Distributed Constrained Optimization for Bayesian Opportunistic Visual Sensing

Akshay A. Morye, Chong Ding, Amit K. Roy-Chowdhury, Jay A. Farrell

***Abstract*—We propose a control mechanism to obtain opportunistic high resolution facial imagery, via distributed constrained optimization of the PTZ parameters for each camera in a sensor network. The objective function of the optimization problem quantifies the per camera per target image quality. The tracking constraints, which are a lower bound on the information about the estimated position for each target, define the feasible PTZ parameter space. Each camera alters its own PTZ settings. All cameras use information broadcast by neighboring cameras such that the PTZ parameters of all cameras are simultaneously optimized relative to the global objective. At certain times of opportunity, due to the configuration of the targets relative to the cameras, and the fact that each camera may track many targets, the camera network may be able to reconfigure itself to achieve the tracking specification for all targets with remaining degrees-of-freedom that can be used to obtain high-res facial images from desirable aspect angles for certain targets. The challenge is to define algorithms to automatically find these time instants, the appropriate imaging camera, and the appropriate parameter settings for all cameras to capitalize on these opportunities. The solution proposed herein involves a *Bayesian* formulation in a game theoretic setting. The Bayesian formulation automatically trades off objective maximization versus the risk of losing track of any target. The article describes the problem and solution formulations, design of *aligned* local and global objective functions and the inequality constraint set, and development of a Distributed Lagrangian Consensus algorithm that allows cameras to exchange information and asymptotically converge on a pair of primal-dual optimal solutions. This article presents the theoretical solution along with simulation results.**

***Index Terms*—Camera Sensor Networks, Cooperative Control, Distributed Constrained Optimization.**

I. Introduction

Static camera networks have a lower per camera cost of installation than pan, tilt, zoom (PTZ) camera sensor networks; however, PTZ camera networks can have lower total installation cost with greater performance. Static camera networks must be designed to achieve the coverage and tracking specifications given worst case target distributions. Installation cost constraints can lead to imagery sequences from static camera network applications that are quite challenging to analyze. PTZ camera networks, with appropriate software and communications, can dynamically reconfigure in response to application events and actual target distributions to optimize the acquired imagery sequences accounting for viewpoints and resolution, to facilitate image analysis and scene understanding.

A prototypical application is a security screening checkpoint at the entrance lobby of a building. Over the course of each day a high volume of people flow through the room. The room is equipped with a fixed number of cameras while the number and location of people in the room varies with time. The objective of the camera network is to track (i.e., state estimation) all persons in the room at a specified accuracy level at all times and to capture high-res facial images for certain persons in the room at opportunistically selected instants.

The challenges of such an application are development of algorithms to ensure accurate propagation of target-related information throughout the distributed network, analysis of the effect of changing network topology on solution convergence, design of distributed PTZ optimization algorithms, and the design of objective functions suitable to solving the specified problem that also have the properties necessary to ensure convergence. All these factors influence the selection of an optimization strategy. In this paper, we formulate the problem within a *Bayesian* game theoretic framework and utilize a distributed constrained optimization approach to compute the optimal PTZ settings that achieve the global camera network objective through optimization of local camera objectives.

II. Literature Review

The research proposed herein falls within the scope of active computer vision [1], which involves research on cooperation and coordination between many cameras in a network, for applications such as autonomous surveillance, simultaneous localization and mapping (SLAM), trajectory planning, etc.

Assuming predesigned static camera placement given certain tasks, the articles [2], [3] define deployment strategies for camera networks. A path-planning based approach is proposed in [4], where static cameras track targets, and PTZ cameras obtain high-res images. Given a target activity map, the Expectation-Maximization algorithm [5] is used to perform area coverage in [6]. Other interesting active vision problems such as object detection across cameras, camera handoff, and camera configuration, with an emphasis on camera networks is addressed in [7]–[10]. Methods for tracking a group of people in a multi-camera setup are addressed in [11], [12]. A review of human motion analysis is performed in [13]. The methods proposed therein dealt with a centralized processing scheme and did not delve into the decentralized organization of cameras. Distributed computer vision algorithms are studied in [14], [15]. Distributed state estimation algorithms over vision networks are studied in [16]–[18]. The above articles do not focus on distributed optimization of the PTZ parameters to optimize the acquired imagery.

Assumptions on the camera network communication topology play a major role in the problem solution. Preliminary work on camera networks using independent cameras that lack coordination is provided in [19]. A machine learning-based method to learn the network topology is described in [20].

A. A. Morye, A. K. Roy-Chowdhury and J. A. Farrell are with the Dept. of EE, U. of CA Riverside, CA, 92521 USA. e-mail: amorye@ee.ucr.edu

C. Ding is with the Dept. of CSE, U. of CA Riverside, CA, 92521 USA

The authors employ unsupervised machine learning to establish links between target activity and the associated camera views to decipher communication topology to determine target tracks. In [21], authors measure the statistical dependence between observations in multiple camera views to quantify a potential interconnecting pathway for target tracks between different camera views with an aim to enable camera handoff. Multi-agent systems with switching topologies are studied in [22], [23]. Though the studies therein are not based on visual sensing applications, [22], [23] provide significant insight on methods potentially applicable to mobile camera networks.

A recent research survey [14] identifies various computer vision problems that can be solved in a distributed fashion within a network topology. Within a distributed framework, problems are often defined as multi-agent tasks that utilize cooperation between agents. A vision-based target pose estimation problem that employs cooperation and uses multi-agent optimization techniques is addressed in [24]. Therein, authors propose a passivity-based target motion observer model, and propose a cooperative estimation algorithm where every agent solves an unconstrained convex minimization problem to minimize estimation error. The passivity based motion observer model is utilized in [25] for camera control, under assumptions on target motion. Though the articles above focus on target and camera pose estimation, the challenges related to obtaining high-res imagery while maintaining state estimation performance are not considered.

A survey of automated visual surveillance systems is provided in [26]. Building scene interpretation in a modular manner on decentralized intelligent multi-camera surveillance systems is described in [27]. For multi-agent network systems that employ a time-varying topology, stability analysis is provided in [28]. The analysis is based on graph-theoretic tools where the authors assume the problem to be convex.

Game-theory as a tool for designing solutions to multi-agent problems is described in [29]. A vehicle-target assignment problem within the game-theoretic framework was proposed in [30]. The standard target assignment problem is different from a camera-target assignment problem in that the problem described therein is a one-to-one mapping problem with the stationary targets, whereas the camera-target assignment problem allows multiple cameras to each track multiple moving targets. Nonetheless, articles [29], [30] provide valuable insight into the challenges faced while designing the camera parameter optimization as a cooperative game played by multiple cameras.

A game-theoretic camera control approach for collaborative sensing is proposed in [31]–[34]. In [31], the agents collaborate to optimize a cost function that is the weighted combination of area coverage over regions of interest while trying to achieve high-res images of specific (highly weighted) targets. In [32]–[34], the agents account for risk and include image quality in a weighted cost function. Collaboration was ensured through a game-theoretic formulation. The quality of the solution was dependent on the user-defined weights.

An automated annealing approach for updating Lagrange multipliers within a constrained optimization framework to reduce dependence on agent inter-communication is provided in [35]. The method uses the probability collectives framework to generate a relation between game theory and statistical physics. The authors use a game-theoretic motivation to develop a parallel algorithm, but consider a non-cooperative game between agents, where the action of one agent is completely independent of the other agents in the network. Such an assumption is not appropriate for our application.

A systematic methodology for designing agent objective functions is outlined in [36], using a hierarchical decoupling of the global objective function into local objective functions that are aligned with the global function. Each agent is modeled as a self-interested decision maker within a game-theoretic environment, then convergence proofs from the game theory literature are utilized. Herein, we utilize the methodology of [36] to decompose a Bayesian value function designed for the opportunistic visual sensing application into local value functions suitable for distributed implementation.

III. Contributions of the Paper

The distributed PTZ camera parameter optimization problem considered herein is similar to the class of problems considered in [32], [33]. The solution methodology herein is significantly enhanced. The cost function in [33] is a weighted summation of competing objectives and the optimization is unconstrained. In [32], the authors use a heuristic to weight the high-res imaging objective dependent on the quality of target tracking performance achieved by the network of cameras, collectively. Although such an approach allows the network of cameras to obtain high-res images only if the tracking heuristic is desirable, there is no guarantee of all targets being tracked to the required tracking specification. Optimization is accomplished via *sequential* utility maximization by each agent. Neither paper includes an analysis of convergence of the distributed implementation. These issues are further discussed in Remark 3 after defining the required notation.

Herein, a method for distributed, cooperative and parallel optimization on a connected camera network is defined, analyzed, and implemented. The camera parameter optimization process maximizes a Bayesian value function that accounts for risk arising from the uncertainty in the estimated target positions, while adhering to Bayesian constraints on the tracking performance. The value function is designed as an *ordinal potential function* [36], such that it can be decoupled into local objectives known to each camera. The tracking constraint is common to all cameras. With reasonable assumptions on network connectivity, utilizing a Bayesian constrained optimization approach, the proposed solution method provides feasible optimal solutions to perform opportunistic visual sensing of targets maneuvering with random trajectories.

A shorter and less comprehensive presentation of this approach is contained in [34]. Relative to [34], this article includes an enhanced discussion and analysis of the proposed approach throughout, a detailed and realistic simulation example with discussion, a statistical analysis of the proposed approach relative to imaging with a network of static PTZ cameras, development of mechanisms for convergence towards the global optima, and a detailed discussion of possible alternative methodologies.

TABLE I: Notation Summary

Parameter	Variable
Pan, Tilt, Zoom	(ρ, τ, ζ)
Min. focal length, Max. focal length	$\underline{F}$, $\overline{F}$
No. of Cameras, No. of Targets in region	N_C, $N_T(t)$
i-th camera, j-th target	C_i, T^j
(ρ, τ, ζ) settings for C_i, all cameras except C_i	$\mathbf{a}_i$, $\mathbf{a}_{-i}$
(ρ, τ, ζ) settings for all cameras	$\mathbf{a}$
Dimension of (ρ, τ, ζ) settings for C_i	$\mathbf{a}_i \in \Re^{n_i}$, $n_i = 3$
Dimension of (ρ, τ, ζ) settings for all cameras	$\mathbf{a} \in \Re^n$, $n = 3N_C$
No. of problem constraints	m
Tracking performance vector for all targets	$\mathbf{U}_T(\mathbf{a})$
Tracking threshold vector for all targets	$\overline{\mathbf{T}}$
Global Bayesian imaging value over all targets	$V_I(\mathbf{a})$
Local Bayesian imaging value over all targets	$V_{I_i}(\mathbf{a}_i)$
Bayesian tracking value vector for all targets	$\mathbf{V}_T(\mathbf{a})$
Achieved global imaging value over all targets	$\bar{V}_I(\mathbf{a})$
Achieved local imaging value over all targets	$\bar{V}_{I_i}(\mathbf{a}_i)$
Achieved tracking value vector for all targets	$\bar{\mathbf{V}}_T(\mathbf{a})$
Lagrange multiplier vector for all targets	$\boldsymbol{\lambda}$
Camera C_i's version of $\boldsymbol{\lambda}$	$\boldsymbol{\lambda}_{(i)}$
Lagrange multiplier vector obtained via consensus	$\bar{\boldsymbol{\lambda}}$
Lagrangian constructed for optimization	$L(\boldsymbol{\lambda}, \mathbf{a})$
Weight for importance of imagery of T^j by C_i	w_i^j
Image resolution obtained for T^j by C_i	$r_i^j(\mathbf{a}_i)$
Relative pose quality factor between C_i and T^j	$\alpha_i^j(\mathbf{a}_i)$
State vector for T^j	$\mathbf{x}^j = [\mathbf{p}^j, \mathbf{v}^j]^\top$
State est., state est. covariance for T^j	$\hat{\mathbf{x}}^j$, $\mathbf{P}^j$
Fisher Information Matrix	$\mathbf{J}$
Measurement Vector, Measurement Covariance	$\mathbf{u}$, $\mathbf{C}$
Rotation Matrix from frame a to frame b	${}^b_a\mathbf{R}$
Entity e before, after new measurement	e^-, e^+
Entity e in global frame, frame defined by C_i	${}^g e$, ${}^i e$
Entity e at time-step t_k	$e(k)$
Entity e for target T^j	e^j
Entity e at local or global optimum	e^*

IV. Problem Description and Solution Overview

Facial detection and recognition processes are greatly facilitated by capture of high resolution images of the face from a desired aspect angle [37], [38]. This paper does not discuss facial recognition methods; instead, it focuses on the means to configure a network of PTZ cameras to opportunistically acquire high-res facial imagery. We use the term 'opportunistic' as each camera must select its parameters to satisfy a tracking constraint at all times and to obtain high-res facial images at times-of-opportunity. Such an opportunity may arise due to the high probability of image capture at a high zoom setting from a superior aspect angle, and when tracking constraints on all the targets can be simultaneously satisfied. The tracking constraints, while useful in their own right, are necessary to enable high-res imaging. Note also that, due to the uncertainty of the target motion, the high zoom setting required by certain cameras to attempt to obtain high-res images, will impose tracking risk that must be accounted for appropriately by the camera network.

The operating environment includes N_C cameras placed at known, fixed locations and a time-varying number of targets $N_T(t)$ with independent and unknown trajectories. It is possible that $N_T(t) > N_C$. All cameras have changeable pan ($\rho \in [-180°, 180°]$), tilt ($\tau \in [-90°, 90°]$), and zoom ($\zeta \in [\underline{F}, \overline{F}]$) parameters. We assume the cameras to have parfocal zoom lenses [39] that maintain focus with changing focal length and have a negligible focus error.

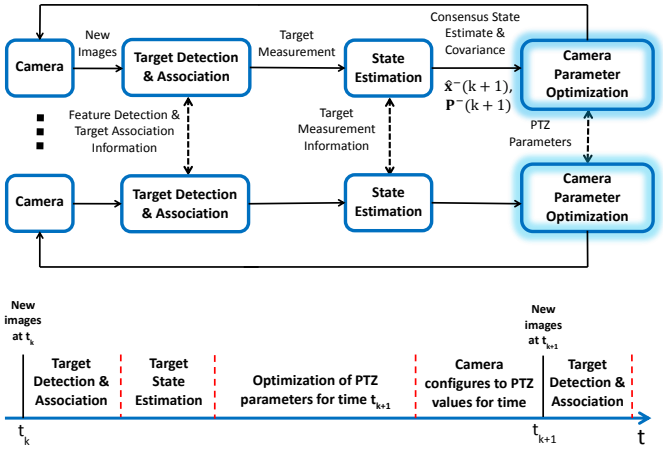


Fig. 1: System Block Diagram and Event Time-line: Note that information exchange shown is only between neighboring cameras. The time-line of procedural events is shown between image sample times.

In a distributed solution framework, using information up to and including the last imaging time t_k, the i-th camera in the network will be required to optimize its parameters $\mathbf{a}_i = [\rho_i, \tau_i, \zeta_i]^\top \in \Re^3$ for the next imaging instant t_{k+1}, in cooperation with the other cameras to maximize an objective function. Any choice of $\mathbf{a}_i$ yields a field-of-view (FoV_i) for the resulting image. The parameters of all cameras are organized into a vector $\mathbf{a} = [\mathbf{a}_1, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_{N_c}]^\top$. The vector containing all parameter vectors except those of C_i is denoted by $\mathbf{a}_{-i}$. Additional notation is summarized in Table I.

Remark 1: By the nature of the visual sensing application, depending on the choice of $\mathbf{a}_i$, Camera C_i may image multiple targets each at a different resolution. The attained resolution, one quantity affecting image quality, is determined by the choice of $\mathbf{a}_i$. In addition, depending on the choice of $\mathbf{a}$ any given target may be imaged by multiple cameras, yielding improved tracking accuracy. $\diamond$

Solution of the overall problem in the time interval $t \in (t_k, t_{k+1})$ involves several processes. Based on the previous images up to and including the image at t_k, the target state estimation process provides a prediction of the mean $\hat{\mathbf{x}}^j(k+1)^-$ and covariance matrix $\mathbf{P}^j(k+1)^-$ for all targets (i.e., $j = 1, \ldots, N_T$) at t_{k+1}. Every camera in the network has its own embedded target detection module [31], [32], [40], [41], a target state tracker, and a distributed camera parameter optimizer, as shown in Fig. 1. This article focuses on distributed camera parameter optimization, not target detection or state estimation. For the results herein, the target state estimate is computed using the Information Weighted *Kalman*-Consensus tracker derived in [18].

In Fig. 1, the top portion illustrates the information flow and the bottom portion illustrates the processing sequence. Images acquired at t_k are first processed for feature detection and target association. The resulting measurements are used for distributed state estimation [18], [42], which ensures that the state of each target is estimated consistently by each camera in the network. Each camera only uses its own imagery and

data communicated from its neighbors' on the communication graph. Consistency and accuracy of state estimation are pre-requisites that enable distributed optimization of the network parameter vector $\mathbf{a}$ for high-res image acquisition at t_{k+1}. Upon completion of the target state estimation process, a prior target state estimate ${}^g\hat{\mathbf{x}}^j(k+1)^-$ and a prior covariance matrix $\mathbf{P}^j(k+1)^-$ are available for each target T^j at the future sampling time t_{k+1}. Subsequently, each camera will optimize its PTZ parameter settings $\mathbf{a}_i$ using ${}^g\hat{\mathbf{x}}^j(k+1)^-$ and $\mathbf{P}^j(k+1)^-$ as inputs. Following computation of an optimal PTZ setting, each camera C_i physically changes its settings to the specified value $\mathbf{a}_i(t_{k+1})$. The focus of this paper is on the algorithms within the optimization block as highlighted in Fig. 1.

Remark 2: If cameras take images with a period T_s, the designer could choose to re-optimize the camera parameter settings $\mathbf{a}$ after every M-th image, resulting in $t_k = MT_s$, while still using all M images for target detection and tracking. In this article, we choose to use $M = 1$. In future work, it could be interesting to adapt M in response to events.$\diamond$

Remark 3: It is useful to compare different optimization approaches. In *centralized optimization*, one entity would receive all the required information and adjust the entire vector $\mathbf{a}$ to maximize the expected value function. Convergence of centralized optimization methods are well understood. In distributed approaches, each agent C_i will only adjust the proposed values of its parameters $\mathbf{a}_i$ for the next imaging time. In *distributed sequential optimization*, while C_i is adjusting $\mathbf{a}_i$, all other cameras C_j for $j \neq i$ are idle. Cameras sitting idle potentially save energy at the expense of time to reach convergence. The convergence of such schemes is straightforward to analyze as each camera is solving a much lower dimension optimization problem. The analysis would be similar to that for the centralized case. In *distributed parallel optimization*, all cameras adjust their parameters simultaneously. Convergence of this case is more complex, requiring results from optimization and game theory, and cost functions meeting certain technical requirements.

V. Background

This section briefly reviews concepts of optimization [43] and game theory [36], [44] necessary for the solution methodology proposed herein.

A. Centralized Constrained Optimization

Consider a standard convex vector optimization (e.g., maximization) problem with a differentiable primal objective function[1] f_o and differentiable inequality constraints g^j

$$\begin{aligned} &\text{maximize} \quad f_o(\mathbf{a}) \\ &\text{subject to} \quad g^j(\mathbf{a}) \geq 0, \qquad j = 1, \cdots, m, \end{aligned} \tag{1}$$

where m is the total number of constraints and $\mathbf{a} \in \Re^n$. The *Lagrangian* $L(\boldsymbol{\lambda}, \mathbf{a})$ augments the primal objective function with the constraints

$$L(\boldsymbol{\lambda}, \mathbf{a}) = f_o(\mathbf{a}) + \sum_{j=1}^{m} \lambda^j g^j(\mathbf{a}) = f_o(\mathbf{a}) + \boldsymbol{\lambda}^\top \mathbf{g}, \tag{2}$$

where $\mathbf{g}$ is the vector of constraint functions and $\boldsymbol{\lambda} \in \Re^m$ is the Lagrange multiplier vector with $\lambda^j \geq 0$.

Since the objective and constraint functions $f_o,\ g^1,\ \cdots,\ g^m$ are differentiable, if an optimum $\mathbf{a}^*$ exists, then the Lagrangian $L(\boldsymbol{\lambda}^*, \mathbf{a}^*)$ attains its maximum at the primal-dual pair $(\mathbf{a}^*, \boldsymbol{\lambda}^*)$ that must satisfy the *KKT* conditions

$$\nabla f_o(\mathbf{a}^*) + \sum_{j=1}^{m} \lambda^{j*}\ \nabla g^j(\mathbf{a}^*) \quad = \quad 0, \tag{3}$$

$$g^j(\mathbf{a}^*) \geq 0, \qquad \lambda^{j*} \quad \geq \quad 0, \tag{4}$$

$$\lambda^{j*} g^j(\mathbf{a}^*) \quad = \quad 0. \tag{5}$$

The KKT conditions provide a certificate for optimality.

In a centralized solution approach, the Lagrangian is maximized by search over the parameters $\mathbf{a}$ and $\boldsymbol{\lambda}$. This requires that all data and all parameters are available at a central controller. Although, proofs of optimality are simpler and well known for this centralized approach, for reasons stated in the introduction, we are interested in decentralized solutions.

B. Game Theory and Ordinal Potential Functions

For a distributed optimization approach, C_i will only adjust $\mathbf{a}_i$ and $\boldsymbol{\lambda}$. A challenge in formulating a distributed optimization problem is the decoupling of the system objective into local objectives, one for each agent. The game theory literature and the concept of *potentiality* provides guidance for addressing this challenge.

Consider $\mathbf{a} \in S$ and $\mathbf{a}_i,\ \mathbf{b}_i \in S_i$, where $S = S_1 \times \ldots \times S_{N_C}$ is the collection of all possible camera parameter settings in the game G, and S_i is the collection of all possible camera parameter settings for C_i. The sets S and S_i for $i = 1$ to N_C, referred to as action sets within the game theory literature, are compact. Let[2] $\phi_i(\mathbf{a}_i : \mathbf{a}_{-i})$ denote the local objective function of C_i. Game G_p is a *potential game* if $\exists$ a potential function $\phi_p : S \mapsto \Re$ such that $\forall \mathbf{a} \in S$ and $\forall \mathbf{a}_i, \mathbf{b}_i \in S_i$,

$$\phi_p(\mathbf{b}_i, \mathbf{a}_{-i}) - \phi_p(\mathbf{a}_i, \mathbf{a}_{-i}) = \phi_i(\mathbf{b}_i : \mathbf{a}_{-i}) - \phi_i(\mathbf{a}_i : \mathbf{a}_{-i}). \tag{6}$$

Game G_o is an *ordinal potential game* if $\exists$ an ordinal potential function $\phi_o : S \mapsto \Re$ such that $\forall \mathbf{a} \in S$ and $\forall \mathbf{a}_i, \mathbf{b}_i \in S_i$,

$$\begin{aligned} \phi_o(\mathbf{b}_i, \mathbf{a}_{-i}) - \phi_o(\mathbf{a}_i, \mathbf{a}_{-i}) \quad &> \quad 0 \\ \Leftrightarrow \phi_i(\mathbf{b}_i : \mathbf{a}_{-i}) \quad &- \quad \phi_i(\mathbf{a}_i : \mathbf{a}_{-i}) > 0. \end{aligned} \tag{7}$$

Potential games and ordinal potential games allow the global utility maximum to be achieved by maximization of the local utilities of each camera. When Eqn. (7) is satisfied, the local objective functions are said to be *aligned* with the global objective. Given $\phi_o(\mathbf{a})$, if the local utilities are defined as $\phi_i(\mathbf{a}_i : \mathbf{a}_{-i}) = \phi_o(\mathbf{a}_i, \mathbf{a}_{-i})$, then it is straightforward to show that the resulting game is a potential game.

[1] The notation $f_o(\mathbf{a})$ is short for $f_o(\mathbf{a} : \mathbf{x}^j, j = 1, \ldots, N_T)$, which is more precise in making explicit the fact that the value depends on the target state; however, it is too cumbersome to be effective.

[2] This notation $\phi_i(\mathbf{a}_i : \mathbf{a}_{-i})$ means that the value of the function ϕ_i may depend on both $\mathbf{a}_i$ and $\mathbf{a}_{-i}$, but that $\mathbf{a}_i$ as treated as an independent variable while $\mathbf{a}_{-i}$ is treated as a constant by C_i.

Thus, by defining the global objective function as an ordinal potential function with the individual local camera objectives aligned to it, the game becomes an ordinal potential game. When the set S is compact, and a game has a continuous potential function, then the game has at least one *Nash* Equilibrium. Therefore, given any feasible initial condition, at each step for which one camera increases its own utility, the global objective function increases correspondingly, due to G being a potential game. If ϕ_o is continuous and S is compact then $\phi_o(\mathbf{a})$ is bounded above; therefore, the optimization converges toward a maxima. At the maxima, no camera can achieve further improvement and thus a Nash equilibrium is reached.

C. Distributed Constrained Optimization

For the distributed approach define the local constrained optimization problem for the i-th camera

$$\begin{aligned} &\text{maximize} \quad f_i(\mathbf{a}_i) \\ &\text{subject to} \quad g^j(\mathbf{a}_i : \mathbf{a}_{-i}) \geq 0, \qquad j = 1, \cdots, m, \end{aligned} \tag{8}$$

where $\mathbf{a}_i \in \Re^{n_i}$, and $\mathbf{a}_{-i} \in \Re^{n-n_i}$. The local Lagrangian is

$$\begin{aligned} L_i(\boldsymbol{\lambda}, \mathbf{a}_i : \mathbf{a}_{-i}) &= f_i(\mathbf{a}_i) + \sum_{j=1}^{m} \lambda^j g^j(\mathbf{a}_i : \mathbf{a}_{-i}) \\ &= f_i(\mathbf{a}_i) + \boldsymbol{\lambda}^\top \mathbf{g}, \end{aligned} \tag{9}$$

where $\mathbf{g}$ is the vector constraint and $\boldsymbol{\lambda} \in \Re^m$ with $\lambda^j \geq 0$.

If we define the global objective function as the sum of local objective functions:

$$f_o(\mathbf{a}) = \sum_{i=1}^{N_C} f_i(\mathbf{a}_i), \tag{10}$$

then from Eqns. (2–5), (9), and (10), $\forall \mathbf{a} \in S$, and $\forall \mathbf{a}_i, \mathbf{b}_i \in S_i$,

$$\begin{aligned} & L(\boldsymbol{\lambda}, \mathbf{b}_i, \mathbf{a}_{-i}) - L(\boldsymbol{\lambda}, \mathbf{a}_i, \mathbf{a}_{-i}) > 0 \\ \Leftrightarrow \quad & L_i(\boldsymbol{\lambda}, \mathbf{b}_i : \mathbf{a}_{-i}) - L_i(\boldsymbol{\lambda}, \mathbf{a}_i : \mathbf{a}_{-i}) > 0, \end{aligned} \tag{11}$$

where the objective of each agent is to maximize its local Lagrangian. Therefore, L_i and L are aligned, and the global Lagrangian L is an ordinal potential function.

Remark 4: In a potential game, C_i can only choose its own action $\mathbf{a}_i$, but will take into account the proposed actions $\mathbf{a}_{-i}$ of all other agents. The actions of all other agents $\mathbf{a}_{-i}$ are determined by the other agents. Because C_i is the only agent able to select $\mathbf{a}_i$, consensus between agents is inappropriate for computation of $\mathbf{a}_i$. Instead, a modified flooding approach will be used, see Section VIII-B. At the same time, all cameras must collaboratively choose actions and $\boldsymbol{\lambda}$ to ensure that all constraints are satisfied. During distributed optimization each C_i has a local version of the Lagrange multiplier vector, denoted as $\boldsymbol{\lambda}_{(i)}$. A consensus algorithm is used to ensure the convergence of $\boldsymbol{\lambda}_{(i)}$ to a single value, see Section VIII-C.$\diamond$

VI. System Model

The continuous-time state space model of target T^j is:

$$\dot{\mathbf{x}}^j(t) = \mathbf{F}\ \mathbf{x}^j(t) + \mathbf{G}\ \boldsymbol{\omega}^j(t) \tag{12}$$

where $j = 1, \ldots, N_T$ is the target number and $\mathbf{x}^j = [{}^g\mathbf{p}^j; {}^g\mathbf{v}^j]$ with ${}^g\mathbf{p}^j$ and ${}^g\mathbf{v}^j$ representing the position and velocity vectors in the global (earth) frame. The process noise vector $\boldsymbol{\omega}^j \in \Re^3$ is assumed to be zero mean Gaussian with power spectral density $\mathbf{Q}$. The discrete-time equivalent model is:

$$\mathbf{x}^j(k+1) = \boldsymbol{\Phi}\mathbf{x}^j(k) + \boldsymbol{\gamma}(k) \tag{13}$$

where $\boldsymbol{\Phi} = e^{\mathbf{F}T}$ is the state transition matrix, $\boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q_d})$ is the process noise, and $T = t_{k+1} - t_k$ is the sampling period.

A. State Estimate Time Propagation

The state estimate and its error covariance matrix are propagated between sampling instants using [45]:

$$\hat{\mathbf{x}}^j(k+1)^- = \boldsymbol{\Phi}\hat{\mathbf{x}}^j(k)^+ \tag{14}$$

$$\mathbf{P}^j(k+1)^- = \boldsymbol{\Phi}\mathbf{P}^j(k)^+\boldsymbol{\Phi}^\top + \mathbf{Q}_d. \tag{15}$$

B. Camera Coordinate Transformations

Target T^j's position in the i-th camera's frame, ${}^{c_i}\mathbf{p}^j$, is related to its position in the global frame ${}^g\mathbf{p}^j$ by:

$${}^g\mathbf{p}^j = {}^g_{c_i}\mathbf{R}\ {}^{c_i}\mathbf{p}^j + {}^g\mathbf{p}_{c_i} \tag{16}$$

$${}^{c_i}\mathbf{p}^j = {}^{c_i}_g\mathbf{R}[{}^g\mathbf{p}^j - {}^g\mathbf{p}_{c_i}], \tag{17}$$

where ${}^g\mathbf{p}_{c_i}$ is the position of C_i in global frame and ${}^{c_i}_g\mathbf{R}$ is a rotation matrix that is a function of the camera mounting angle and $\boldsymbol{a}_i$.

C. Measurement Model

Camera measurement models are derived in various references [32], [46]. The following presents final results only for the expressions needed in this article. The derivation in [32] uses a notation similar to that herein.

Let the coordinates of target T^j in camera C_i's frame be ${}^{c_i}\mathbf{p}^j = \left[{}^{c_i}x^j, {}^{c_i}y^j, {}^{c_i}z^j\right]^\top$. The standard pin-hole perspective projection camera model for C_i and T^j, assuming that T^j is in the FoV of camera C_i is,

$${}^{i_i}\mathbf{u}^j = \begin{bmatrix} \frac{F_i}{s_x}\frac{{}^{c_i}x^j}{{}^{c_i}z^j} + o_x \\ \frac{F_i}{s_y}\frac{{}^{c_i}y^j}{{}^{c_i}z^j} + o_y \end{bmatrix} + {}^{i_i}\boldsymbol{\eta}^j, \tag{18}$$

where s_x and s_y give the effective size of a pixel in (m/pixel) measured in the horizontal and vertical directions, respectively; F_i is the focal length setting defined by $\mathbf{a}_i$; the point (o_x, o_y) gives the coordinates of the image plane center in pixels; and the measurement noise ${}^{i_i}\boldsymbol{\eta}^j \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_i^j)$ with $\mathbf{C}_i^j(\mathbf{a}_i) \in \Re^{2\times 2}$. The fact that the measurement noise covariance $\mathbf{C}_i^j$ is dependent on $\mathbf{a}_i$ is important. Note, for example, that as the focal length increases, the size of a target in the image increases and the pixel uncertainty of its location changes.

Given the estimated state and the camera model, the predicted measurement is

$${}^{i_i}\hat{\mathbf{u}}^j = \begin{bmatrix} \frac{F_i}{s_x}\frac{{}^{c_i}\hat{x}^j}{{}^{c_i}\hat{z}^j} + o_x \\ \frac{F_i}{s_y}\frac{{}^{c_i}\hat{y}^j}{{}^{c_i}\hat{z}^j} + o_y \end{bmatrix}. \tag{19}$$

The measurement residual ${}^{i_i}\tilde{\mathbf{u}}^j$ is

$$ {}^{i_i}\tilde{\mathbf{u}}^j = {}^{i_i}\mathbf{u}^j - {}^{i_i}\hat{\mathbf{u}}^j. \tag{20}$$

D. Observation Matrix $\boldsymbol{H}_i^j$

The linearized relationship between the residual and the position error vector is

$$ {}^{i_i}\mathbf{u}^j - {}^{i_i}\hat{\mathbf{u}}^j \approx \mathbf{H}_i^j \left({}^{g}\mathbf{p}^j - {}^{g}\hat{\mathbf{p}}^j \right), \tag{21}$$

where $\mathbf{H}_i^j = \left. \frac{\partial^{i_i}\mathbf{u}^j}{\partial {}^{g}\mathbf{p}^j} \right|_{{}^{g}\hat{\mathbf{p}}^j} \in \Re^{2\times 3}$ is defined in [34]. Note that $\mathbf{H}_i^j$ is a function of both ${}^{g}\mathbf{p}^j$ and $\mathbf{a}_i$. When T^j is not in FoV_i, then $\mathbf{H}_i^j = \mathbf{0} \in \Re^{2\times 3}$.

E. Measurement Update

Using $\mathbf{H}_i^j$ and Eqn. (20), a measurement update for the state estimates and error covariances for targets in the area is performed using the Information-Consensus Filter from [18].

VII. Image Optimization Methodology

The goal is to track all targets at all times to a specified accuracy $\bar{\mathbf{T}}$, to discover times-of-opportunity at which improved imagery is obtainable, and to determine the sequence of PTZ parameters for each camera to achieve these goals. *Improved imagery* means that the system should only attempt to acquire a high-res image of any target T^j at the next imaging instant if that image is expected to yield higher imaging value for that target than is already available. The method implements distributed, simultaneous optimization by all agents to compute the optimal camera parameters $\mathbf{a}^*$ relative to Eqn. (1) at each imaging time-instant.

This section starts with specification of the global objective and constraints that are subsequently decoupled into local objectives for each camera. The global objective function for the constrained optimization problem is designed as a Bayesian imaging value function that accounts for the risk in imaging the target. Risk will be formulated using the Fisher information matrix defined as $\mathbf{J}^j(k+1)^- = \left(\mathbf{P}^j(k+1)^-\right)^{-1}$. The prior covariance matrix $\mathbf{P}^j(k+1)^-$ computed using Eqn. (15) can be written in block form as[3]:

$$ \mathbf{P}^{j-} = \begin{bmatrix} \mathbf{P}_{\mathbf{pp}}^{j-} & \mathbf{P}_{\mathbf{pv}}^{j-} \\ \mathbf{P}_{\mathbf{vp}}^{j-} & \mathbf{P}_{\mathbf{vv}}^{j-} \end{bmatrix}, \tag{22}$$

where $\mathbf{P}_{\mathbf{pp}}^{j-}$ is the prior, position error-covariance matrix.

A. Global Imaging Value Function

This section discusses the design and desired properties of the global imaging value function $V_I\left(\mathbf{a} : {}^{g}\hat{\mathbf{p}}^{j-}, \mathbf{P}_{\mathbf{pp}}^{j-}\right)$ and the global constraint set. The notation $V_I\left(\mathbf{a} : {}^{g}\hat{\mathbf{p}}^{j-}, \mathbf{P}_{\mathbf{pp}}^{j-}\right)$ with $j = 1, \cdots, N_T$ makes explicit that the optimization variable is $\mathbf{a}$, while the value also depends on the distribution of target T^j which is parameterized by $\left({}^{g}\hat{\mathbf{p}}^{j-}, \mathbf{P}_{\mathbf{pp}}^{j-}\right)$. For ease of notation, from this point in the paper, we will drop dependence of $V_I\left(\mathbf{a}\right)$ on ${}^{g}\hat{\mathbf{p}}^{j-}$ and $\mathbf{P}_{\mathbf{pp}}^{j-}$, unless needed for clarity.

[3]The time argument $(k+1)$ is dropped for ease of notation.

1) Value Function Properties: The imaging value function should have the following properties:

Continuously differentiable: This is necessary for proofs of convergence, and greatly facilitates numeric optimization.

Increases with image quality: Herein, image quality is defined by two parameters: image resolution and relative pose between the imaging camera and the imaged target.

Image resolution $r_i^j(\mathbf{a}_i, {}^{g}\hat{\mathbf{p}}^j)$, which is a positive real number, will be quantified by the number of pixels occupied by T^j on camera C_i's image plane. Given ${}^{g}\hat{\mathbf{p}}^j$, the resolution increases monotonically with zoom ζ of the imaging camera.

Relative pose between camera C_i and target T^j will be quantified by the scalar quality factor $\alpha_i^j(\mathbf{a}_i)$. Let vector $\mathbf{o}_{v^j}$ be the target's velocity vector. Define the vector $\mathbf{o}_{C_i}$ to be the i-th camera's optical axis direction in the global frame

$$ \mathbf{o}_{C_i} = {}^{g}_{c_i}\mathbf{R}\ {}^{c_i}\mathbf{e}_3, \tag{23}$$

where $\mathbf{e}_3 = [0,0,1]^\top$. Define $\mathbf{o}_{T^j}$ to be the vector from camera C_i's position to target T^j's estimated position. Using the vectors $\mathbf{o}_{v^j}$, $\mathbf{o}_{C_i}$, and $\mathbf{o}_{T^j}$ we define the scalars

$$ o_c = \frac{\mathbf{o}_{C_i} \cdot \mathbf{o}_{T^j}}{\| \mathbf{o}_{C_i} \| \| \mathbf{o}_{T^j} \|}, \text{ and } o_o = \frac{\mathbf{o}_{C_i} \cdot \mathbf{o}_{v^j}}{\| \mathbf{o}_{C_i} \| \| \mathbf{o}_{v^j} \|}. \tag{24}$$

The scalar $o_c \in [-1,1]$ yields the maximum possible positive value of 1 if camera C_i images target T^j such that T^j is at the center of its FoV. The scalar $o_o \in [-1,1]$ has maximum magnitude when T^j's motion vector $\mathbf{o}_{v^j}$ is pointing directly toward or away from camera C_i.

To define $\alpha_i^j(\mathbf{a}_i)$, we use the following assumption.

Assumption 1: (Facial Direction) Target T^j faces in the direction indicated by vector $\mathbf{o}_{v^j}$.

From Assumption 1 and Eqn. (24), when the scalar $o_o < 0$, T^j is likely to be facing camera C_i. This condition differentiates between targets facing C_i and those facing away from it. The relative pose quality factor is thus defined as

$$ \alpha_i^j(\mathbf{a}_i) = \begin{cases} (o_c\ o_o)^2 & \text{if } o_o < 0 \\ 0 & \text{otherwise.} \end{cases} \tag{25}$$

Hence when $\alpha_i^j \in [0,1]$ is large, it is likely that T^j is facing C_i and at the center of C_i's FoV. The imaging value obtained by camera C_i for imaging T^j located at ${}^{g}\mathbf{p}^j$ is defined as

$$ V_{I_i}^j(\mathbf{a}_i, {}^{g}\mathbf{p}^j) = r_i^j(\mathbf{a}_i)\ \alpha_i^j(\mathbf{a}_i). \tag{26}$$

Balanced Risk: Risk is defined as the probability that the target is outside of the FoV of the cameras that are expected to image it. Risk increases monotonically with zoom ζ, because the ground-plane area within the FoV decreases as ζ increases. Herein, we will address risk by using the expected value of the tracking constraints and the imaging value.

To understand the issues involved, it is informative to briefly consider the simple case where $N_T = 1$ and $\alpha_i^1 > 0$ for $i = 1, \ldots, N_C$. For this case, if risk was neglected and V_I was defined with the properties mentioned above, then each camera would maximize its focal length and select its pan and tilt parameters to center on the expected target location. If instead, the value accounted appropriately for risk, then one or more camera might significantly increase its zoom parameter, while at least one of the remaining cameras would use lower zoom

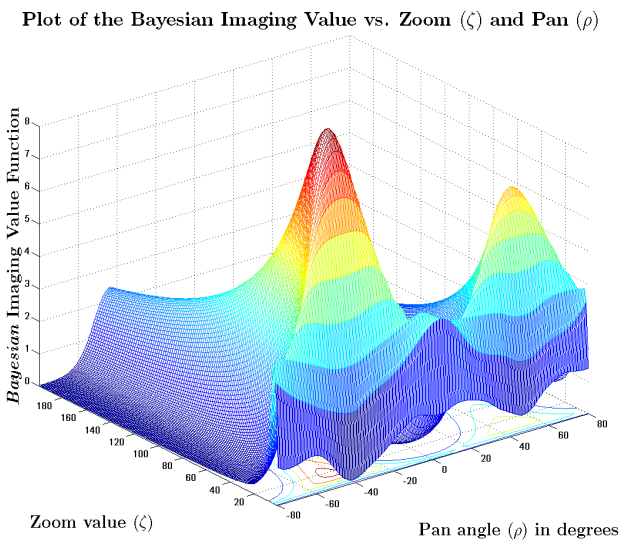


Fig. 2: Non-convexity of the Bayesian Imaging Value Function: The figure is a plot of an example Bayesian imaging value function that highlights the multimodal nature of the function for scenarios where $N_T > 1$.

parameters, to decrease the tracking risk due to the uncertainty in the estimated target position. The camera at the highest zoom setting would be the one at the best aspect angle.

Remark 5: Fig. 2 depicts the appearance of an example Bayesian Imaging Value function for a scenario with $N_T = 2$ targets in an area monitored by $N_C = 1$ camera. As shown in the figure, when $N_T > 1$, the summation of the per target expected imaging value across the targets for any camera will typically yield a multimodal (i.e., nonconvex) objective function. Given the expected target positions and their respective distributions, for a constant tilt angle τ, the plot shows how the value of the objective function changes versus zoom ($\zeta \in [\underline{\zeta}, \bar{\zeta}]$), and pan angle ($\rho \in [\underline{\rho}, \bar{\rho}]$), It can be seen that the multi-modal nature of the function is exaggerated for higher values of ζ. Thus, the possibility of multiple targets makes the visual sensing problem inherently non-convex and provides challenges in achieving optimal solutions. Non-convexity is further discussed in Section IX.$\diamond$

2) Imaging Value $V_I(\boldsymbol{a} : {}^g\hat{\mathbf{p}}^{j-}, \mathbf{P}^{j-}_{\mathbf{pp}})$: We define the global Bayesian image value function as

$$V_I(\mathbf{a}) = \sum_{i=1}^{N_C} \sum_{j=1}^{N_T(t)} w_i^j(t)\ E\left\langle V_{I_i}^j\left(\mathbf{a}_i,\ {}^g\mathbf{p}^j\right)\right\rangle \quad (27)$$
$$= \sum_{i=1}^{N_C} \int_{\text{FoV}_i} \left(\sum_{j=1}^{N_T(t)} w_i^j(t)\ V_{I_i}^j(\mathbf{a}_i,\ \mathbf{z})\ p_{\mathbf{p}^j}(\mathbf{z}) \right) d\mathbf{z},$$

where $w_i^j(t)$ is a time-varying local dynamic imaging weight that magnifies the importance of imaging certain targets relative to other targets[4]. Given the assumptions herein, the probability distribution $p_{\mathbf{p}^j}(\mathbf{z})$ of the position of T^j in the global frame at the next imaging instant is the Normal distribution $\mathcal{N}({}^g\hat{\mathbf{p}}^{j-}, \mathbf{P}^{j-}_{\mathbf{pp}})$. The dummy variable $\mathbf{z}$ representing target position is used for integration over the ground plane, where the region of integration is the i-th camera's FoV.

Each camera integrates over its own FoV. The integral of image quality over FoV_i as a function of probability weighted target position yields the Bayesian value function, which provides the desired tradeoff between image quality and risk.

3) Performance Constraints: The performance constraints will be defined as a function of the posterior *Fisher Information Matrix* $\mathbf{J}^{j+}\left(\mathbf{a} : {}^g\hat{\mathbf{p}}^{j-}, \mathbf{P}^{j-}_{\mathbf{pp}}\right)$.

Fisher Information: The Fisher Information $\mathbf{J}^j$ for T^j in block form is

$$\mathbf{J}^j = \begin{bmatrix} \mathbf{J}^j_{\mathbf{pp}} & \mathbf{J}^j_{\mathbf{pv}} \\ \mathbf{J}^j_{\mathbf{vp}} & \mathbf{J}^j_{\mathbf{vv}} \end{bmatrix}, \quad (28)$$

where, $\mathbf{J}^j_{\mathbf{pp}}$ represents the position information matrix. The posterior position information matrix $\mathbf{J}^{j+}_{\mathbf{pp}}$ is given by

$$\mathbf{J}^{j+}_{\mathbf{pp}} = \mathbf{J}^{j-}_{\mathbf{pp}} + \sum_{i=1}^{N_C} \mathbf{H}_i^{j\top} \left(\mathbf{C}_i^j\right)^{-1} \mathbf{H}_i^j, \quad (29)$$

where, $\mathbf{J}^{j-}_{\mathbf{pp}}$ is the prior information about T^j. As was shown in Section VI-D, $\mathbf{H}_i^j$ and $\mathbf{C}_i^j$ are functions of $\mathbf{a}_i$ and $\mathbf{H}_i^j$ is a function of the target position. Therefore, $\mathbf{J}^{j+}_{\mathbf{pp}}$ depends on $\mathbf{a}$ and on the target position. Computation of the expected tracking accuracy should account for this variation and for the probability that $T^j \in FoV_i$.

Tracking Performance: We define a vector $\mathbf{U}^j_T(\mathbf{a} : {}^g\hat{\mathbf{p}}^{j-}, \mathbf{P}^{j-}_{\mathbf{pp}})$ as a measure of tracking performance for each target in the area. One example is $\mathbf{U}^j_T(\mathbf{a}) = \text{diag}\left(\mathbf{J}^{j+}_{\mathbf{pp}}\right)$. Because the quantity $\mathbf{U}^j_T(\mathbf{a})$ depends on whether T^j is within the FoV of each camera that is expected to image it, we define the global Bayesian tracking value vector $\mathbf{V}^j_T(\mathbf{a})$ as the expected value of the tracking performance vector $\mathbf{U}^j_T(\mathbf{a})$ over the position of T^j computed across all the camera's FoVs:

$$\mathbf{V}^j_T(\mathbf{a}) = E_{\mathbf{p}^j}\left\langle \mathbf{U}^j_T(\mathbf{a})\right\rangle = \int \left(\mathbf{U}^j_T(\mathbf{a})\ p_{\mathbf{p}^j}(\mathbf{z})\right) d\mathbf{z}, \quad (30)$$

where all variables are as defined in Eqn. (27) and the summation over all cameras is accounted for already in eqn. (29), which also accounts for prior information.

Tracking Constraint: Each target's tracking constraint is

$$\mathbf{V}^j_T(\mathbf{a}) \succeq \bar{\mathbf{T}}^j, \quad (31)$$

where $\bar{\mathbf{T}}^j$ is the user specified lower bound on the tracking information about target T^j. Due to the reciprocal relation between (scalar) information and covariance, the reciprocal of $\bar{\mathbf{T}}^j$ is the upper bound on the covariance of target T^j's state estimate. The tracking performance threshold $\bar{\mathbf{T}}^j$ is hence measured in m^{-2}. The notation '$\succeq$' in Eqn. (31) indicates a per-element vector inequality. Stacking the Bayesian tracking value vectors for each target, we obtain

$$\mathbf{V}_T(\mathbf{a}) = \left[\mathbf{V}^1_T, \cdots, \mathbf{V}^j_T, \cdots, \mathbf{V}^{N_T}_T\right]^\top, \quad (32)$$

and rewrite Eqn. (31) for all targets presently in the area as:

$$\mathbf{V}_T(\mathbf{a}) \succeq \bar{\mathbf{T}}, \quad (33)$$

where $\mathbf{V}_T(\mathbf{a}), \bar{\mathbf{T}}, \mathbf{0} \in \Re^m$ with $m = N_T(t)\text{dim}({}^g\mathbf{p}^j)$. Eqn. (33) is the global tracking constraint.

[4]Specification of $w_i^j(t)$ is application dependent. It could be constant, user-specified, or could increase as the target approaches a specified location such as an exit. See the example in eqn. (49).

4) Global Problem Summary: The constrained global imaging value maximization problem can be written as

$$\begin{aligned} &\text{maximize} \quad V_I\left(\mathbf{a}:\ {}^g\hat{\mathbf{p}}^{j-},\ \mathbf{P}^{j-}_{\mathbf{pp}}\right) \\ &\text{subject to} \quad \left[\mathbf{V}_T\left(\mathbf{a}:\ {}^g\hat{\mathbf{p}}^{j-},\ \mathbf{P}^{j-}_{\mathbf{pp}}\right) - \bar{\mathbf{T}}\right] \succeq \mathbf{0}. \end{aligned} \tag{34}$$

The global Lagrangian $L(\boldsymbol{\lambda}, \mathbf{a})$ is

$$L(\boldsymbol{\lambda}, \mathbf{a}) = V_I\left(\mathbf{a}\right) + \boldsymbol{\lambda}^\top \left[\mathbf{V}_T(\mathbf{a}) - \bar{\mathbf{T}}\right], \tag{35}$$

where $L : (\boldsymbol{\lambda}, \mathbf{a}) \mapsto \Re$, and $\boldsymbol{\lambda} \in \Re^m$ is the Lagrange multiplier vector. Thus, to find the optimal primal-dual pair of solutions $(\mathbf{a}^*, \boldsymbol{\lambda}^*)$ through a central controller, the global unconstrained problem given by the Lagrangian in Eqn. (35) would be solved.

B. Decoupling the Global Problem

Due to the desired distributed nature of our solution, we need to decompose the global problem into smaller local problems that are solvable by each camera.

In our problem formulation, we allow camera C_i to optimize only its own camera parameter settings $\mathbf{a}_i$. Using this system restriction, we define the local Bayesian imaging value function for C_i as

$$V_{I_i}(\mathbf{a}_i) = \int_{FoV_i} \sum_{j=1}^{N_T(t)} \left(w_i^j(t)\ V_{I_i}^j(\mathbf{a}_i,\ \mathbf{z})\ p_{\mathbf{p}^j}\left(\mathbf{z}\right)\right) d\mathbf{z}. \tag{36}$$

Define $\mathbf{V}_{T_i}(\mathbf{a}_i) = \mathbf{V}_T(\mathbf{a}_i\ :\ \mathbf{a}_{-i})$. This notation concisely indicates that C_i can only alter $\mathbf{a}_i$, where for the purpose of its local optimization $\mathbf{a}_{-i}$ is fixed. Note that $\mathbf{V}_T(\mathbf{a}_i\ :\ \mathbf{a}_{-i})$ is distinct from $\mathbf{V}_T(\mathbf{a}_i, \mathbf{a}_{-i}) = \mathbf{V}_T(\mathbf{a})$ and that $\max_{\mathbf{a}_i \in S_i} \mathbf{V}_{T_i}(\mathbf{a}_i) \leq \max_{\mathbf{a} \in S} \mathbf{V}_T(\mathbf{a})$. Each agent will have the constraint $\mathbf{V}_T(\mathbf{a}_i : \mathbf{a}_{-i}) \succeq \bar{\mathbf{T}}$. While C_i is changing $\mathbf{a}_i$, the other agents are simultaneously changing their sub vectors of $\mathbf{a}_{-i}$ and all agents are broadcasting their current locally optimal values through the network. Thus the tracking constraint for camera C_i is

$$\mathbf{V}_{T_i}(\mathbf{a}_i) \succeq \bar{\mathbf{T}}. \tag{37}$$

Note that,

$$\mathbf{V}_{T_i}(\mathbf{a}_i) \succeq \bar{\mathbf{T}} \quad \Leftrightarrow \quad \mathbf{V}^j_{T_i}(\mathbf{a}_i) \succeq \bar{\mathbf{T}}^j \text{ for } j = 1, \ldots, N_T(t).$$

The *Fisher* Information given in Eqn. (29) can be reorganized as:

$$\mathbf{J}^{j+}_{\mathbf{pp}} \quad = \quad \left[\mathbf{J}^{j-}_{\mathbf{pp}} + \mathbf{H}^{j\top}_{-i}\left(\mathbf{C}^j_{-i}\right)^{-1}\mathbf{H}^j_{-i}\right] + \mathbf{H}^{j\top}_i\left(\mathbf{C}^j_i\right)^{-1}\mathbf{H}^j_i.$$

For the process of C_i optimizing its parameter vector $\mathbf{a}_i$, the contribution from prior information and all other cameras (term in brackets) is independent of $\mathbf{a}_i$ and considered by C_i to be constant and known. The term $\left[\mathbf{H}^{j\top}_{-i}\left(\mathbf{C}^j_{-i}\right)^{-1}\mathbf{H}^j_{-i}\right]$ is computed from $\mathbf{a}_{-i}$ which will be available through the distributed optimization process discussed in Section VIII.

Thus from Eqns. (29-30), we can write

$$E_{\mathbf{p}^j}\left\langle \text{diag}\left(\mathbf{J}^j_i\right)\right\rangle \succeq \bar{\mathbf{T}} - E_{\mathbf{p}^j}\left\langle \text{diag}\left(\mathbf{J}^{j-}_{\mathbf{pp}} + \mathbf{J}^j_{-i}\right)\right\rangle, \tag{38}$$

where $\mathbf{J}^j_i = \mathbf{H}^{j\top}_i\left(\mathbf{C}^j_i\right)^{-1}\mathbf{H}^j_i$ and $\mathbf{J}^j_{-i} = \mathbf{H}^{j\top}_{-i}\left(\mathbf{C}^j_{-i}\right)^{-1}\mathbf{H}^j_{-i}$. The right hand side of this inequality represents, for the current proposed settings of the other cameras $\mathbf{a}_{-i}$, the expected improvement in tracking accuracy required from C_i for imaging T^j to have a feasible global solution. Targets for which the right-hand side of Eqn. (38) is negative can be removed from the set of tracking constraints for C_i.

From Eqns. (36) and (37), the local imaging value maximization problem can be written as

$$\begin{aligned} &\text{maximize} \quad V_{I_i}\left(\mathbf{a}_i:\ {}^g\hat{\mathbf{p}}^{j-},\ \mathbf{P}^{j-}_{\mathbf{pp}}\right) \\ &\text{subject to} \quad \mathbf{V}_{T_i}\left(\mathbf{a}_i:\ \mathbf{a}_{-i},\ {}^g\hat{\mathbf{p}}^{j-},\ \mathbf{P}^{j-}_{\mathbf{pp}}\right) \succeq \bar{\mathbf{T}}. \end{aligned} \tag{39}$$

The local Lagrangian $L_i(\boldsymbol{\lambda}_{(i)}, \mathbf{a}_i)$ is

$$L_i\left(\boldsymbol{\lambda}_{(i)}, \mathbf{a}_i\right) \quad = \quad V_{I_i}(\mathbf{a}_i) + \boldsymbol{\lambda}^\top_{(i)}\left[\mathbf{V}_{T_i}(\mathbf{a}_i) - \bar{\mathbf{T}}\right]. \tag{40}$$

Thus, for camera C_i to find its local optimal primal-dual pair of solutions $(\mathbf{a}^*_i, \boldsymbol{\lambda}^*_{(i)})$, C_i will maximize the local unconstrained Lagrangian given in Eqn. (40).

In this approach, all cameras in the network optimize simultaneously. The subscript (i) on $\boldsymbol{\lambda}_{(i)}$ in Eqn. (40) indicates that the Lagrange multiplier vector picked by camera C_i to solve the problem is a local variable and may not be globally the same throughout the network. In order to overcome this predicament, cameras in the network employ a variant of the algorithm described in [47], [48] to perform dynamic average consensus over the local Lagrange multiplier vectors. This results in a consensus-step after each optimization-step The algorithm is explained in detail in Section VIII.

C. Lagrangian as an Ordinal Potential Function

For the problem stated in Eqn. (34), note that the global objective of the multi-camera network defined in Eqn. (27) is the sum over the local objectives defined in Eqn. (36)

$$V_I(\mathbf{a}) \quad = \quad \sum_{i=1}^{N_C} V_{I_i}(\mathbf{a}_i). \tag{41}$$

At each optimization step κ, the i-th camera adjusts $\boldsymbol{\lambda}_{(i)}(\kappa)$ and $\mathbf{a}_i(\kappa)$, leaving $\mathbf{a}_{-i}(\kappa)$ fixed, to solve the problem in Eqn. (39) with $V_{I_i}(\mathbf{a}_i)$ defined in Eqn. (36). Dynamic average consensus over $\boldsymbol{\lambda}_{(i)}$ between optimization steps forces each agent's local value toward a non-negative consensus agreement vector $\bar{\boldsymbol{\lambda}}(\kappa) = \frac{1}{N_c}\sum_i \boldsymbol{\lambda}_{(i)}(\kappa)$. Convergence of the dynamic game is assured when the local Lagrangians $L_i(\boldsymbol{\lambda}_{(i)}, \mathbf{a}_i)$, and the global Lagrangian $L(\boldsymbol{\lambda}, \mathbf{a})$ form an ordinal potential game.

From Eqns. (35), (40),and (41), $\forall \boldsymbol{\lambda}_b,\ \boldsymbol{\lambda}_a \succeq \mathbf{0}$, let $\tilde{L} = L(\boldsymbol{\lambda}_b,\ \mathbf{b}_i,\ \mathbf{a}_{-i}) - L(\boldsymbol{\lambda}_a,\ \mathbf{a}_i,\ \mathbf{a}_{-i})$. Thus,

$$\begin{aligned} \tilde{L} \quad &= \quad \sum_{i=1}^{N_C} V_{I_i}(\mathbf{b}_i) + \boldsymbol{\lambda}_b^\top\left[\mathbf{V}_T\left(\mathbf{b}_i,\ \mathbf{a}_{-i}\right) - \bar{\mathbf{T}}\right] \\ &\quad - \sum_{i=1}^{N_C} V_{I_i}(\mathbf{a}_i) - \boldsymbol{\lambda}_a^\top\left[\mathbf{V}_T\left(\mathbf{a}_i,\ \mathbf{a}_{-i}\right) - \bar{\mathbf{T}}\right] \\ &= \quad V_{I_i}(\mathbf{b}_i) + \sum_{l\neq i} V_{I_l}(\mathbf{a}_l) + \boldsymbol{\lambda}_b^\top\left[\mathbf{V}_T\left(\mathbf{b}_i,\ \mathbf{a}_{-i}\right) - \bar{\mathbf{T}}\right] \\ &\quad - V_{I_i}(\mathbf{a}_i) - \sum_{l\neq i} V_{I_l}(\mathbf{a}_l) - \boldsymbol{\lambda}_a^\top\left[\mathbf{V}_T\left(\mathbf{a}_i,\ \mathbf{a}_{-i}\right) - \bar{\mathbf{T}}\right] \\ &= \quad L_i(\boldsymbol{\lambda}_b,\ \mathbf{b}_i :\ \mathbf{a}_{-i}) - L_i(\boldsymbol{\lambda}_a,\ \mathbf{a}_i :\ \mathbf{a}_{-i}). \end{aligned}$$

Hence, $\forall \mathbf{a} \in S$, $\forall \mathbf{a}_i, \mathbf{b}_i \in S_i$, and $\forall \boldsymbol{\lambda}_b,\ \boldsymbol{\lambda}_a \succeq \mathbf{0}$,

$$\begin{aligned} & L(\boldsymbol{\lambda}_b,\ \mathbf{b}_i,\ \mathbf{a}_{-i}) - L(\boldsymbol{\lambda}_a,\ \mathbf{a}_i,\ \mathbf{a}_{-i}) > 0 \\ \Leftrightarrow\quad & L_i(\boldsymbol{\lambda}_b,\ \mathbf{b}_i :\ \mathbf{a}_{-i}) - L_i(\boldsymbol{\lambda}_a,\ \mathbf{a}_i :\ \mathbf{a}_{-i}) > 0. \end{aligned}$$

Therefore, as explained in Sections V-B and V-C, Eqns. (35) and (40) form an ordinal potential game.

VIII. Distributed Optimization

The distributed optimization process can be broken down into three separate steps, where κ denotes the iteration counter:

1) **Camera Parameter Optimization:** Each camera C_i computes $(\mathbf{a}_i, \boldsymbol{\lambda}_{(i)})$ to increase $L_i\left(\boldsymbol{\lambda}_{(i)}, \mathbf{a}_i : \mathbf{a}_{-i}\right)$ while holding $\mathbf{a}_{-i}$ constant. It then communicates the newly computed local primal-dual pair estimates $\left(i, \kappa, \mathbf{a}_i(\kappa), \boldsymbol{\lambda}_{(i)}(\kappa)\right)$ and new portions of $\mathbf{a}_{-i}$ to its neighbors $\mathcal{N}_i$.
2) **Camera Parameter Replacement:** Each camera C_n that is a neighbor of C_i (i.e. $C_n \in \mathcal{N}_i$) receives $\left(i, \kappa, \mathbf{a}_i(\kappa), \boldsymbol{\lambda}_{(i)}(\kappa), \mathbf{a}_{-i}\right)$. It replaces its previous value of $(\mathbf{a}_i, \mathbf{a}_{-i})$ using the rules of replacement described in Section VIII-B.
3) **Consensus on Lagrange Multipliers:** C_i performs dynamic average consensus on its local Lagrange multiplier vector $\boldsymbol{\lambda}_{(i)}(\kappa)$ and the Lagrange multiplier vectors received from cameras in $\mathcal{N}_i$ to converge towards a consensus Lagrange multiplier vector $\bar{\boldsymbol{\lambda}}$, using the Lagrange multiplier update law in Eqn. (43), defined in Section VIII-C.

This distributed optimization process is then iterated over κ until a stopping criteria is achieved.

Since the optimization problem described by Eqn. (39) is non-convex, any solution found may only be locally optimal. It is assumed that all agents start with identical values of $\mathbf{a}(\kappa)$ for $\kappa = 0$ and that $\mathbf{a}(0)$ is not on the separatrix dividing the domain of attraction of one local optimum from another.

A. Connectivity, Communication, and Consensus

The approach requires the following standard assumptions on the camera communication graph.

Assumption 2: (Connectivity) The camera communication graph is undirected, and *connected*, i.e. there exists at least one communication path from each agent to every other agent in the network.

Remark 6: In [48], each agent changes $\boldsymbol{\lambda}$ and the entire vector $\mathbf{a}$ while computing a dual solution, then using consensus on both $\mathbf{a}$ and $\boldsymbol{\lambda}$. Herein, agent C_i only optimizes $\boldsymbol{\lambda}$ and $\mathbf{a}_i$, which is a subvector of $\mathbf{a}$. When the subvector $\mathbf{a}_i$ is broadcast to the neighbors of C_i, they pass it to their neighbors. Each agent receiving a newer value of $\mathbf{a}_i$ replaces their older value. Thus, for the approach herein, each camera C_i need only perform dynamic average consensus on $\boldsymbol{\lambda}_{(i)}$ and the set of Lagrange multiplier vectors $\{\boldsymbol{\lambda}_{(n)}\}$ for $C_n \in \mathcal{N}_i$. Connectivity ensures that the changes to $\mathbf{a}_i$ and $\boldsymbol{\lambda}_{(i)}$ by each C_i eventually affect all agents in the network. The convergence of consensus is asymptotic, but becomes trivial for strictly feasible solutions, which have $\boldsymbol{\lambda} = 0$. The effects of a change in any $\mathbf{a}_i$ are fully distributed throughout the network in a finite number of steps, which is less than the diameter of the network.$\diamond$

Assumption 3: (Weights Rule) There exists a scalar $\beta > 0$ such that for each $i \in [1, N_c]$, $\omega_{ii}(\kappa) \geq \beta$, and $\omega_{in}(\kappa) \in [\beta, 1]$ for $C_n \in \mathcal{N}_i$. If cameras C_i and C_n are not directly connected, then $\omega_{in}(\kappa) = 0$.

Assumption 4: (Double Stochasticity) Let $B_i = C_i \bigcup \mathcal{N}_i$, and $\sum_{l \in B_i} \omega_{il}(\kappa) = 1$ and $\sum_{i \in B_i} \omega_{il}(\kappa) = 1$.

Assumption 3 ensures that all cameras are *influential* [47] while performing consensus on the local Lagrange multiplier vectors, and Assumption 4 ensures that all cameras asymptotically converge to a consensus Lagrange multiplier vector $\bar{\boldsymbol{\lambda}} \succeq \mathbf{0}$ [49].

B. Camera Parameter Replacement Rule

We use a variant of the flooding algorithm [50] to propagate the local variables through the network of cameras. After C_i computes $\mathbf{a}_i(\kappa)$, it delivers the information $\{i, \kappa, \mathbf{a}_i(\kappa), \boldsymbol{\lambda}_{(i)}(\kappa)\}$ to its neighbors $\mathcal{N}_i$, and will rebroadcast to its neighbors any *updated PTZ information*, $\{l, \mathbf{a}_l, \kappa_l\}$ for $l \neq i$, that it received since the last broadcast. Using rebroadcast, each agent's parameter updates travel throughout a connected network exactly one time. For the l-th subvector in $C_i's$ version of $\mathbf{a}$, C_i has a value $\mathbf{a}_l(\kappa_l)$ and a time-stamp κ_l both computed by C_l, even if C_i and C_l are not neighbors. Because the network may contain loops, C_i may receive information about other cameras via multiple paths. C_i will replace its l-th subvector with the received information only if the time-stamp in $\{l, \mathbf{a}_l, \kappa_l\}$ is more recent than the time stamp corresponding to the value it is currently using. Otherwise, the message is discarded without rebroadcast.

C. Distributed Lagrangian Consensus

At iteration κ camera C_i receives the set of Lagrange multiplier vectors $\{\boldsymbol{\lambda}_{(n)}(\kappa)\}$ for $C_n \in \mathcal{N}_i$. It also has its local copy of $\mathbf{a}$. Its local computations must jointly optimize $\mathbf{a}_i$ and $\boldsymbol{\lambda}$ (given $\mathbf{a}_{-i}$) while also converging toward agreement across the network on the value of $\boldsymbol{\lambda}$. This section describes dynamic average consensus on the local versions of Lagrange multiplier vectors. Following the notation in [48], we refer to this as a distributed Lagrangian consensus algorithm.

Camera C_i iteratively optimizes using the update law [47]:

$$\mathbf{a}_i(\kappa+1) = -s(\kappa)\ \mathbf{D}_{\mathbf{a}_{(i)}}(\kappa) \tag{42}$$

$$\boldsymbol{\lambda}_i(\kappa+1) = \boldsymbol{\nu}_{\boldsymbol{\lambda}_{(i)}}(\kappa) - s(\kappa)\ \mathbf{D}_{\boldsymbol{\lambda}_{(i)}}(\kappa), \tag{43}$$

where the scalar $s(\kappa) > 0$ is the step-size,

$$\mathbf{D}_{\mathbf{a}_i} = \nabla_{\mathbf{a}_i}\ L_i\left(\boldsymbol{\lambda}_{(i)}(\kappa),\ \mathbf{a}_i(\kappa) :\ \mathbf{a}_{-i}(\kappa)\right),$$

and

$$\begin{aligned} \mathbf{D}_{\boldsymbol{\lambda}_{(i)}} &= \nabla_{\boldsymbol{\lambda}_{(i)}}\ L_i\left(\boldsymbol{\lambda}_{(i)}(\kappa),\ \mathbf{a}_i(\kappa) :\ \mathbf{a}_{-i}(\kappa)\right) \\ &= \left[\mathbf{V}_{T_i}\left(\mathbf{a}_i^*\right) - \bar{\mathbf{T}}\right]. \end{aligned}$$

The first term in Eqn. (43) is the consensus term, which is a convex combination of $\boldsymbol{\lambda}_{(i)}(\kappa)$ and $\{\boldsymbol{\lambda}_{(n)}(\kappa)\}$:

$$\boldsymbol{\nu}_{\boldsymbol{\lambda}_{(i)}}(\kappa) = \sum_{l \in B_i}\ \omega_{il}(\kappa)\ \boldsymbol{\lambda}_{(l)}(\kappa), \tag{44}$$

which always yields $\boldsymbol{\nu}_{\boldsymbol{\lambda}_{(i)}}(\kappa) \in \Re^m$ as a non-negative vector. The second term is the gradient descent term, which adjusts $\boldsymbol{\lambda}_{(i)}$ in a coordinated fashion with the change in $\mathbf{a}_i$ to converge toward an optimal and feasible solution relative to the local optimization problem. The step-size $s(\kappa) > 0$ can be adjusted to maintain component-wise non-negativity of $\boldsymbol{\lambda}_{(i)}$.

With Assumptions 2, 3 and 4 , it is shown in [47]–[49] that for all $i = 1, \ldots, N_C$, there exist $\bar{\boldsymbol{\lambda}} \succeq \mathbf{0}$ such that $\lim_{\kappa \to \infty} \|\bar{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_{(i)}(\kappa)\| = 0$.

D. Certificate for Optimality

For the unconstrained maximization problem defined by Eqn. (40) for each agent, the optimal primal-dual pair $\left(\mathbf{a}_i^*, \boldsymbol{\lambda}_{(i)}^*\right)$ must satisfy the *KKT* conditions:

$$\nabla V_{I_i}(\mathbf{a}_i^*) + \left[\nabla \mathbf{V}_{T_i}(\mathbf{a}_i^*)\right]^\top \boldsymbol{\lambda}_{(i)}^* = 0, \quad (45)$$

$$\mathbf{V}_{T_i}(\mathbf{a}_i^*) - \bar{\mathbf{T}} \succeq \mathbf{0}, \quad \boldsymbol{\lambda}_{(i)}^* \succeq \mathbf{0}, \quad (46)$$

$$\boldsymbol{\lambda}_{(i)}^{*\top} \left[\mathbf{V}_{T_i}(\mathbf{a}_i^*) - \bar{\mathbf{T}}\right] = 0, \quad (47)$$

which provide a certificate of optimality at each agent.

All cameras optimize in parallel. Camera C_i broadcasts $\mathbf{a}_i^*$ and $\boldsymbol{\lambda}_{(i)}^*$, and new portions of $\mathbf{a}_{-i}$, to its neighbors who propagate them through the network. While C_i is locally optimizing its settings, it is accounting for an updated $\boldsymbol{\lambda}_{(i)}$, and for each target, the prior information $\mathbf{J}^{j-}$ and expected new information based on the currently best settings of all the other cameras $\mathbf{a}_{-i}$.

Optimization stops when either an optimum is achieved, a user-defined stopping condition is met, or the time interval allotted for optimization elapses (see Fig 1). The solution approach described in [32] optimized a weighted combination of tracking and imaging; whether or not an optimum was achieved, there was no guarantee that the tracking specification was achieved. For the approach herein, the *KKT* conditions described in Eqns. (45 - 47) provide a certificate on optimality and feasibility (i.e., satisfaction of the tracking specification). Numeric algorithms to solve the constrained optimization problem defined in Eqn. (39), to which the KKT conditions of Eqns. (45–47) apply, first find a feasible solution, then search within the feasible set for the optimal feasible solution. Thus, when the time interval allotted for optimization elapses, even if the solution is sub-optimal, the solution obtained is guaranteed to be feasible. This results in all targets being tracked to the specified tracking accuracy at all times, while procuring high-res imagery when opportunity arises. After optimization, the cameras physically alter their settings to the optimal values in readiness for upcoming images at t_{k+1}.

Thus, by using the replacement step in Section VIII-B and the Lagrange multiplier update law from Eqn. (43), at each consensus iteration κ, every camera maintains an estimate of the primal-dual pairs of all cameras.

IX. Implementation

This section describes a *Matlab* implementation of the proposed approach. The goal of the simulation is to evaluate the performance of a distributed PTZ camera network using the methods described herein to obtain opportunistic high-res facial imagery of targets moving in a region, while tracking all targets at all times to a specified tracking accuracy.

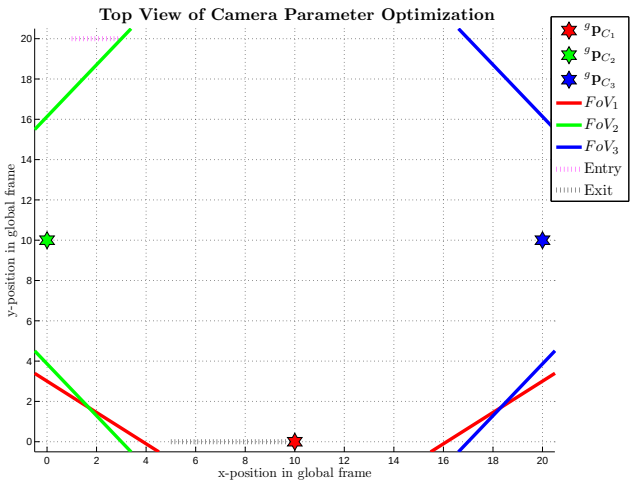


Fig. 3: Top-view of 20×20 surveillance area at $t = 0$, prior to target entry. Camera locations are indicated by colored stars. The camera's FoV boundary is drawn on the ground plane using the same color as its star. The FoV of the camera is the convex area interior of this polygon.

A. Scenario, Setup and Experiment Details

Fig. 3, shows a 400 m^2 area being monitored by $N_C = 3$ calibrated cameras located at $C_1 = [10, 0, 3]^\top$, $C_2 = [0, 10, 3]^\top$, and $C_3 = [20, 10, 3]^\top m$. Camera locations are indicated by colored stars. The boundary of the FoV for each camera is drawn as a wide solid line in a color coordinated with the color of the position marker of the camera. Note that the FoV is the area in the interior of this polygon.

Every target T^j is modeled as a circular disc of negligible height and a radius of 30 cm. All target discs are coplanar to the ground plane. The entrance to the area is located at $y = 20, x \in [1, 3]$ and indicated by the pink hash marks in Fig. 3. Targets enter through the entrance at random times; therefore, the total number of targets in the area is time variant. When a target T^j enters the area, its position coordinates are randomly initialized in $[^g x^j, ^g y^j, 0]^\top$, where $^g x^j \in [1, 3]$ and $^g y^j = 20$. When a new target is detected, the number of targets $N_T(t)$ is increased, and the target state is augmented to the state vector and included in the imaging and tracking value functions. The maximum number of targets permissible in the area was limited such that $0 \leq N_T(t) \leq \overline{N}_T$ where $\overline{N}_T = 10$. To ensure that targets entering the area are detected, the entrance must be constantly monitored. This is achieved by inserting an artificial stationary target at $(2, 20, 0)$ with constant position uncertainty of $2m^2$. Once a target is in the room, its motion is generated using the model in Eqn. (13). The exit to the room is located at $y = 0, x \in [5, 10]$ and indicated by the black hash marks in Fig. 3. If the target trajectory intersects the wall in this region, then the target has exited the room, in which case, the target state is removed from the state vector, excluded from the imaging and tracking value functions, and the number of targets $N_T(t)$ is decreased.

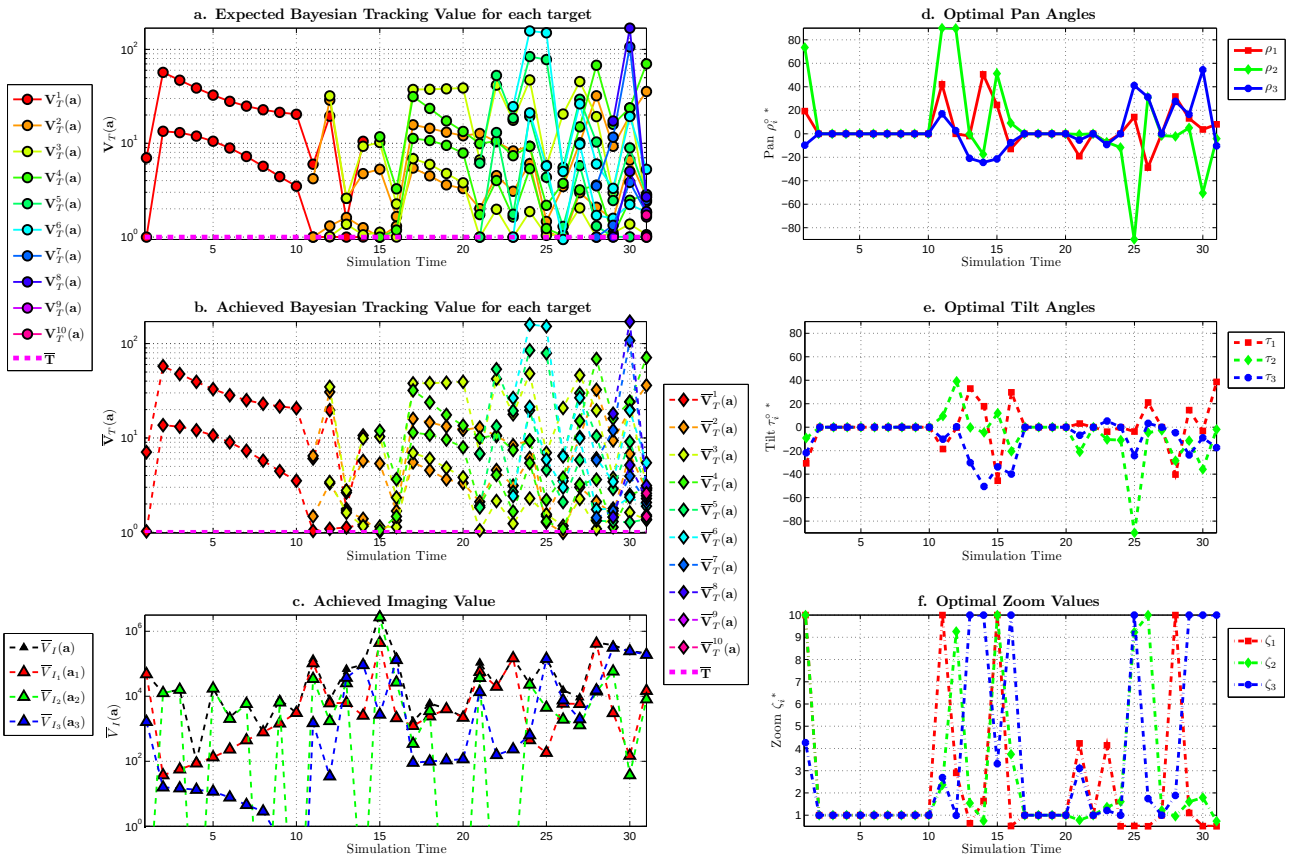


Fig. 4: *Bayesian* Tracking and Imaging Values: Fig. a (top left) shows that the camera network expects to successfully and co-operatively satisfy the tracking constraint $\bar{\mathbf{T}} = 1.0$ m^{-2} for every target, at all times. Fig. b (middle left) shows that the achieved tracking values satisfy the tracking constraint $\bar{\mathbf{T}}$. Fig. c (bottom left), plots the achieved local imaging value $\bar{V}_{I_i}(\mathbf{a}_i)$ and the achieved global imaging value $\bar{V}_I(\mathbf{a})$ (i.e., sum of the local values). Figs. d (top right), e (middle right) and f (bottom right) show the per camera optimal pan angle $\rho_i^{\circ\,*}$, tilt angle $\tau_i^{\circ\,*}$ and zoom $\zeta_i{}^*$ values, respectively.

Remark 7: Note that the target trajectory from Eqn. (13) may intersect a wall. If the point of intersection is the exit, then the target exits the area as described above. If the point of intersection is not the exit, then the target trajectory reflects off the wall.$\diamond$

As discussed in Section VI-C, the measurement model depends on the camera parameters. In addition, while the image processing algorithms may compute the centroid of the feature region in the image plane to subpixel resolution, the covariance matrix used in the state estimation routine must account for the the uncertainty in the computed centroid relative to the "actual target centroid." Let $n_i^j(\mathbf{a}_i)$ represent the area occupied by T^j's image on C_i's image plane measured in sq. pixels. For this simulation, the estimation routine models the covariance of the measurement of T^j by C_i as

$$\mathbf{C}_i^j(\mathbf{a}_i) = \begin{bmatrix} \frac{n_i^j(\mathbf{a}_i)}{p_i}\sigma_x^2 & 0 \\ 0 & \frac{n_i^j(\mathbf{a}_i)}{p_i}\sigma_y^2 \end{bmatrix}, \qquad (48)$$

where p_i is the pixel resolution of C_i's image plane (in sq. pixels) and σ_x^2 and σ_y^2 (in sq. pixels) are positive constants. For this simulation, each camera C_i was set to an image resolution p_i of 800×600 sq. pixels, with $\sigma_x = \sigma_y = 5$ pixels.

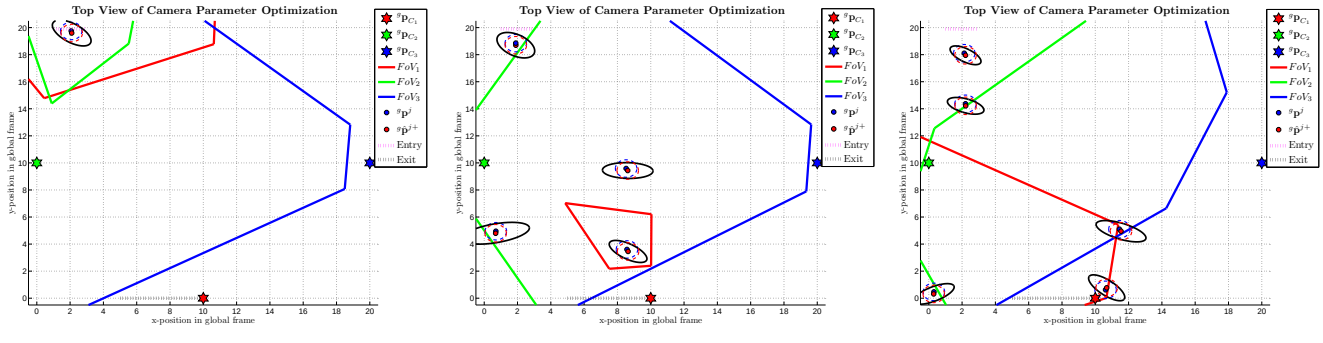
The results in Figs. 4 - 6 correspond to a 31 sec. simulation. All cameras image at a frequency of 1 Hz, with the first images obtained at time $t = 1$ second.

All cameras optimize simultaneously, using an interior-point method [51]. The tracking constraint in Eqn. (37) uses $\bar{\mathbf{T}} = 1.0$ m^{-2}. C_i receives camera parameters $\mathbf{a}_{-i}^*$ through its neighbors, and uses its current parameters $\mathbf{a}_i$ to implement the method described in Section VIII.

Define $\bar{V}_{I_i}^j$ to be the imaging value *achieved* by camera C_i for imaging T^j. The weight $w_i^j(t)$, in Eqn. (27), is defined as the continuously differentiable and bounded function:

$$w_i^j(t) = \sigma_d\left(d^j(t)\right)\ \sigma_v\left(\bar{V}^j, V_{I_i}^j(t)\right), \qquad (49)$$

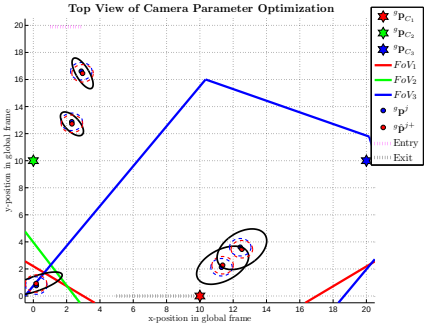
where $\sigma_d = 1 + \frac{1}{1+\exp[l_d d^j(t)]}$, and $\sigma_v = \frac{1}{1+\exp[l_v(\bar{V}^j - V_{I_i}^j(t))]}$. The symbol $\bar{V}^j = \max_{\tau<t,\ i\in[1,N_c]}\left(\bar{V}_{I_i}^j(\tau)\right)$ is the best image quality for target T^j for any camera and any prior image. The symbol $d^j(t)$ is the distant between T^j's estimated position at time t and the exit. With these definitions $\sigma_d(t) \in [1, 2]$ and $\sigma_v(t) \in [0, 1]$. This definition of $w_i^j(t)$ gives higher value (i.e., emphasizes) those targets nearest to the exit and those targets for which the value of the next image is expected to improve the most relative to prior imagery.



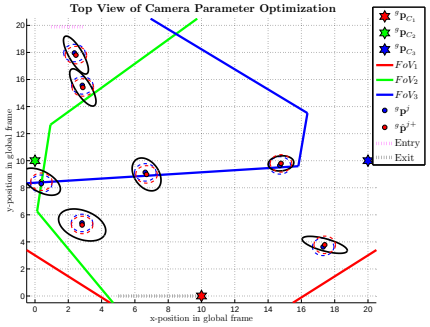
(a) Optimized FoVs at $t = 1$.

(b) Optimized FoVs at $t = 21$.

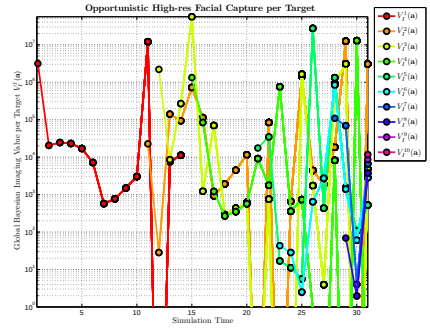
(c) Optimized FoVs at $t = 24$.



(d) Optimized FoVs at $t = 25$.



(e) Optimized FoVs at $t = 30$.



(f) Opportunistic Hi-res Capture.

Fig. 5: Top-view of Opportunistic High-res Facial Image Capture: Figs. a (top left), b (top center), c (top right), d (bottom left), and e (bottom center) show the optimized FoVs at times of opportunity: $t = 1$, $t = 21$, $t = 24$, $t = 25$ and $t = 30$, respectively. Each figure shows the optimized FoVs of the cameras after feasible optimal solutions are achieved. Fig. f (bottom right) shows the expected per target imaging value $V_I^j(\mathbf{a}(t))$ from Eqn. (50).

B. Single Trial Results

For this simulation, targets T^1 to T^{10} entered the area at times 0.1, 10.2, 11.1, 14.2, 20.4, 22.9, 27.2, 28.8, 30.1, and 30.4 seconds, respectively. Target T^1 left the area at time 14.4 seconds. No other targets left the area. When T^j enters, a camera monitoring the entrance images it, detects the new target and augments it to its state vector. Other cameras add the new target to their state vector as they receive the new target information at the state estimation stage.

1) Bayesian Imaging and Tracking Performance: Cameras maximize their local Lagrangians $L_i(\boldsymbol{\lambda}^*_{(i)}, \mathbf{a}^*_i)$ to satisfy the tracking spec and maximize their local Bayesian imaging values. Fig. 4a shows that the expected Bayesian tracking value $\mathbf{V}_T^j(\mathbf{a}^*)$ is greater than the tracking spec, at all times; therefore, all primal-dual solutions $(\mathbf{a}^*_i, \boldsymbol{\lambda}^*_{(i)})$ obtained through local optimization are *expected* to be feasible at all imaging instants. Because the solutions are strictly feasible, using Eqn. (47), it is trivial to prove that the dual optimal Lagrange multiplier vectors for all cameras are $\boldsymbol{\lambda}^*_{(i)}(t) = \bar{\boldsymbol{\lambda}} = \mathbf{0} \in \Re^m$, where $m = 2N_T(t)$.

The proposed approach utilizes predicted target motion based on state estimates from the last imaging time. Estimation error or unexpected maneuvers by targets, such as a simulated target reflecting of a wall, can lead to a drop in the accuracy actually *achieved*. Fig. 4b shows the tracking value $\bar{\mathbf{V}}_T^j(\mathbf{a}^*)$ actually achieved by the network. Various instances of differences between the expected and achieved accuracy can be observed through the simulation time. Since the target motion is a random process, there is no deterministic guarantee that the achieved accuracy meets the specification.

Fig. 4c shows the achieved imaging values of each camera and of the network of cameras. The peak values occur at those opportunistic times at which the cameras procure high-res facial images of targets, while the tracking constraints on all targets are satisfied. A high value for $\bar{V}_{I_i}(\mathbf{a}^*_i)$ indicates a high-res facial capture by camera C_i. Given the target trajectories of this simulation, all cameras availed opportunities for high-res image capture throughout simulation time.

Figs. 4d - 4f show the per camera optimized PTZ values versus time. Top-views of the camera FoVs for a selection of high-res imaging opportunities is shown in Fig. 5.

Figs. 5a, 5b, 5c, 5d, and 5e show the post-optimization FoVs of the cameras for time-steps $t = 1$, $t = 21$, $t = 24$, $t = 25$, and $t = 30$, respectively. The prior estimate of the position of the centroid of each target is marked by a red dot. The actual position of the centroid of each target is marked by a blue dot. A red dashed curve is drawn to indicate the surface area occupied by a target on the ground plane, relative to the target's estimated centroid position. Similarly, a blue dashed curve indicates the surface area occupied by a target, relative to the actual target centroid position. The posterior $1-\sigma$ position error ellipse corresponding to the estimated position of each target is drawn as a wide black curve.

Target T^1 enters at time $t = 0.1$. Cameras collaboratively image T^1 at time $t = 1$, where C_2 images T^1 with the highest imaging value among all cameras (see Fig. 4c). Similarly,

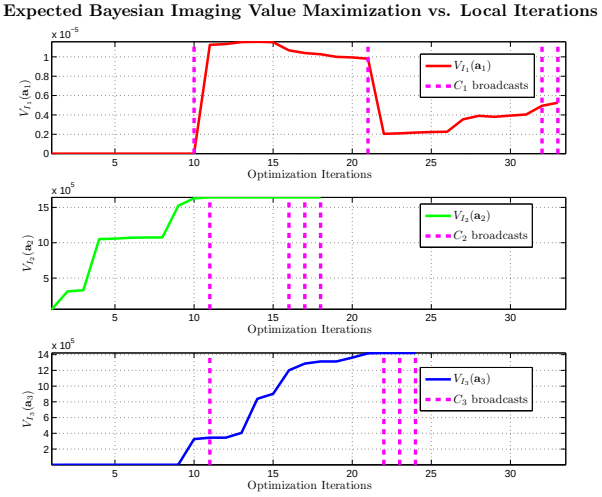


Fig. 6: Optimization: The maximization of $V_{I_i}(\mathbf{a}_i)$ versus the number of local iterations at time-instant $t = 25$. The vertical pink dashed lines indicate local iterations κ_i at which C_i broadcast parameters as described in Section VIII-B. After the cameras have collaboratively found a feasible PTZ configuration, C_2 and C_3 capitalize on the target configuration to obtain images expected to have very high values.

camera C_1 obtains an opportunistic high-res image of T^2 at time $t = 21$. Note that in all cases, the entrance and all targets are within at least one FoV.

Fig. 5f plots the time history of the expected imaging value per target acquired by all cameras in the network:

$$V_I^j\big(\mathbf{a}(t)\big) = \sum_i w_i^j(t)\; E\left\langle V_{I_i}^j\left(\mathbf{a}_i(t),\; {}^g\mathbf{p}^j(t)\right)\right\rangle, \qquad (50)$$

where $V_{I_i}^j(\mathbf{a}_i, {}^g\mathbf{p}^j)$ is defined in eqn. (26). The number of curves is different at each time because the number of targets is time varying. A high-res image capture of T^j by any camera C_i is indicated by a spike in the global *Bayesian* imaging Value function $V_I^j\big(\mathbf{a}(t)\big)$. The figure shows that for this simulation run, the camera network obtained at least one high-res facial image of each target in the area, at times-of-opportunity distributed throughout the time period of the simulation. Combining the information from this figure with that from Fig. 4c, we see for example that as T^2 moves through the room, at various times, different cameras have opportunities to image it.

Fig. 6 is an example, using time-instant $t = 25$, of the optimization process that each camera performs prior to each imaging instance. All cameras simultaneously perform a few optimization iterations, then broadcast their (approximate) primal-dual solutions, update their local estimates of $\mathbf{a}^*$ using the sub-vectors received from their neighbors and resume the optimization process. The broadcast instances are indicated by the pink dashed vertical lines. This process repeats till an optimum is reached or time expires.

Remark 8: We start each optimization iteration with a wide FoV. This choice of initial condition facilitates the search for a feasible solution. This is similar to using a metaheuristic [52], [53] to aid computation of a feasible solution. $\diamond$

Remark 9: Camera FoV's alter significantly between successive time instants to achieve high-res imagery and satisfy the tracking spec. One such example can be seen in Figs. 5c and 5d as all the camera FoVs change considerably from time $t = 24$ to those at $t = 25$ (also see Figs. 4d-4f). Such rapid motion can hamper image quality due to motion blurring and may also cause mechanical wear. Model Predictive Control [54] based approaches to enforce constraints on the PTZ parameters are interesting for future research. $\diamond$

Remark 10: As seen in Fig. 5f, in spite of the formulation of Eqn. (49), it is still possible that the camera will attempt to acquire images of targets with lower image value than was previously obtained. There are at least two explanations. First, this can occur inadvertently because a previously imaged target is sometimes visible in the FoVs of cameras that have been optimized for imaging other targets. Second, as long as the expected imaging value V_I^j is finite, which it always is, the optimization still receives some value for new imagery, even if it is not of higher quality than previous imagery. $\diamond$

C. Multi-Trial Performance Analysis

This section provides an analysis of the performance of the proposed PTZ camera network approach using data from $N = 100$ *Matlab* simulation runs. Across all simulation runs, target trajectories and target times of entry were independent, with the target times of entry designed such that target T^j always entered before target T^{j+1}. To make results comparable, all other parameters (e.g. camera locations, image resolution, pixel noise, area entrances and exits, etc.) were defined to be the same for all simulation runs, as defined in Section IX-B.

For a dynamic PTZ camera network, define $\bar{V}_D^j(n)$ to be the maximum global imaging value achieved for target T^j during simulation run n. Similarly, for a constant (static) PTZ camera network, let $\bar{V}_S^j(n)$ be the maximum global imaging value achieved for target T^j during simulation run n. Define a *performance ratio* $\bar{V}_B^j(n)$ as

$$\bar{V}_B^j(n) = \frac{\bar{V}_D^j(n)}{\bar{V}_S^j(n)}, \qquad (51)$$

where $\bar{V}_B^j(n)$ provides a measure of the relative gain in imaging value achieved by utilizing a dynamic PTZ configuration rather than a static PTZ configuration.

Fig. 7 shows the distribution (histogram) of the performance ratio $\bar{V}_B^j$ over $N = 100$ simulation runs. The network of dynamic PTZ cameras consistently outperforms the static camera configuration by procuring images of higher quality. Fig. 7 uses a semilog horizontal axis with a maximum of 10^6. The performance improvement ratio actually varies from 1 to 10^{12}. Such enhanced image quality is better suited for image analysis and scene understanding.

Fig. 8 displays the histogram of the per target achieved imaging value, which is denoted by $\bar{V}_I^j(\mathbf{a})$. The number of opportunistic high-res images obtained for target T^1 is greater than those obtained for T^2, and so on. There are at least two explanations. First, T^j always enters before target T^{j+1}; therefore, the cameras likely have more opportunities to image T^j at a higher resolution than T^{j+1}. Second, the difficulty in acquiring high-res images increases as the number of targets

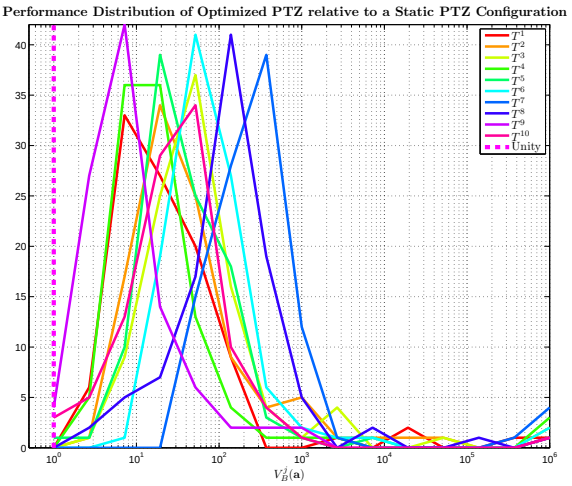


Fig. 7: Performance Improvement Distribution: Imaging performance improvement ratio of dynamic PTZ camera configuration relative to a static PTZ configuration. Each colored line corresponds to a distinct target, showing the distribution of the per target performance ratio $\bar{V}_B^j$ (see Eqn. (51)) over 100 simulation runs. The range of $\bar{V}_B^j$ values is plotted as bins on the horizontal axis. The dynamic PTZ config. significantly outperforms the static PTZ config.

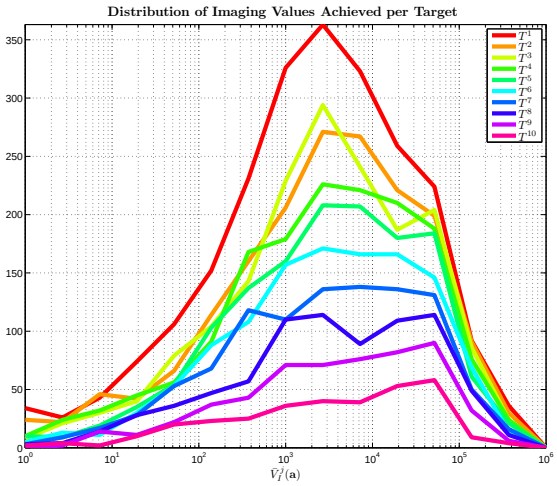


Fig. 8: Opportunistic Imaging Distribution: Fig. shows the distribution of the per target achieved image value over $N = 100$ simulation runs. Each colored line corresponds to a distinct target, showing the histogram of values.

in the area increases, due to the increase in the number of feasibility constraints (see Eqns.(33), (37), and (39)).

Remark 11: The camera positions across all simulation runs (static and dynamic) were left unchanged. The locations were selected so that, for the static configuration, all locations within the entire area were within the FoV of at least one camera. Altering the positions of entrances, exits, or the static parameter cameras may provide different performance than the static configuration used herein. ⋄

This section demonstrates that the proposed method causes the cameras to cooperate to ensure that all targets are expected to be tracked to an accuracy better than $\bar{\mathbf{T}}$, and that high-res target images are obtained at times-of-opportunity implicitly defined by the feasibility constraints. The statistical analysis provides a measure of the increase in imaging performance obtained while using the proposed method.

D. Discussion of Implementation Issues

For convex problems, the proposed distributed optimization methodology would converge to the unique global optimum for each imaging time instant. As with many practical applications, visual sensing problems such as the one considered herein are inherently non-convex (refer to Fig. 2), and thus the solution obtained may only be locally optimal. The large variable space makes design of an exhaustive search impractical.

In addition to being non-convex, the local imaging value and the constraint functions are nonlinear. Our implementation used the Matlab function '*fmincon*', which is offered as part of the Optimization Toolbox and is designed to solve nonlinear optimization problems with nonlinear constraints.

To facilitate the search for a feasible solution, at the start of each optimization interval, we initiated each camera using the optimal pan and tilt values from the end of the prior optimization interval, but reset the zoom parameter to its minimum value (i.e., widest FoV). The wide FoV initialization was preferred as it enhances feasibility and convexity of the value function, see Fig. 2 and Remark 8. This initialization method ensures that all agents begin the optimization process from the same value of $\mathbf{a}$. This initialization worked well in the sense that a feasible solution was found for every imaging instant of every trial; nonetheless, alternative initialization and relaxation techniques, could be investigated.

Finally, it is important to note that if the initial parameters of the cameras were not identical and were distributed about a saddle point of the value function, such that some initial parameter vectors were in the domains of attraction (DOA) of different local optima, then different agents could conceivably converge toward different locally optimal points prior to communicating their new settings. After the communication, there would be no guarantee that the camera parameter settings of different agents are all within the DOA of the same locally optimal point. We ensured that all agents start with the same value for $\mathbf{a}$. This issue and methods to ensure convergence to the global optimum are interesting areas for future research.

X. Conclusion and Future Work

This article addressed the design of a method for a distributed network of smart imaging sensors to collaboratively track all targets to a specified accuracy while also acquiring high resolution images at times-of-opportunity. The solution uses a *Bayesian* framework that trades off higher imaging value versus increased risk of the target not being in the field-of-view. The Bayesian imaging value depends on the target's expected position, direction-of-motion, image resolution and camera relative pose. The approach includes a dynamic target weighting scheme. In the example, we demonstrate the utility of this feature in two ways. First, the importance of a target increases as the target approaches the exit, to help ensure that all targets are imaged at least once. Second, the weight at any time instant for a target is dependent on the image quality previously acquired for that target; therefore, subsequent images

of each target receive little value unless a better quality image is expected to be acquired.

The method proposed herein allows all agents to optimize simultaneously. The global optimization problem is formulated as a *potential game* with the global objective decoupled into smaller local problems with aligned local objectives. A *Lagrangian* consensus algorithm is used to perform distributed, co-operative and simultaneous optimization across all cameras in the network. This formulation enables use of existing convergence proofs from the game theory literature. Convergence of the Lagrange multiplier vector is achieved by consensus methods and of the camera PTZ parameters is achieved by a modified flooding algorithm. Future research could explore probabilistic communication schemes [55] to decrease communication loading and latency issues.

It is possible to design alternative optimization methods that use a combination of parallel and sequential processing. Graph partitioning [56] is a branch of optimization that decouples an existing communication graph into smaller subgraphs, often dynamically, subject to problem constraints. A property implicit to a visual sensing application is that the communication graph and the vision graph [15] are often different. This occurs since every camera in the network may not obtain a measurement of every target in the area, which often results in cameras [18] 'naive' with respect to certain targets. Using vision graph discovery methods to design a set of constraints, the existing communication graph G_c could be decoupled into smaller subgraphs G^j, on the basis of measurements on target T^j. Agents that are nodes of subgraph G^j may optimize sequentially, while subgraph G^j optimizes in parallel with respect to other subgraphs in the network. This topic is largely unexplored and could provide beneficial results, especially in scenarios where a network of cameras is assigned to survey a large area with camera visibility constraints.

The method proposed herein can be extended to locally convex discontinuous functions using the subgradient-based distributed constrained optimization approaches described in [47], [48]. Though preliminary work on modifying the approach therein for application to a visual sensing problem was done in [57], further research on the method is required.

Extension of the proposed method to optimize the PTZ settings over a time horizon would enhance continuity of the PTZ settings, see Remark 9. This would reduce mechanical wear and enhance image quality which facilitates image analysis. Design of alternative imaging value functions and constraints that account for occlusion or are dependent on advanced image processing techniques to aid target gait [58], gesture [59], and activity [60] recognition is another interesting topic. As discussed in Remark 2, various alternative designs are possible for $M > 1$. For example, a sequence of high rate images could be taken with various focal lengths allowing a complete reformulation of the trade-off between risk and image quality. Implementation on a Camera Network test-bed is also of interest and in process.

ACKNOWLEDGEMENT

The authors thank the Office of Naval Research for supporting the research under award no.: N000140910666 for the project titled 'Distributed Dynamic Scene Analysis in a Self-Configuring Multimodal Sensor Network'.

REFERENCES

[1] Ballard, D. H., "Animate Vision," in *Art. Intell.*, vol. 48, no. 1, Feb. 1991, pp. 57–86.

[2] Zhao, J. and Cheung, S. -C. and Nguyen, T., "Optimal Camera Network Configurations for Visual Tagging," in *J. of Sel. Top. in Sig. Proc.*, vol. 2, no. 4, Aug. 2008, pp. 464–479.

[3] Erdem, M. and Sclaroff, S., "Automated Camera Layout to Satisfy Task-specific and Floor Plan-specific Coverage Requirements," in *Comp. Vis. and Img. Und.*, vol. 103, no. 3, Sep. 2006, pp. 156–169.

[4] Qureshi, F. Z. and Terzopoulos, D., "Planning ahead for PTZ camera assignment and handoff," in *Int. Conf. on Dist. Smart Cam.*, Sep. 2009, pp. 1–8.

[5] Moon, T. K., "The Expectation-Maximization Algorithm," in *Sig. Proc. Mag.*, Nov. 1996, pp. 47–60.

[6] Piciarelli, C. and Micheloni, C. and Foresti, G. L., "PTZ Camera Network Reconfiguration," in *Int. Conf. on Dist. Smart Cam.*, Sep. 2009, pp. 1–7.

[7] Alahi, A. and Marimon, D. and Bierlaire, M. and Kunt, M., "A Master-Slave Approach for Object Detection and Matching with Fixed and Mobile Cameras," in *Int. Conf. on Img. Proc.*, Oct. 2008, pp. 1712–1715.

[8] Khan, S. and Javed, O. and Rasheed, Z. and Shah, M., "Human Tracking in Multiple Cameras," in *Int. Conf. on Comp. Vis.*, vol. 1, Jul. 2001, pp. 331–336.

[9] Li, Y. and Bhanu, B., "Utility-based Dynamic Camera Assignment and Hand-off in a Video Network," in *Int. Conf. on Dist. Smart Cam.*, Sep. 2008, pp. 1–9.

[10] Stancil, B. A. and Zhang, C. and Chen, T., "Active Multicamera Networks: From Rendering to Surveillance," in *J. of Sel. Top. in Sig. Proc.*, vol. 2, no. 4, Aug. 2008, pp. 597–605.

[11] Du, W. and Piater, J., "Multi-camera People Tracking by Collaborative Particle Filters and Principal Axis-based Integration," in *Asian Conf. on Comp. Vis.*, Nov. 2007, pp. 365–374.

[12] Khan, S. and Shah, M., "A Multiview Approach to Tracking People in Crowded Scenes Using a Planar Homography Constraint," in *ECCV*, May 2006, pp. 133–146.

[13] J. K. Aggarwal and Q. Cai, "Human Motion Analysis: A Review," in *Comp. Vis. and Img. Und.*, Sep. 1999, pp. 428–440.

[14] Tron, R. and Vidal, R., "Distributed Computer Vision Algorithms," in *Sig. Proc. Mag.*, vol. 28, no. 3, May 2011, pp. 32–45.

[15] ——, "Distributed Computer Vision Algorithms through Distributed Averaging," in *CVPR*, Jun. 2011, pp. 57–63.

[16] Taj, M. and Cavallaro, A., "Distributed and Decentralized Multicamera Tracking," in *Sig. Proc. Mag.*, vol. 28, no. 3, May 2011, pp. 46–58.

[17] Song, B. and Ding, C. and Kamal, A. T. and Farrell, J. A. and Roy-Chowdhury, A. K., "Distributed Camera Networks," in *Sig. Proc. Mag.*, May. 2011, pp. 20–31.

[18] Kamal, A. T. and Farrell, J. A. and Roy-Chowdhury, A. K., "Information Wieghted Consensus," in *IEEE CDC*, Dec. 2012.

[19] Stillman, S. and Tanawongsuwan, T., "Tracking Multiple People with Multiple Cameras," in *Int. Conf. on Aud. and Vid. based Biometric Person Auth.*, Mar. 1999, pp. 1–6.

[20] Markis, D. and Ellis, T. and Black, J., "Bridging the Gap Between Cameras," in *CVPR*, vol. 2, Jul. 2004, pp. 205–210.

[21] Tieu, K. and Dalley, G. and Grimson, W. E. L., "Inference of Non-Overlapping Camera Network Topology by Measuring Statistical Dependence," in *Int. Conf. on Comp. Vis.*, vol. 2, Oct. 2005, pp. 1842–1849.

[22] Ren, W. and Beard, R. W., "Consensus Seeking in Multiagent Systems under Dynamically Changing Interaction Topologies," in *IEEE T-AC*, vol. 50, no. 5, May 2005, pp. 655–661.

[23] Xiao, F. and Wang, L., "Asynchronous Consensus in Continuous-Time Multi-Agent Systems With Switching Topology and Time-Varying Delays," in *IEEE T-AC*, vol. 53, no. 8, Sep. 2008, pp. 1804–1816.

[24] Hatanaka, T. and Fujita, M. and Bullo, F., "Vision-based cooperative estimation via multi-agent optimization," in *IEEE CDC*, Dec. 2010, pp. 2492–2497.

[25] Ibuki, T. and Hatanaka, T. and Fujita, M., "Passivity-based Visual Pose Regulation for a Moving Target Object in Three Dimensions: Structure Design and Convergence Analysis," in *IEEE CDC*, Dec. 2012, pp. 5655–5660.

[26] Valera, M. and Velastin, S. A., "Intelligent Distributed Surveillance Systems: A Review," in *P. on Vis., Img. and Sig Proc.*, vol. 152, no. 2, Apr. 2005, pp. 192–204.

[27] Remagnino, P. and Shihab, A. I. and Jones, G. A., "Distributed Intelligence for Multi-Camera Visual Surveillance," in *Pat. Recog.*, vol. 37, no. 4, Apr. 2004, pp. 675–689.

[28] Moreau, L., "Stability of Multiagent Systems with Time-dependent Communication Links," in *IEEE T-AC*, vol. 50, no. 2, Feb. 2005, pp. 169–182.
[29] Vincent, T. L., "Game Theory as a Design Tool," in *J. of Mech., Trans. and Auto. in Des.*, vol. 105, no. 2, Jun. 1983, pp. 165–170.
[30] Arslan, G. and Marden, J. and Shamma, J., "Autonomous Vehicle-Target Assignment: A Game-Theoretical Formulation," in *J. of Dyn. Sys., Meas. and Cont.*, vol. 129, no. 5, Sep. 2007, pp. 584–596.
[31] Soto, C. and Song, B. and Roy-Chowdhury, A. K., "Distributed Multi-target Tracking in a Self-configuring Camera Network," in *CVPR*, Jun. 2009, pp. 1486–1493.
[32] Morye, A. A. and Ding, C. and Song, B. and Roy-Chowdhury, A. K. and Farrell, J. A., "Optimized Imaging and Target Tracking within a Distributed Camera Network," in *ACC*, Jun. 2011, pp. 474–480.
[33] Song, B. and Ding, C. and Morye, A. A. and Farrell, J. A. and Roy-Chowdhury, A. K., "Collaborative Sensing in a Distributed PTZ Camera Network," in *IEEE T-IP*, Jul. 2012, pp. 3282–3295.
[34] Morye, A. A. and Ding, C. and Roy-Chowdhury, A. K. and Farrell, J. A., "Constrained Optimization for Opportunistic Distributed Visual Sensing," in *ACC*, Jun. 2013, pp. ??–??
[35] Macready, W. and Wolpert, D., "Distributed Constrained Optimization," in *Int. Conf. on Complex Sys.*, vol. Bar-Yam (Ed.). Perseus books, 2004, pp. 1–12.
[36] Li, N. and Marden, J. R., "Designing Games for Distributed Optimization," in *IEEE CDC*, Dec. 2011, pp. 2434–2440.
[37] Hjelmas, E. and Low, B. K., "Face Detection: A Survey," in *Comp. Vis. and Img. Und.*, vol. 83, no. 3, Sep. 2001, pp. 236–274.
[38] Zhao, W. and Chellappa, R. and Phillips, P. J. and Rosenfeld, A., "Face Recognition: A Literature Survey," in *ACM Comp. Sur.*, vol. 35, no. 4, Dec. 2003, pp. 399–458.
[39] Choate, A. G., "Telecentric, Parfocal, Multiple Magnification Optical System for Video-inspection Apparatus," in *US Patent*, no. 5668665, Sep. 1997, pp. 1–9.
[40] Song, B. and Soto, C. and Roy-Chowdhury, A. K. and Farrell, J. A., "Decentralized Camera Network Control using Game Theory," in *Int. Conf. on Dist. Smart Cam.*, Sep. 2008, pp. 1–8.
[41] Song, B. and Roy-Chowdhury, A. K., "Stochastic Adaptive Tracking in a Camera Network," in *Int. Conf. on Comp. Vis.*, Oct. 2007, pp. 1–8.
[42] Olfati-Saber, R., "Kalman-Consensus Filter: Optimality, Stability, and Performance," in *IEEE CDC*, Dec. 2009, pp. 7036–7042.
[43] Boyd, S. and Vandenberghe, L., "Convex Optimization." Cambridge Uni. Press, 2004, pp. 241–246.
[44] Monderer, D. and Shapley, L. S., "Potential Games," in *Games and Econ. Behavior*, vol. 14, no. 0044, 1996, pp. 124–143.
[45] Farrell, J. A., "Aided Nav.: GPS w/ High Rate Sensors." Mcgraw Hill, 2008, pp. 72–75.
[46] Trucco, E. and Verri, A., "Introductory Techniques for 3-D Computer Vision." Prentice Hall, 1998, pp. 26–27.
[47] Nedic, A. and Ozdaglar, A. and Parrilo, P. A., "Constrained Consensus and Optimization in Multi-Agent Networks," in *IEEE T-AC*, vol. 55, Apr. 2010, pp. 922–938.
[48] Zhu, M. and Martinez, S., "On Distributed Optimization under Inequality Constraints via Lagrangian Primal-Dual Methods," in *ACC*, Jul. 2010, pp. 4863–4868.
[49] Olfati-Saber, R. and Fax, J. A. and Murray, R. M., "Consensus and Cooperation in Networked Multi-Agent Systems," in *Proc. IEEE*, vol. 95, no. 1, Jan. 2007, pp. 215–233.
[50] Tanenbaum, A. S. and Wetherall, D. J., "Computer Networks (5th Edition)." Pearson Edu. Inc., 2010, pp. 368–370.
[51] Boyd, S. and Vandenberghe, L., "Convex Optimization." Cambridge Uni. Press, 2004, pp. 609–615.
[52] Ibaraki, T. and Nonobe, K. and Yagiura F., "Metaheuristics: Progress as Real Problem Solvers." Springer Sci. + Busi. Med., 2005, pp. 1–28.
[53] van Laarhoven, P. J. and Aarts, E. H., "Simulated Annealing: Theory and Applications." Kluver Acad. Pub., 1987, pp. 1–6.
[54] Garcia, C. E. and Prett, D. M. and Morari, M., "Model Predictive Control: Theory and Practice - A survey," in *Automatica*, vol. 25, no. 3, May. 1989, pp. 335 – 348.
[55] Sasson, Y. and Cavin, D. and Schiper, A., "Probabilistic broadcast for flooding in wireless mobile ad hoc networks," in *Wireless Comm. and Netw.*, vol. 2, Mar. 2003, pp. 1124–1130.
[56] Lee, L. and Grimson, W. E. L., "Continuous Graph Partitioning for Camera Network Surveillance," in *Wrk. Dist. Est. and Cont. in Netw. Sys.*, vol. 3, no. 1, Sep. 2012, pp. 228–233.
[57] Morye, A. A. and Ding, C. and Roy-Chowdhury, A. K. and Farrell, J. A., "Cooperative, Opportunistic Imaging within a Bayesian Distributed Constrained Optimization Framework," in *BCOE Res.*, Sep. 2012.
[58] Lee, L. and Grimson, W. E. L., "Gait Analysis for Recognition and Classification," in *Int. Conf. on n Auto. Face and Gesture Recog.*, May 2002, pp. 148–155.
[59] Wu, Y. and Huang, T. S., "Vision-Based Gesture Recognition: A Review." Springer-Verlag Berlin Heidelberg, 1999, pp. 103–115.
[60] Bodor R. and Jackson B. and Papanikolopoulos N., "Vision-based Human Tracking and Activity Recognition," in *Medit. Conf. on Cont. and Auto.*, Jun. 2003, pp. 18–20.

Akshay A. Morye received the Diploma in Industrial Electronics from the Maharashtra State Board of Technical Education, India, the B.E. degree in Electronics Engineering from the University of Mumbai, India, and the M.S degree in Electrical Engineering from the University of California, Riverside, where he is pursuing a Ph.D. degree in Electrical Engineering. His broad research interests include Cooperative Control, Advanced Robotics, Distributed Optimization, Autonomous Systems, and Computer Vision.

Chong Ding received the B.S. degree in Computer Science from the University of California, Riverside, in 2008, where he is currently pursuing the Ph.D. degree in the Department of Computer Science. His main research interests include Intelligent Camera Networks, Wide-area Scene Analysis, and Distributed and Real-time Systems.

Amit K. Roy-Chowdhury received the Bachelors degree in Electrical Engineering from Jadavpur University, Calcutta, India, the Masters degree in systems science and automation from the Indian Institute of Science, Bangalore, India, and the Ph.D. degree in Electrical Engineering from the University of Maryland, College Park. He is a Professor of Electrical Engineering at U. of California, Riverside. His research interests include image processing and analysis, computer vision, and video communications and statistical methods for signal analysis. His current research projects include intelligent camera networks, wide-area scene analysis, motion analysis in video, activity recognition and search, video-based biometrics (face and gait), biological video analysis, and distributed video compression. He is coauthor of "The Acquisition and Analysis of Videos over Wide Areas." He is the editor of the book Distributed Video Sensor Networks. He has been on the organizing and program committees of multiple conferences and serves on the editorial boards of a number of journal.

Jay A. Farrell received B.S. degrees in physics and electrical engineering from Iowa State University, and M.S. and Ph.D. degrees in electrical engineering from the University of Notre Dame. From '89-'94 he was at Charles Stark Draper Lab receiving the Engineering Vice President's Best Technical Publication Award in 1990, and Recognition Awards for Outstanding Performance and Achievement in 1991 and 1993. He is a Professor and two time Chair of the Dept. of Electrical Engineering at U. of California, Riverside. He has served on the IEEE Control Systems Society (CSS) Board of Governors for two terms ('03-'06, '12-'14), as Vice Pres. of Fin. ('05-'06), as Vice Pres. of Tech. Act. ('07-'08), as Gen. Chair of IEEE CDC 2012, and will be Pres. in 2014. He is a Fellow of the IEEE, a Fellow of AAAS, a Dist. Member of IEEE CSS, and was recognized as a GNSS Leader to Watch by GPS World Magazine in 2009. He is author of over 200 technical publications and (co-)author of three books.