necessary. A second child can be generated in the same way by swapping parents. The

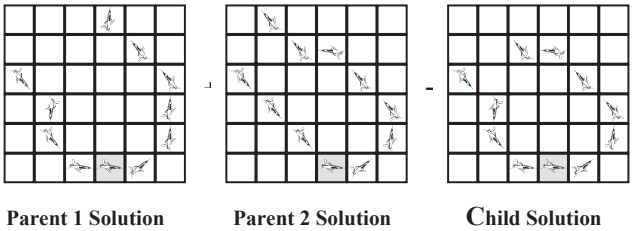

Figure 5. Crossover operator X_path mating Parent 1 and 2 to generate a new child solution. Parent trajectories are shown to intersect at a cross-over

point depicted by the shaded cell. The last control action inherited from Parent 2 is deleted to maintain solution consistency.

operator mainly relies on the key notion of parent compatibility. In its simplest form, two parents are compatible if their paths cross each other at least once, while exhibiting dissimilar subroutes prior and posterior to the crossing point (avoiding parent duplication). However, local repair of the offspring path solution may be necessary to make it feasible.

4) Mutation: Mutation is a natural evolution process modifying some individual's genes more or less frequently. Two mutation operators are proposed, namely, M_path and M_path_local_repair. M_path first consists in selecting a specific move (index t>2) composing a path solution, modifying it randomly with an alternate action and then reconstructing a feasible remaining solution from that point, as shown in Figure 6. From the altered move (gene at index t) on, the last portion of the path (chromosome) is generated,

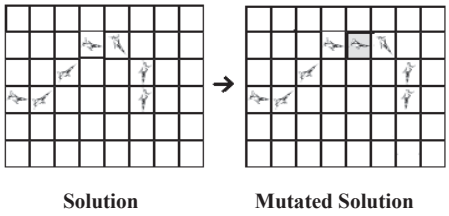

Figure 6. Mutation: a move from an individual solution is randomly selected and mutated. A new solution is then reconstructed from that point (shaded move).

by consecutively and randomly adding new moves (genes) until a complete solution emerges. Move probability selection is proportional or biased toward the best expected one-step information gain.

The M_path_local_repair mutation operator randomly removes a short path segment (few consecutive moves, at least two steps from the path endpoint) from a solution and builds an alternate path segment fragment to locally repair or bridge the disconnected components.

IV. COMPUTATIONAL EXPERIMENT

A. Methodology

A computational experiment has been conducted to illustrate the value of the proposed approach, showing the value of action control coordination and information-sharing. The cooperative path planning and information-sharing algorithms were first implemented in a decision support simulation prototype capturing multi-agent behaviors co-evolving in a time-constrained uncertain environment setting. Performance comparison over a variety of alternate techniques (e.g. tree search heuristic, self-interested agent) and key problem parameters (e.g. range, bandwidth) for a typical set of scenarios involving 3-10 agents on a 400-cell (20x20) grid has been carried out. A series of 30 simulation runs were conducted for each scenario and performance compared based on statistical analysis. Run-time 200-step simulations turned out to range between 1 and 4 minutes (Pentium II) for 30-solution agent populations. Typical scenario results are presented next.

B. Results

The value of action coordination of the proposed co-evolutionary approach over the baseline approximate dynamic programming (limited look-ahead) method reported in [10] for a 5-UAV team scenario is shown in Figure 7.

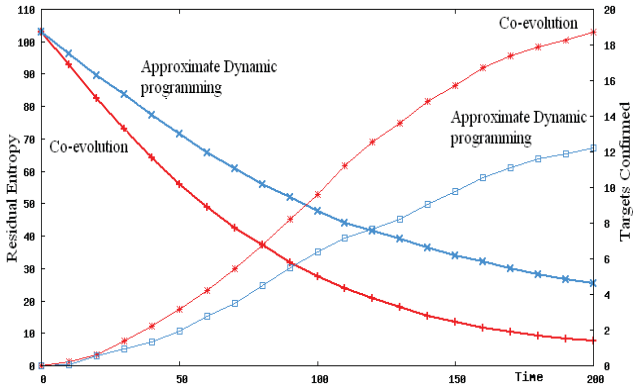

Figure 7. Entropy and confirmed targets over time: co-evolution Vs. approximate dynamic programming from [10].

In [10], cooperative path planning mainly ignore explicit communication to coordinate actions but rather relies on local environment sensor information and potential fields to avoid multi-agent interference in concurrently visiting the same cell. Performance gap is explained by explicit action coordination taking place between agents following information-sharing during a cycle. However, as shown in Figure 8 this gain comes at a higher communication cost, mainly due to continuous neighboring agent interaction in periodically exchanging revised intents which is overlooked in [10]. In contrast, the need to communicate observations decreases over time, as cells status increasingly gets confirmed. The longer the simulation, the heavier the communication burden associated to intents exchange. This could be reduced by restricting the exchange of intents to a maximum frequency and distance between agents, in proportion to the likelihood of interference with one another, while allowing communication of observations.

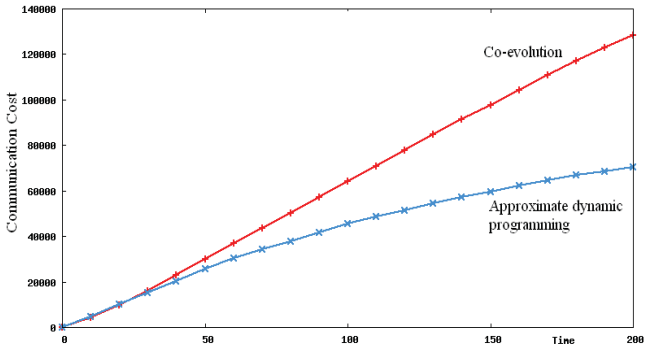

Figure 8. Communication cost over time: co-evolution Vs. approximate dynamic programming from [10].

The divergence-based information-sharing policy proposed in III A has been compared to the policy used in

[10] which relies on fixed periodic exchanges and full cognitive map exchange when meeting team members. Both policies have been implemented in the co-evolutionary framework and simulated. Depicting similar performance (no significant statistical difference) in terms of entropy and targets discovered, both policies differ on communication cost as shown in Figure 9 for a 5-UAV team scenario.

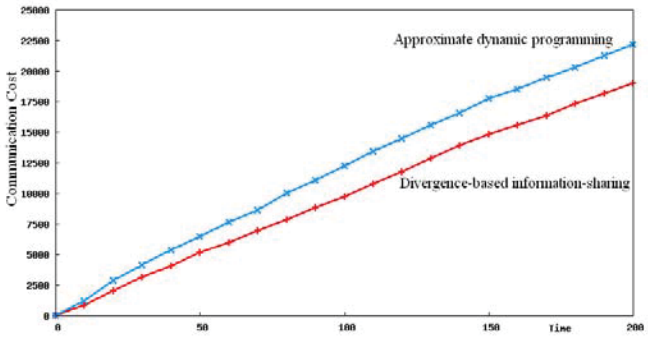

Figure 9. Communication cost over time: Divergence-based Vs. Periodic/opportunistic [10] information-sharing policies.

Both information-sharing policies are equally efficient in timely providing needed information, but exchanging observations based on belief divergence turns out to be significantly cheaper (~ 20%). The relative savings is even higher when considering non-intents (belief, observations) communication cost alone as emphasized in [10]. The difference lies on two reasons. First, agents exchange observations only within each other's field of interest. Even though this requires some overhead in the form of belief exchange, this overhead is counterbalanced by the drastically reduced number of observations communicated. Second, divergence-based information-sharing enables limited consensus through initial belief-sharing on mutually relevant regions. High-value observations can then be exchanged while better using available communication bandwidth. Alternatively, neighboring agents reported in [10] periodically synchronize or reach consensus either by sharing latest sensor readings recorded during the last period in the case of current neighbors, or blindly exchanging as much as possible the entire content of their cognitive map in the case of new neighbors meeting casually (opportunistic). As the policy proposed in [10] does not take advantage of potentially shared beliefs over cells or regions, it keeps consuming limited bandwidth on low utility observations and possibly multiple times.

V. Conclusion

A new information-theoretic co-evolutionary approach has been proposed to learn multi-agent resource-bounded coordination in the context of an extended multi-UAV search path planning problem explicitly taking into account available limited communication bandwidth and computational resources. It emphasizes co-evolution of team member path plans to achieve system-wide global objective. The technique allows distributed planning over rolling horizons, and exhibits adaptive and self-organizing team behavior. The anytime algorithm is combined to a belief divergence -based information-sharing policy to handle available limited communication bandwidth constraints.

Future directions will focus on new energy-efficient information-sharing strategies to minimize communication cost while maximizing team performance over multiple search-and-respond tasks with growing complexity.

References

[1] Benkoski, S., Monticino, M., and Weisinger, J., 1991. A survey of the search Theory Literature, *Nav. Res. Log*, 38.

[2] Stone, L. 1989. What's happened in search theory since the 1975 Lanchester Prize? *Operations Research*, 37 (3).

[3] Lau H. 2007. Optimal Search in Structured Environments, *PhD Thesis*, Univ. of Technology Sydney.

[4] How, J. et al., 2009. Increasing Autonomy of UAVs – Decentralized CSAT Mission Management Algorithm, *IEEE Robotics & Automation Magazine,* June 09, 43:51.

[5] Agmon, N. et al., 2008. The giving tree: constructing trees for efficient offline and online multi-robot coverage, *Ann Math Artif. Intell.* 52:143–168.

[6] Mathews, G., Durrant-Whyte H. 2006. Scalable Decentralised Control for Multi-Platform Reconnaissance and Information Gathering Tasks. In *9th International Conference on Information Fusion.*

[7] Jin, Y., Liao, Y., Minai, A., Polycarpou, M. 2006. Balancing search and target response in cooperative unmanned aerial vehicle (UAV) Teams *IEEE Trans on Sys Man and Cybern. Part B*, 36(3).

[8] Flint, F., E. Fernandez-Gaucherand, E., and M. Polycarpou, M., 2004. Efficient Bayesian Methods for Updating and Storing Uncertain Search Information for UAV's. *Proc 43rd IEEE Conf on Decision and Control*.

[9] Finn, A. et al., 2007. Design Challenges for an Autonomous Cooperative of UAVs, In *Proc. of the International Conf on Information, Decision and Control*.

[10] Liao, Y., Jin, Y., Minai, A. and Polycarpou, M. 2005. Information Sharing in Cooperative Unmanned Aerial Vehicle Teams, *Decision and Control.*

[11] Velagapudi, Prokopiev, Sycara, Scerri. Maintaining Shared Belief in a Large Mutiagent Team, Fusion 2007.

[12] Sycara, K. et al.2007. An Analysis and Design Methodology for Belief Sharing in Large Groups, *Fusion 2007*, Quebec.

[13] Intanagonwiwat, C. et al. 2003. Directed diffusion for wireless sensor networking. *IEEE/ACM Trans. on Networking*, 11(1):2–16.

[14] Xuan, P., Lesser, V., Zilberstein, S., 2001. Communication decisions in multi-agent cooperation: Model and experiments. In *Proc. of the 5th International Conference on Autonomous Agents*, Montreal, Canada.

[15] Brooks, A., Makarenko, A., Williams, S., Durrant-Whyte, H. 2006. Parametric POMDPs for Planning in Continuous State Spaces In *Robot. & Auton. Sys.*, 54(11).

[16] Bernstein, D. et al. 2002. The complexity of decentralized control of Markov decision processes. *Mathematics of Operations Research,* 27 (4), 819-840.

[17] Yang, Y., Minai, A., Polycarpou, M. 2005. Evidential Map Building Approaches for Multi-UAV Cooperative Search. *Proceedings of the American Control Conf.*

[18] Yang, Y., Minai, A., and Polycarpou, M. 2004. Decentralized Cooperative Search by Networked UAVs in an Uncertain Environment, *American Control Conf.*

[19] Yang, Y., Polycarpou, M., Minai, A. 2007. Multi-UAV Cooperative Search Using an Opportunistic Learning Method. *Journ. of Dyn. Sys., Meas., and Control,* 129(5).

[20] Cover, T. and Thomas, J., 2006. *Elements of Information-Theory*, 2nd edition, Wiley.

[21] Potvin, J.-Y. and Bengio, S., 1996. The Vehicle Routing Problem with Time Windows Part II: Genetic Search, *INFORMS Journal on Computing* 8, 165–172.

[22] Goldberg, D., 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning.* New York: Addison-Wesley.