SAND2019-5803C

Distributed Pursuit-Evasion Games for Mobile Monitoring and Surveillance

†Zvonimir Stojanovski, †Pingping Zhu, ‡Keith LeGrand, and †Silvia Ferrari
†Laboratory for Intelligent Systems and Controls (LISC),
Sibley School of Mechanical and Aerospace Engineering,
Cornell University, Ithaca (NY); ‡Sandia National Laboratories, Albuquerque (NM).

***Abstract*— This paper presents a novel pursuit-evasion problem for systems of agents tasked with pursuing moving targets while minimizing exposure to multiple observers. Given target position information from remote sensor data, each agent seeks to keep targets inside a bounded field-of-view known from the on-board sensor. Exposure to the network of observers is minimized by avoiding their sensor fields-of-view while pursuing the targets of interest. The optimality conditions for the agent trajectory optimization are derived by reducing the calculus of variations problem to an initial value problem to be solved for the agents' initial positions. The method is shown to be effective at controlling the agents in numerical simulations involving two observers programmed to cover the region of interest, and four targets with randomized motion patterns.**

I. Introduction

Many formulations of pursuit-evasion games have been proposed to date in the literature because of their applicability to a wide range of problems with competing objectives, such as rescue missions in contested environments [1], spacecraft rendezvous and missile defense [2], and mobile robotics [3]. A comprehensive survey of pursuit-evasion games and solution strategies for different types of agents, targets, and environments can be found in [4]. More recently, new methods for solving classic pursuit-evasion problems have also been developed, including potential-based trajectory optimization [5] and distributed optimal control (DOC) [6]. Other successful solution approaches include but are not limited to model predictive control (MPC) [7], reinforcement learning (RL) [8], and deep learning [9].

This paper presents a new formulation of pursuit-evasion problems that is inspired by sensing and surveillance applications such as space situational awareness (SSA) and fence-line monitoring. In the proposed pursuit-evasion problem, multiple mobile agents equipped with on-board sensors are tasked with monitoring multiple mobile targets using a bounded field-of-view, thus, also requiring the agents to pursue the targets so as to maintain visibility over time. In many modern sensing applications, agents must closely monitor moving targets in a contested environment, where exposure to multiple observers also is to be minimized. This paper presents a novel pursuit-evasion problem and solution based on recent techniques for sensor modeling [10]–[18] and calculus of variations. The paper shows that the optimal trajectories can be obtained from an initial value problem derived from the Euler-Lagrange equations, where the numerical solution can be cast as the optimization of the agents' initial positions. The problem formulation and approach are demonstrated via numerical simulations in which a network of four agents seeks to pursue four moving targets while avoiding detection by two mobile directional sensors.

This work was supported in part by the Office of Naval Research Code 321 and by the Laboratory Directed Research and Development program at Sandia National Laboratories, a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

II. Problem Formulation

Consider a rectangular region of interest (ROI) denoted by $\mathcal{W} \subset \mathbb{R}^2$ and populated by N agents, M targets, and P observers that the agents must avoid at all times. The agents, targets, and observers are all mobile with respect to $\mathcal{W}$. The agents and the observers are equipped with on-board sensors that allow them to obtain measurements of targets and agents, respectively, within a bounded field-of-view (FOV). The agents are each equipped with an omnidirectional sensor that is characterized by a circular FOV, $\mathcal{S}[\mathbf{x}_i(t)]$, centered at the agent position, $\mathbf{x}_i(t) \in \mathcal{W}$, and abbreviated by $\mathcal{S}_i(t)$. The observers are equipped with directional sensors, such as cameras, whose FOV can be modeled as a sector with origin at the observer position, $\mathbf{s}_k(t) \in \mathcal{W}$, and an aperture, $\beta \in [0, \pi/2)$, that is assumed constant and known *a priori*. The N agents to be controlled are tasked with observing moving targets at known time-varying positions $\boldsymbol{\xi}_j(t) \in \mathcal{W}$, and thus, must pursue the targets so as to maintain them inside the FOV, while avoiding the observers' FOVs. The N agents' trajectories, $\mathbf{x}_i(t) = [x_i(t)\ y_i(t)]^T$ $(i = 1, \ldots, N)$, are to be optimized in an attempt to provide persistent target observations during a fixed time interval $[t_0, t_f]$, while avoiding detection by the observers, and minimizing energy consumption. For simplicity, in this paper, it is assumed that the target and observer positions are known *a priori*, based on remote sensing data, and they are denoted by $\boldsymbol{\xi}_j(t)$ and $\mathbf{s}_k(t)$, respectively, where $j = 1, \ldots, M$ and $k = 1, \ldots, P$. Figure 1 shows a simple example of a workspace with targets, agents, observers, and corresponding FOVs.

It is assumed that each sensor is equipped with an omnidirectional sensor with range r. Then, the indicator function

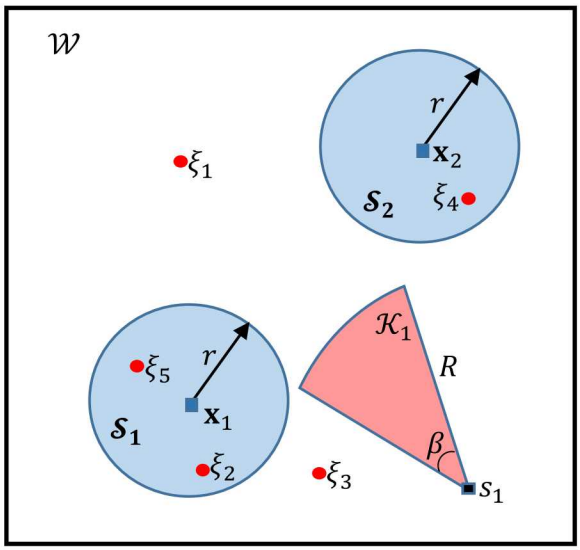

Fig. 1. Schematic of problem formulation for two agents (blue squares), five targets (red circles), one observer (black square), and corresponding FOV geometries.

can be used to model the FOV, as follows:

$$1_{\mathcal{S}_i}[\boldsymbol{\xi}_j(t)] \triangleq \begin{cases} 1, & \|\mathbf{x}_i(t) - \boldsymbol{\xi}_j(t)\|^2 \leq r^2 \\ 0, & \text{otherwise} \end{cases}, \ \forall i, j \tag{1}$$

This implies that the number of targets that are *not* covered by at least one agent is

$$m(t) = \sum_{j=1}^{M} \prod_{i=1}^{N} \{1 - 1_{\mathcal{S}_i}[\boldsymbol{\xi}_j(t)]\} \tag{2}$$

Similarly, the observer FOV is modeled as a sector of a circle centered at $\mathbf{s}_k(t)$, with constant and known radius R and central angle β. Because the orientation of the sector can change over time, e.g. as a result of pan and tilt parameters, the observer FOV is also a function of a unit vector, $\mathbf{b}_k(t) \in \mathbb{R}^2$, chosen to represent the direction of the bisecting ray. Then, the indicator function representing the observer FOV, denoted by $\mathcal{K}_k[\mathbf{s}_k(t), \mathbf{b}_k(t)]$ and abbreviated by $\mathcal{K}_k(t)$, is defined as,

$$1_{\mathcal{K}_k}[\mathbf{x}_i(t)] \triangleq \begin{cases} 1, & \|\mathbf{x}_i(t) - \mathbf{s}_k(t)\|^2 \leq R^2 \text{ and} \\ & \cos^{-1}\left(\frac{\mathbf{b}_k(t)\cdot[\mathbf{x}_i(t) - \mathbf{s}_k(t)]}{\|\mathbf{x}_i(t) - \mathbf{s}_k(t)\|}\right) \leq \beta/2 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

for all i and k, where $(\cdot)$ denotes the dot product. In other words, an agent i is in the FOV of observer k when the distance between the agent and the observer is less than or equal to the range R, and the angle between $\mathbf{b}_k(t)$ and the vector $[\mathbf{x}_i(t) - \mathbf{s}_k(t)]$ is less than half the sensor aperture. The number of observers able to detect agents at any time $t \in [t_0, t_f]$ is given by,

$$n(t) = \sum_{i=1}^{N} \sum_{k=1}^{P} 1_{\mathcal{K}_k}[\mathbf{x}_i(t)] \tag{4}$$

where each observer and agent pair are accounted for separately. Then, the time-averaged values of $m(t)$ and $n(t)$,

$$\bar{m} \triangleq \frac{1}{t_f - t_0} \int_{t_0}^{t_f} m(t)\ dt \tag{5}$$

and

$$\bar{n} \triangleq \frac{1}{t_f - t_0} \int_{t_0}^{t_f} n(t)\ dt, \tag{6}$$

respectively, can be optimized with respect to time.

The instantaneous energy required by agent i at time t can be assumed to be proportional to the square of its speed, i.e.,

$$e_i(t) = \kappa \|\dot{\mathbf{x}}_i(t)\|^2 \tag{7}$$

where $\|\cdot\|$ denotes the L_2 norm, and κ is a positive constant, as shown in [5]. Then, the total energy required by agent i is obtained by integrating (7) over time,

$$E_i = \kappa \int_{t_0}^{t_f} \|\dot{\mathbf{x}}_i(t)\|^2\ dt \tag{8}$$

and the total energy required by all agents,

$$E = \kappa \sum_{i=1}^{N} \int_{t_0}^{t_f} \|\dot{\mathbf{x}}_i(t)\|^2\ dt \tag{9}$$

can be minimized over time.

Finally, assuming a scalar weighted combination of multiple objectives can be used to represent the agents' performance, the objective function to be minimized is,

$$J = w_E E + w_m \bar{m} + w_n \bar{n} \tag{10}$$

where w_E, w_m, and w_n are positive weights chosen by the user or by a Pareto optimal trade off. Since each of the objective functions is a definite integral over the same range of time, the objective function can be written as

$$J = \int_{t_0}^{t_f} L[\mathbf{x}_i(t), \dot{\mathbf{x}}_i(t)]\ dt \tag{11}$$

where,

$$L = w_E \kappa \|\dot{\mathbf{x}}(t)\|^2 + \frac{1}{(t_f - t_0)}[w_m m(t) + w_n n(t)] \tag{12}$$

is the Lagrangian of the optimization problem. This paper presents an approach for optimizing the agents' trajectories over the time interval $[t_0, t_f]$, such that the agents are able to pursue and observe the targets with a bounded FOV, while simultaneously avoiding the set of observers.

III. Trajectory Optimization Solution

An approach based on calculus of variations [19] is proposed here for networks of agents of small to medium size. In future work, the approach will be extended to large collaborative networks by using distributed optimal control (DOC) [6]. Based on calculus of variations, when given an integral objective function in the form (11), the necessary conditions for optimality are provided by the Euler-Lagrange (EL) equations [20],

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}_i}\right) - \frac{\partial L}{\partial x_i} = 0 \tag{13}$$

and

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{y}_i}\right) - \frac{\partial L}{\partial y_i} = 0 \tag{14}$$

with boundary conditions,

$$\left.\frac{\partial L}{\partial \dot{x}_i}\right|_{t=t_0} = \left.\frac{\partial L}{\partial \dot{x}_i}\right|_{t=t_f} = 0 \tag{15}$$

and

$$\left.\frac{\partial L}{\partial \dot{y}_i}\right|_{t=t_0} = \left.\frac{\partial L}{\partial \dot{y}_i}\right|_{t=t_f} = 0 \tag{16}$$

Evaluating the derivatives of L and rewriting the equations in vector form provides extremals that satisfy the equation of motion,

$$\ddot{\mathbf{x}}_i = \frac{1}{2\kappa w_E(t_f - t_0)}\left[w_m \nabla_i(m) + w_n \nabla_i(n)\right] \tag{17}$$

where $\nabla_i(\cdot)$ denotes the gradient with respect to $\mathbf{x}_i$, or

$$\nabla_i(\cdot) \triangleq \left[\frac{\partial(\cdot)}{\partial x_i} \quad \frac{\partial(\cdot)}{\partial y_i}\right]^T \tag{18}$$

Applying (15)-(16) to the problem formulation in Section II, boundary conditions (BCs) for the EL equations are reduced to:

$$\dot{\mathbf{x}}_i(t_0) = \dot{\mathbf{x}}_i(t_f) = \mathbf{0} \tag{19}$$

Together with the above BCs, the EL equations amount to two-point boundary value problem (TPBV) that can be solved numerically only for small N. When the kinodynamic constraints of the agents can be neglected, the optimal agent trajectories can be found by converting the TPBV problem to an initial value problem (IVP) comprised of the equation of motion (17) and the initial conditions,

$$\mathbf{x}_i(t_0) = \mathbf{x}_i^* \tag{20}$$

and

$$\dot{\mathbf{x}}_i(t_0) = \mathbf{0} \tag{21}$$

where $\mathbf{x}_i^*$ is the unknown. The above IVP is much easier to solve and is known to have a unique solution [21]. Thus, the trajectory optimization problem presented in Section II is reduced to optimizing the initial positions of the agents, $\mathbf{x}_i^*$, for $i = 1, \ldots, N$.

A. Evaluating the Indicator Functions and their Gradients

The indicator functions used to model the sensor FOVs $\mathcal{S}_i(t)$ and $\mathcal{K}_k(t)$ can be expressed in terms of the Heaviside function, $H(\cdot)$, as follows,

$$1 - 1_{\mathcal{S}_i}[\boldsymbol{\xi}_j(t)] = H(\|\mathbf{x}_i(t) - \boldsymbol{\xi}_j(t)\|^2 - r^2) \tag{22}$$

and,

$$\begin{aligned} 1_{\mathcal{K}_k}[\mathbf{x}_i(t)] &= H(R^2 - \|\mathbf{x}_i(t) - \mathbf{s}_k(t)\|^2) \\ &\times H\{\mathbf{b}_k(t) \cdot [\mathbf{x}_i(t) - \mathbf{s}_k(t)] - \|\mathbf{x}_i(t) - \mathbf{s}_k(t)\| \cos(\beta/2)\} \end{aligned} \tag{23}$$

where

$$H(\zeta) \triangleq \begin{cases} 1, & \zeta \geq 0 \\ 0, & \zeta < 0 \end{cases} \tag{24}$$

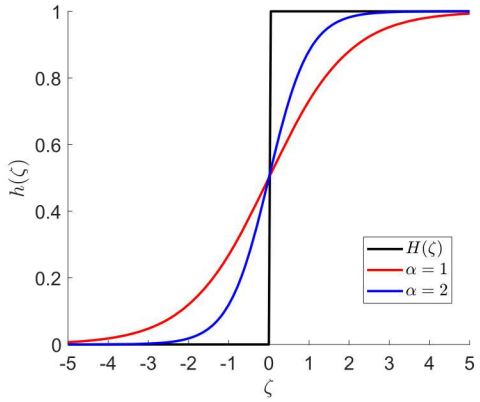

Fig. 2. Sigmoidal approximation of Heaviside function.

Substituting the above expressions into objective functions (2) and (4) allows one to re-write these objectives as,

$$m(t) = \sum_{j=1}^{M}\prod_{i=1}^{N} H(\|\mathbf{x}_i(t) - \boldsymbol{\xi}_j(t)\|^2 - r^2) \tag{25}$$

and

$$\begin{aligned} n(t) = &\sum_{k=1}^{P}\sum_{i=1}^{N} H(R^2 - \|\mathbf{x}_i(t) - \mathbf{s}_k(t)\|^2) \\ &\times H\{\mathbf{b}_k(t) \cdot [\mathbf{x}_i(t) - \mathbf{s}_k(t)] - \|\mathbf{x}_i(t) - \mathbf{s}_k(t)\| \cos(\beta)\} \end{aligned} \tag{26}$$

Because the Heaviside function $H(\zeta)$ is non-differentiable at $\zeta = 0$, an approach is proposed to obtain the gradients of m and n in (17) by implementing a smooth approximation of $H(\zeta)$ provided by the sigmoidal function,

$$h(\zeta) = \frac{1}{1 + \exp(-\alpha\zeta)} \tag{27}$$

where α is a positive constant chosen here by the user. The above function is infinitely differentiable and its first derivative is given by:

$$h'(\zeta) = \frac{\alpha}{2 + \exp(-\alpha\zeta) + \exp(\alpha\zeta)} \tag{28}$$

Figure 2 shows how the value of α can be adjusted to improve the approximation of the Heaviside function. Because the derivative of the sigmoidal function approaches infinity as the approximation improves, a suitable tradeoff must be chosen to represent a gradual deterioration in performance as a target or agent approaches the boundary of the FOV.

Substituting the sigmoidal function approximation (27) in objective functions (25) and (26), the following differentiable objectives,

$$m(t) = \sum_{j=1}^{M}\prod_{i=1}^{N} h(\|\mathbf{x}_i(t) - \boldsymbol{\xi}_j(t)\|^2 - r_j^2) \tag{29}$$

and

$$\begin{aligned} n(t) = &\sum_{k=1}^{P}\sum_{i=1}^{N} h(R^2 - \|\mathbf{x}_i(t) - \mathbf{s}_k(t)\|^2) \\ &\times h\{\mathbf{b}_k(t) \cdot [\mathbf{x}_i(t) - \mathbf{s}_k(t)] - \|\mathbf{x}_i(t) - \mathbf{s}_k(t)\| \cos(\beta)\} \end{aligned} \tag{30}$$

are obtained. This makes it possible to compute the gradients of the target-pursuit objective with respect to $\mathbf{x}_l$,

$$\begin{aligned}\nabla_l(m) =& 2\sum_{j=1}^{M}(\mathbf{x}_l - \boldsymbol{\xi}_j)\frac{h'(\|\mathbf{x}_l - \boldsymbol{\xi}_j\|^2 - r^2)}{h(\|\mathbf{x}_l - \boldsymbol{\xi}_j\|^2 - r^2)} \\ &\times \prod_{i=1}^{N} h(\|\mathbf{x}_i - \boldsymbol{\xi}_j\|^2 - r^2)\end{aligned} \tag{31}$$

and the gradient of the observer-avoidance objective with respect to $\mathbf{x}_l$,

$$\nabla_l(n) = \sum_{k=1}^{P} \nabla_l[1_{\mathcal{K}_k}(\mathbf{x}_l)] \tag{32}$$

where,

$$\begin{aligned}&\nabla_l[1_{\mathcal{K}_k}(\mathbf{x}_l)] = -2[\mathbf{x}_l - \mathbf{s}_k]h'(R^2 - \|\mathbf{x}_l - \mathbf{s}_k\|^2) \\ &\times h\{\mathbf{b}_k \cdot [\mathbf{x}_l - \mathbf{s}_k] - \|\mathbf{x}_l - \mathbf{s}_k\|\cos(\beta)\} \\ &+ 2\left[\mathbf{b}_k - \frac{\mathbf{x}_l - \mathbf{s}_k}{\|\mathbf{x}_l - \mathbf{s}_k\|}\cos(\beta)\right] h(R^2 - \|\mathbf{x}_l - \mathbf{s}_k\|^2) \\ &\times h'\{\mathbf{b}_k \cdot (\mathbf{x}_l - \mathbf{s}_k) - \|\mathbf{x}_l - \mathbf{s}_k\|\cos(\beta)\}\end{aligned}$$

and time arguments are omitted for brevity.

IV. Simulation and Results

The methodology presented in the previous section is demonstrated via numerical simulations involving multiple agents, targets, and observers. The workspace is given by $\mathcal{W} = [-L/2,\ L/2] \times [-L/2,\ L/2]$, where $L = 4$ Km. The target trajectories are parameterized by quadratic functions of time,

$$\boldsymbol{\xi}_j(t) = \mathbf{a}_j t^2 + \mathbf{b}_j t + \mathbf{c}_j, \quad j = 1, \ldots, 4 \tag{33}$$

where $\mathbf{a}_j$, $\mathbf{b}_j$, and $\mathbf{c}_j$ are (2×1) constant vectors that are randomly generated for every value of j. The trajectories of the observers are chosen to be along circles with a constant radius S, centered at the origin of an inertial reference frame $\mathcal{F}_{\mathcal{W}}$ embedded in $\mathcal{W}$ (Fig. 3). These motion patterns are motivated by monitoring applications in which the observers must periodically cover a pre-defined ROI by means of a bounded FOV. For $P = 2$ observers that move at a constant speed, the trajectories are chosen to perform one revolution per hour, such that the observers are located on opposite sides of the circle at any given time (Fig. 3).

Based on the above description, the observers' trajectories can be obtained from the equations of a circle,

$$\mathbf{s}_k(t) = \begin{bmatrix} S\cos(\Omega t + k\pi) \\ S\sin(\Omega t + k\pi) \end{bmatrix}, \quad k = 1, 2 \tag{34}$$

where Ω is a constant angular rate chosen by the user. The observers' boresights spin at a constant rate and are in-phase and, therefore, the FOV orientation is given by,

$$\mathbf{b}_k(t) = \begin{bmatrix} \cos(\omega t) \\ \sin(\omega t) \end{bmatrix}, \quad k = 1, 2 \tag{35}$$

where ω is a constant angular rate chosen by the user. The angular rates and other constant parameters used in the simulations are listed in Table I. The objective function weights w_m and w_n are chosen to be 100 times greater than w_E in order to prioritize target tracking and observer avoidance over energy consumption.

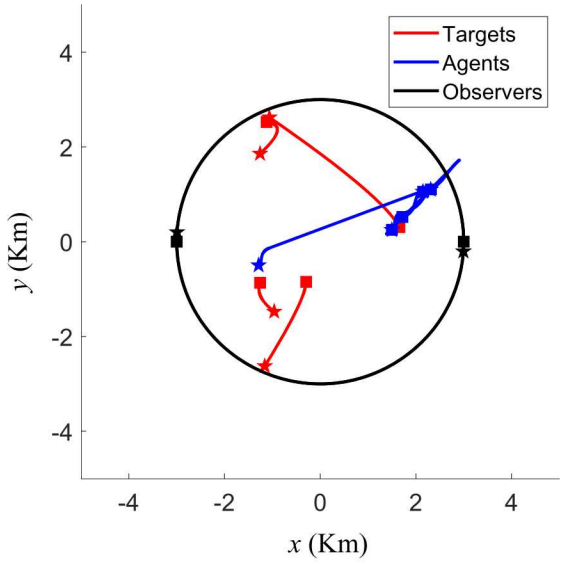

Fig. 3. Optimal agent trajectories (blue) given the known motion patterns of four targets (red) and two observers (black), with initial and final positions labeled by squares and stars, respectively.

TABLE I
VALUES OF SCALAR PARAMETERS USED IN NUMERICAL SIMULATIONS.

Parameter	Symbol	Value	Units
Number of targets	M	4	—
Number of agents	N	4	—
Number of observers	P	2	—
Observer FOV radius	R	3	km
Agent FOV radius	r	1	km
Observer path radius	S	3	km
Workspace dimension	L	4	km
Observer FOV aperture	β	$\pi/6$	rad
Start time	t_0	0	h
End time	t_f	1	h
Power coefficient	κ	1	W/(km/h)^2
Energy consumption weight	w_E	1	J^{-1}
Missed target weight	w_m	100	—
Agent detection weight	w_n	100	—
Observer path angular speed	Ω	2π	rad/h
Observer boresight angular speed	ω	24π	rad/h
Sigmoidal function parameter	α	1	—

The optimal agent trajectories are plotted by blue lines in Figure 3, along with the paths of the targets and observers simulated based on the laws in (33)-(35), and the parameters in Table I. Figure 4 shows the resulting numbers of missed targets and agent detections over time. The time history of the total energy consumption is plotted in Figure 5.

The results in Figure 4 show that the agents are fairly successful in pursuing the targets and avoiding observers. There are times when one or two targets escape the agents by exiting their FOVs, but this happens only sporadically and, then, only for a short time interval, on the order of few minutes. Also, agents are detected by observers only twice toward the end of the time interval. Furthermore,

TABLE II

VALUES OF MATRIX PARAMETERS USED IN TARGET SIMULATIONS.

Parameter	Value	Units
$\mathbf{a}$	$\begin{bmatrix} -0.831 & -1.156 & -2.003 & 0.520 \\ -0.979 & -0.534 & 0.964 & -0.020 \end{bmatrix}$	km/h^2
$\mathbf{b}$	$\begin{bmatrix} -0.035 & 1.019 & -0.715 & -0.225 \\ -0.798 & -0.133 & 1.351 & -0.589 \end{bmatrix}$	km/h
$\mathbf{c}$	$\begin{bmatrix} -0.294 & -1.120 & 1.655 & -1.257 \\ -0.848 & 2.526 & 0.308 & -0.865 \end{bmatrix}$	km

Notation: $\mathbf{a} = [\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_M]$, etc.

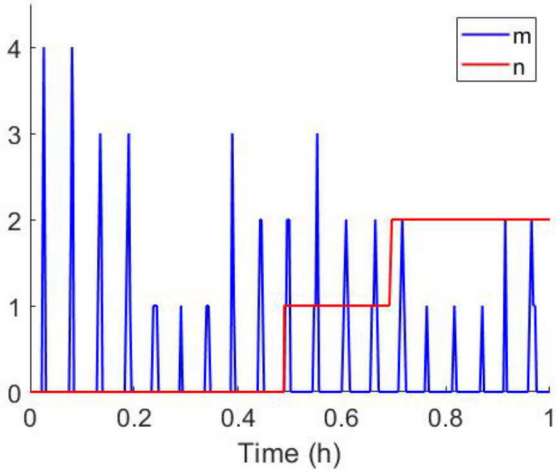

Fig. 4. Number of missed targets (m) and agent detections (n) over time resulting from the optimal agent trajectories in Fig. 3.

while the minimization of energy consumption has lower priority relative to the pursuit-evasion objectives, the agent trajectories are very efficient and require a remarkably low amount of energy (Figure 5).

V. CONCLUSION

This paper presents an approach for optimizing the trajectories of multiple distributed agents tasked with pursuing multiple moving targets, while avoiding being detected by multiple moving observers. In this problem, the motion of targets and observers is assumed known from remote sensor data, while the controllable agents are in charge of obtaining *in-situ* high-quality measurements that can allow, for example, to monitor, model, and classify the targets of interest. The approach could be modified and this assumption relaxed by including objectives related to state estimation and prediction.

The approach presented in this paper shows that by neglecting kinodynamic constraints, the optimal trajectories can be obtained by reducing the Euler-Lagrange equations to an initial value problem in which the variables correspond to the initial positions of the agents. The simulation results show that this approach can optimize both pursuit and evasion objectives, while requiring a minimal amount of energy expenditure. Future work will investigate the use of other objective functions and the presence of kinodynamic constraints in order to improve the pursuit-evasion performance as well as guarantee reachability of the optimal trajectories.

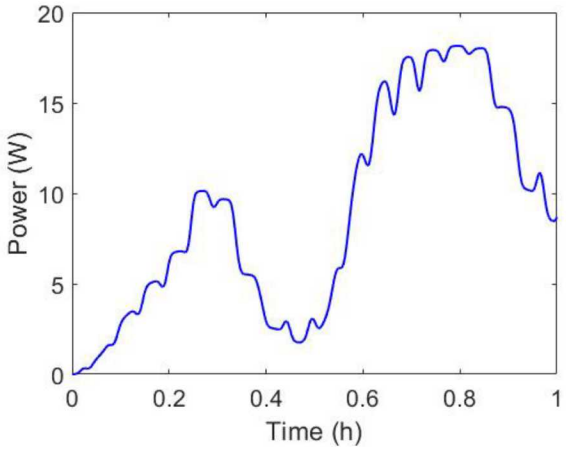

Fig. 5. Total energy consumption required by the agents when following the optimal trajectories in Fig. 3.

REFERENCES

[1] D. W. Oyler, P. T. Kabamba, and A. R. Girard, "Pursuitevasion games in the presence of obstacles," *Automatica*, vol. 65, pp. 1–11, 2016.

[2] K. D. Aures-Cavalieri, "Incomplete information pursuit-evasion games with applications to spacecraft rendezvous and missile defense," Ph.D. dissertation, 2014.

[3] N. Noori, A. Beveridge, and V. Isler, "Pursuit-evasion: A tool kit to make applications more accessible," *IEEE Robotics and Automation Magazine*, vol. 23, no. 4, pp. 138–149, 2016.

[4] T. H. Chung, G. A. Hollinger, and V. Isler, "Search and pursuit-evasion in mobile robotics: A survey," *Autonomous Robots*, vol. 31, no. 4, pp. 299–316, 2011.

[5] S. Ferrari and G. Foderaro, "A Potential Field Approach to Finding Minimum-Exposure Paths in Wireless Sensor Networks," *IEEE International Conference on Robotics and Automation*, 2010.

[6] G. Foderaro, S. Ferrari, and T. A. Wettergren, "Distributed optimal control for multi-agent trajectory optimization," *Automatica*, vol. 50, no. 1, pp. 149–154, 2014.

[7] D. De Simone, N. Scianca, P. Ferrari, L. Lanari, and G. Oriolo, "Mpc-based humanoid pursuit-evasion in the presence of obstacles," vol. 2017-, 2017, pp. 5245–5250.

[8] I. J. Sledge, M. S. Emigh, and J. C. Principe, "Guided policy exploration for markov decision processes using an uncertainty-based value-of-information criterion," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2080–2098, 2018.

[9] M. Httenrauch, A. oic, and G. Neumann, "Deep reinforcement learning for swarm systems," *JOURNAL OF MACHINE LEARNING RESEARCH*, vol. 20, 2019.

[10] C. Cai and S. Ferrari, "On the development of an intelligent computer player for CLUE®: a case study on preposterior decision analysis," in *American Control Conference*, Minneapolis, MN, 2006, pp. 4350–4355.

[11] ——, "Information-driven sensor path planning by approximate cell decomposition," *IEEE Transactions on Systems, Man, and Cybernetics - Part B*, vol. 39, no. 3, pp. 672–689, 2009.

[12] ——, "A *Q*-learning approach to developing an automated neural computer player for the board game of CLUE®," in *International Joint Conference on Neural Networks*, Hong Kong, 2008.

[13] ——, "Comparison of information-theoretic objective functions for decision support in sensor systems," in *Proc. American Control Conference*, New York, NY, 2007, pp. 63–133.

[14] H. Wei, W. Lu, P. Zhu, S. Ferrari, R. H. Klein, S. Omidshafiei, and J. P. How, "Camera control for learning nonlinear target dynamics via bayesian nonparametric dirichlet-process gaussian-process (dp-gp) models," in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*. IEEE, 2014, pp. 95–102.

[15] H. Wei, P. Zhu, M. Liu, J. P. How, and S. Ferrari, "Automatic pan–tilt camera control for learning dirichlet process gaussian process (dpgp) mixture models of multiple moving targets," *IEEE Transactions on Automatic Control*, vol. 64, no. 1, pp. 159–173, 2019.

[16] H. Wei, W. Lu, P. Zhu, S. Ferrari, M. Liu, R. H. Klein, S. Omidshafiei, and J. P. How, "Information value in nonparametric dirichlet-process gaussian-process (dpgp) mixture models," *Automatica*, vol. 74, pp. 360–368, 2016.

[17] H. Wei, W. Lu, P. Zhu, G. Huang, J. Leonard, and S. Ferrari, "Optimized visibility motion planning for target tracking and localization," in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, 2014, pp. 76–82.

[18] T. A. Wettergren, "Statistical analysis of detection performance for large distributed sensor systems," Naval Undersea Warfare Center, Newport, RI, Tech. Rep. ADA417136, June 2003. [Online]. Available: http://stinet.dtic.mil

[19] C. Fox, *An introduction to the calculus of variations*. Courier Corporation, 1987.

[20] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*. Dover, 2017.

[21] D. G. Schaeffer and J. W. Cain, *Ordinary Differential Equations: Basics and Beyond*. Springer, 2016.