

Prioritized target tracking with active collaborative cameras

Yiming Wang, Andrea Cavallaro

Centre for Intelligent Sensing, Queen Mary University of London

Email: {yiming.wang, a.cavallaro}@qmul.ac.uk

Abstract

Mobile cameras on robotic platforms can support fixed multi-camera installations to improve coverage and target localization accuracy. We propose a novel collaborative framework for prioritized target tracking that complements static cameras with mobile cameras, which track targets on demand. Upon receiving a request from static cameras, a mobile camera selects (or switches to) a target to track using a local selection criterion that accounts for target priority, view quality and energy consumption. Mobile cameras use a receding horizon scheme to minimize tracking uncertainty as well as energy consumption when planning their path. We validate the proposed framework in simulated realistic scenarios and show that it improves tracking accuracy and target observation time with reduced energy consumption compared to a framework with only static cameras and compared to a state-of-the-art motion strategy.

1. Introduction

Collaboration between static and mobile sensors has been employed for sensor placement [16] and path planning to extend the coverage for target detection [11]. Moreover, mobile cameras (i.e. autonomous robotic platforms equipped with cameras) can help static cameras by actively tracking targets to extend the duration of the observation and to improve localization accuracy [20].

The problem of assigning multiple robots to track moving targets can be decomposed in two tasks, namely robot-task assignment and motion planning. Cooperative Multi-robot Observation of Multiple Moving Targets (CMOMMT) uses local force vectors to maximize the average observation time of targets [9, 15]. The vectors are obtained from predefined distance-based functions and targets are generally treated equally. To enable a collaborative motion strategy, each robot shares its scene knowledge with all the other robots [2, 9, 15]. However, sharing local knowledge globally may be unrealistic when robots with limited communication range are far apart. Moreover, depending on the specific context, targets may have

application-related priorities. For example, some targets with higher priority should be tracked first, with higher accuracy or for longer.

In addition to the above, while *energy efficiency* is considered when assigning robots to a set of goal positions [17] or when planning paths towards goal positions [12], it has not yet been taken into account for motion planning when assigning mobile sensors to follow targets [20].

In this paper, we propose a collaborative framework for prioritized target tracking with both static and mobile cameras. Each static camera detects targets and dynamically joins tracking groups to *fuse* neighborhood-related and target-related information. Each mobile camera moves on demand after accepting a request from static cameras. Mobile cameras *select* a target to track, *plan* their motion for active tracking and *fuse* their target state estimate with that of static cameras. We propose a target selection strategy for mobile cameras based on local knowledge on the scene dynamics received from static cameras. Moreover, we define (i) an energy-efficient robot-target assignment that combines a distance-based criterion with target priority information; and (ii) an energy-efficient motion strategy that maximizes view quality and minimizes energy consumption based on a realistic energy model.

2. Problem formulation

Let $\mathbf{C} = \mathbf{C}^S \cup \mathbf{C}^M$ be a mixed camera network, where \mathbf{C}^S is a set of static cameras and \mathbf{C}^M is a set of mobile cameras. Each camera $c_i \in \mathbf{C}$ has a directional field of view (FoV) defined by its viewing angle ϕ and viewing range r_v .

Let $\mathcal{N}_i(t) = \{c_{i'} : d_{ii'}(t) < r_c\}$ define the single-hop communicative neighborhood of c_i , where $d_{ii'}(t)$ is the distance between camera c_i and camera $c_{i'}$, and r_c is their communication range.

Let N be the total number of targets. Each target o_j is assigned a priority $w_j \in (0, 1]$ and is uniquely identifiable from the other targets by discriminative features (e.g. color).

Let $\mathbf{s}_j(t) = [x_j(t), y_j(t), \dot{x}_j(t), \dot{y}_j(t)]^T$ be the state of target o_j on the ground plane, where $\mathbf{p}_j(t) = [x_j(t), y_j(t)]^T$ is its position and $\mathbf{v}_j(t) = [\dot{x}_j(t), \dot{y}_j(t)]^T$

is its velocity. Similarly, $\mathbf{s}_{ij}(t)$ is the estimated state of target o_j on the ground plane when seen from camera c_i .

We assume that a target is tracked by at most one mobile camera. A mobile camera c_i (i.e. a camera mounted on a robot) is defined at time t by the position of the robot, $[x_i(t), y_i(t)]^T$, the heading direction of the robot, $\varphi_i(t)$, and the orientation of the camera, $\theta_i(t)$.

Finally, let $\mathbf{C}_j(t)$ be the set of cameras that are jointly tracking o_j at t . The goal is to update over time the members of $\mathbf{C}_j(t)$, with $j = 1, \dots, N$, and to assign mobile cameras to targets to minimize the tracking error, maximize the target observation period and minimize energy consumption.

3. Static cameras

Static cameras perform target detection (e.g. using [4]) and assign a priority to each target based on application-specific knowledge¹. We assume that the inter-camera target association is solved using uniquely identifiable features for each target. Moreover, we assume that cameras are calibrated (e.g. using [1]) with knowledge of both intrinsic (e.g. focal length) and extrinsic (e.g. location and orientation) parameters, so that image-plane observations can be mapped onto a common ground plane.

Cameras self-organize in tracking groups via messaging with neighbors and perform distributed tracking via lossless wireless communication links whose delays are negligible with respect to the target speed.

3.1. Distributed tracking

We use the Iterative Covariance Intersection (ICI) algorithm for distributed tracking as it does not require prior knowledge of the network connectivity [6, 8]. Each camera $c_i \in \mathbf{C}_j(t)$ performs local tracking based on the Information Filter (IF) and iteratively exchanges in a synchronous manner² its local update with neighbors to agree on a state estimate [8].

Let $\mathbf{z}_{ij}(t) \in \mathbb{R}^2$ be the measurement generated by target o_j in c_i at t . The measurement model is $\mathbf{z}_{ij}(t) = \mathbf{H}_i \mathbf{s}_{ij}(t) + \nu_{ij}(t)$, where $\mathbf{H}_i \in \mathbb{R}^{2 \times 4}$ is the projection matrix from the measurement space of c_i to the state space, and $\nu_{ij}(t)$ is an additive Gaussian noise with covariance matrix $\mathbf{R}_{ij} \in \mathbb{R}^{2 \times 2}$ [7]. Each camera $c_i \in \mathbf{C}_j(t)$ first calculates the information vector $\mathbf{i}_{ij}(t) \in \mathbb{R}^4$ and the information matrix $\mathbf{I}_{ij}(t) \in \mathbb{R}^{4 \times 4}$ for target o_j :

$$\begin{aligned}\mathbf{I}_{ij}(t) &= \mathbf{H}_i^T \mathbf{R}_{ij}(t)^{-1} \mathbf{H}_i \\ \mathbf{i}_{ij}(t) &= \mathbf{H}_i^T \mathbf{R}_{ij}(t)^{-1} \mathbf{z}_{ij}(t).\end{aligned}\quad (1)$$

$\mathbf{I}_{ij}(t)$ and $\mathbf{i}_{ij}(t)$ are set to $\mathbf{0}$ when camera c_i does not observe o_j at t .

¹While the priority can be time varying, for simplicity in this work we consider it to be a constant.

²For asynchronous distributed tracking, please refer to [5].

After K iterations (we set $K = 3$ as in [8]), each camera $c_i \in \mathbf{C}_j(t)$ has a local state estimate on the ground plane $\mathbf{s}_{ij}(t)$ and a corresponding covariance matrix $\mathbf{P}_{ij}(t)$ based on the following procedure: c_i computes the initial local updates $\Gamma_{ij}^0(t) \in \mathbb{R}^{4 \times 4}$ and $\psi_{ij}^0(t) \in \mathbb{R}^4$ that are to be exchanged with neighbors:

$$\begin{aligned}\Gamma_{ij}^0(t) &= \mathbf{I}_{ij}(t) + \mathbf{P}_{ij}^-(t)^{-1} \\ \psi_{ij}^0(t) &= \mathbf{i}_{ij}(t) + \mathbf{P}_{ij}^-(t)^{-1} \mathbf{s}_{ij}^-(t),\end{aligned}\quad (2)$$

where $\mathbf{s}_{ij}^-(t)$ is the predicted target state from the previous estimation and $\mathbf{P}_{ij}^-(t) \in \mathbb{R}^{4 \times 4}$ is the corresponding covariance matrix.

At each iteration k , $c_i \in \mathbf{C}_j(t)$ sends $\Gamma_{ij}^{k-1}(t)$ and $\psi_{ij}^{k-1}(t)$ to its neighbors $\mathcal{N}_i(t)$ that are tracking o_j at t and updates $\Gamma_{ij}^k(t)$ and $\psi_{ij}^k(t)$ as:

$$\begin{aligned}\Gamma_{ij}^k(t) &= \sum_{c_{i'} \in \mathbf{C}_j(t) \cap \mathcal{N}_i(t)} \mu_{i'j}^k(t) \Gamma_{i'j}^{k-1}(t) \\ \psi_{ij}^k(t) &= \sum_{c_{i'} \in \mathbf{C}_j(t) \cap \mathcal{N}_i(t)} \mu_{i'j}^k(t) \psi_{i'j}^{k-1}(t),\end{aligned}\quad (3)$$

where $\mu_{i'j}^k(t)$ is the weight for fusion calculated as in [6].

3.2. Messaging

Static cameras that are observing the target o_j at t become members of the tracking group $\mathbf{C}_j(t)$ and inform their neighbors via a message that contains (i) the target description for identity association, (ii) the viewing utility, which defines how well the target is viewed by the camera, and (iii) the states of neighboring mobile cameras.

Messaging makes each member camera aware of the viewing utility of other member cameras as well as the presence and states of mobile cameras within their communication range.

If member cameras in the same tracking group cannot directly communicate because of their limited radio range, other cameras may join the group and behave as *bridge* to enable communication among cameras involved in tracking the same target. A camera identifies itself as bridge based on the received messages. A member camera that receives other member messages sends a confirmation message to help cameras that are not involved in tracking understand whether they are needed as bridge cameras. A camera becomes a bridge when it receives multiple member messages but less confirmation messages than member messages. The bridge camera then informs neighbors via a bridge member message, which contains the state of neighboring cameras and the information from all the received member messages.

If a member camera does not receive any member messages or bridge member messages, the camera is either the

only camera viewing that target or it is isolated (i.e. no member neighbors are present within a 2-hop communication range). This camera will therefore perform individual tracking instead of distributed tracking.

When a target has to be tracked with priority and there is no mobile camera tracking that target, the closest static camera to a mobile camera sends a request message. This message contains the target description, the estimated target state, the target priority and the states of mobile cameras that will receive the same request.

4. Mobile cameras

Mobile cameras select a target to track based on information received from static cameras. With the proposed messaging scheme, a mobile camera in a tracking group is aware of the presence of other mobile cameras and their viewing utility. If there are multiple mobile cameras in the same tracking group, the mobile camera with the highest viewing utility continues tracking while the other mobile cameras switch to idle or start tracking another target.

We assume that mobile cameras are aware of their own locations via communication with static cameras and on-board odometry measurements [3]. When there is movement, each mobile camera updates its connectivity table with neighboring cameras via radio signaling [18].

4.1. Target selection

After receiving multiple tracking requests, a mobile camera selects one target to track and starts moving to track that target jointly with static cameras. We treat the target that a mobile camera is currently tracking as important as other candidate targets that the mobile camera might receive requests to track. This allows the mobile camera to switch to track a target with higher priority.

Let $\mathbf{C}_j^{M,R}(t)$ be the set of mobile cameras that receive the request to track the same target o_j and let $\Omega_i(t)$ be the set of candidate targets that mobile camera c_i receives the requests to track.

For each $o_j \in \Omega_i(t)$, the mobile camera c_i locally decides to track the target with the highest utility $a_{ij}(t)$. The value of $a_{ij}(t)$ is proportional to the target priority w_j and inversely proportional to the energy cost for the mobile camera to capture the target.

While the energy cost is often approximated as the robot-target distance $d_{ij}(t)$ between c_i and o_j [13], we also take into account $\mathbf{v}_{ij}^r(t)$, the relative velocity between the mobile camera and the target. We approximate the energy cost as the expected robot-target distance \hat{d}_{ij} after a time period Δt (see Fig. 1), assuming that $\mathbf{v}_{ij}^r(t)$ is constant during Δt :

$$\hat{d}_{ij}(t) = |d_{ij}(t) - v_{ij}^{r,p}(t)\Delta t|, \quad (4)$$

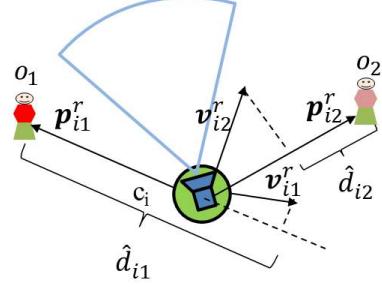


Figure 1. The distances between a mobile camera c_i and two targets, o_1 and o_2 . \mathbf{p}_{ij}^r and \mathbf{v}_{ij}^r are the relative position and velocity between c_i and o_j .

where $\Delta t = 1$ second in our experiments and $v_{ij}^{r,p}(t)$ is the relative speed of the mobile camera toward the target, which is computed as the scalar projection of the relative velocity $\mathbf{v}_{ij}^r(t)$ onto the relative position $\mathbf{p}_{ij}^r(t)$:

$$v_{ij}^{r,p}(t) = \frac{\mathbf{v}_{ij}^r(t)^T \cdot \mathbf{p}_{ij}^r(t)}{|\mathbf{p}_{ij}^r(t)|}. \quad (5)$$

Let $g_{ij}(t)$ account for the energy cost as follows:

$$g_{ij}(t) = 1 - \frac{\hat{d}_{ij}(t)}{r_v + r_c}, \quad (6)$$

where the normalization by the sum of the viewing range, r_v , and the communication range, r_c , takes into account the largest possible distance between a target and a mobile camera that receives a request to track that target.

Because multiple mobile cameras may receive the same target-tracking request. In our scenario mobile cameras do not communicate with each other, a mobile camera should discount the utility for the target to reduce the probability of selecting the same target as another mobile camera. Thus, we define the utility $a_{ij}(t)$ as:

$$a_{ij}(t) = w_j \frac{g_{ij}(t)^2}{\sum_{c_{i'} \in \mathbf{C}_j^{M,R}(t)} g_{i'j}(t)}. \quad (7)$$

Let $\Lambda_i(t)$ be the set containing the indices of all the targets in $\Omega_i(t)$. The mobile camera c_i will select the target o_{j^*} with the highest utility, defined by the index

$$j^* = \arg \max_{j' \in \Lambda_i(t)} a_{ij'}(t). \quad (8)$$

4.2. Motion strategy

Once a target is selected, the mobile camera c_i computes the control $\mathbf{a}_i(t)$ for updating the camera state in order to improve tracking accuracy in an energy-efficient manner. Let $\mathbf{u}_i(t) = [v_i(t), \alpha_i(t), \beta_i(t)]^T$ be the velocity of the

mobile camera c_i , where $v_i(t)$ is the speed along the heading direction $\varphi_i(t)$, $\alpha_i(t)$ is the steering angular speed of the robot and $\beta_i(t)$ is the panning angular speed of the camera. The control $\mathbf{a}_i(t)$ is the acceleration $\mathbf{a}_i(t) = \dot{\mathbf{u}}_i(t)$.

We consider that the best view of a target is in the center of the FoV. Let $\delta_{ij}(t)$ be the angular difference between the camera orientation and the target bearing angle. When the target is observed by the mobile camera, the target-robot distance $d_{ij}(t)$ and $\delta_{ij}(t)$ can be inferred from the image plane with the knowledge of camera calibration or target dimensions. Otherwise, $d_{ij}(t)$ and $\delta_{ij}(t)$ can be computed with the mobile camera state and ground-plane target state received from static cameras.

Let $\rho_{ij}^d(t)$ and $\rho_{ij}^\delta(t)$ be the ratio of the distance and angular difference to the center of the FoV with respect to half of the viewing range, $\frac{r_v}{2}$, and half of the viewing angle, $\frac{\phi}{2}$:

$$\rho_{ij}^d(t) = \frac{2|d_{ij}(t) - \frac{1}{2}r_v|}{r_v} \quad (9)$$

and

$$\rho_{ij}^\delta(t) = \frac{2\delta_{ij}(t)}{\phi}. \quad (10)$$

The closer $\rho_{ij}^d(t)$ and $\rho_{ij}^\delta(t)$ are to 0, the better the view of the target is.

We design two cost functions (see Fig. 2) to encode the objectives of maintaining the target at a desired position in the FoV that produces the best view on the target (J_1) and of generating an energy-efficient path toward the target (J_2). J_1 penalizes the deviation of the target position from the center of the FoV:

$$J_1 = \exp\left(\sqrt{\rho_{ij}^d(t)^2 + \rho_{ij}^\delta(t)^2}\right), \quad (11)$$

whereas J_2 penalizes the energy consumption during the motion of c_i :

$$J_2 = \kappa \exp(E_{N,i}(t)), \quad (12)$$

where $E_{N,i}(t)$ is a normalized energy consumption that takes into account the kinetic energy, $E_{k,i}$, and the energy to overcome resistance, $E_{f,i}$ [12]; the constant κ constrains J_2 so that J_1 always plays a major role during motion planning.

The kinetic energy can be represented as:

$$\begin{aligned} E_{k,i}(t) &= m_i \max(\dot{v}_i(t)d_i^v(t), 0) \\ &+ I_i^p \max(\dot{\alpha}_i(t)d_i^\alpha(t), 0) \\ &+ I_i^c \max(\dot{\beta}_i(t)d_i^\beta(t), 0), \end{aligned} \quad (13)$$

where m_i is the mass of the mobile camera; I_i^p and I_i^c are the moment of inertia around the rotation axis for the robotic platform and camera, respectively; and $d_i^v(t)$, $d_i^\alpha(t)$ and $d_i^\beta(t)$ are the camera displacements.

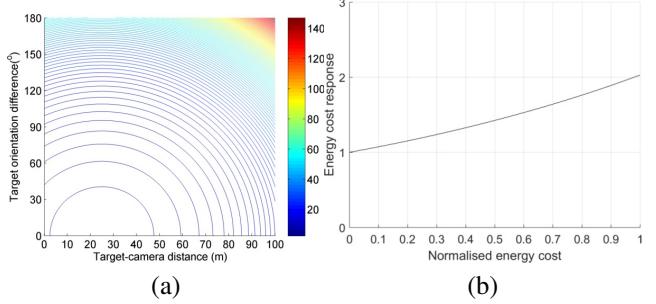


Figure 2. The response of the two cost functions: (a) J_1 when $r_v = 50m$ and $\phi = 90^\circ$; (b) J_2 when $\kappa = \sqrt{2}/2$.

The energy to overcome resistance is $E_{f,i}(t) = \mu F |d_i^v(t)|$, where μ is the coefficient of friction that depends on the contacting materials and F is the pressure force between the two contacting materials [12].

We define the normalized energy cost as:

$$E_{N,i}(t) = e_i^k \frac{E_{k,i}(t)}{E_{k,i}^{\max}} + e_i^f \frac{E_{f,i}(t)}{E_{f,i}^{\max}}, \quad (14)$$

where $E_{k,i}^{\max}$ and $E_{f,i}^{\max}$ are the maximum energy consumptions for c_i at each component; and e_i^k and e_i^f are the weights for each component. Each energy component will contribute equally in our experiments.

We use a model predictive controller (MPC) with a time horizon T_h for active tracking [14]. A mobile camera c_i at t computes a control sequence $\mathbf{a}_i^{T_h}(t) = \{\mathbf{a}_i(\tau) : \tau \in [t, t + T_h - 1]\}$ by minimizing $\lambda_1 J_1 + \lambda_2 J_2$, subject to

$$\begin{cases} \mathbf{u}_i(\tau) \leq \mathbf{u}_{\max}, & \tau \in [t, t + T_h] \\ \mathbf{a}_i(\tau) \leq \mathbf{a}_{\max}, & \tau \in [t, t + T_h]. \end{cases} \quad (15)$$

For simplicity, the disturbances from the environment and noises of the kinematic model are not considered in this work.

To solve this problem we employ the *fmincon* tool of MATLAB [14]. A global optimal solution for this nonlinear constrained optimization problem is hard to achieve with a non-convex objective function and the solution depends on the initial searching point. We accept local optimality and always initialize the starting point as $\mathbf{a}_i(t) = [0, 0, 0]$.

5. Validation

We validate the proposed framework in terms of tracking accuracy, target observation time and energy efficiency. We first quantify the improvement on tracking accuracy and energy efficiency achieved by the proposed MPC controller (MobStaMPC) and compare it with distributed tracking with only static cameras [19] (OnlyStatic) and distributed tracking with both static and mobile cameras using a one-step-ahead optimal controller [20] (MobStaOSA). We then

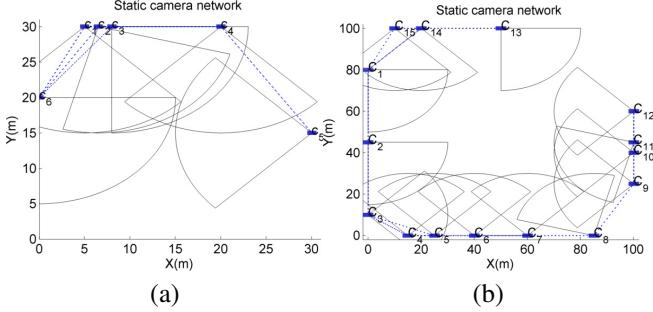


Figure 3. (a) Scenario I: 6 static cameras in a campus square. (b) Scenario II: 15 static cameras in a public square. Dashed lines indicate when cameras can communicate (1 hop) with each other.

evaluate the robot-target assignment strategy in terms of the target observation time and energy cost, and compare it with a centralized assignment (the performance upper-bound) using the Hungarian algorithm [10] and a distributed assignment considering only distance as the cost [13].

We simulate two scenarios with static cameras positioned as in real deployments in a campus square and a public square (Fig. 3). Scenario I is a $30\text{ m} \times 30\text{ m}$ campus square with 6 static cameras deployed to cover the entrance of shops and buildings (Fig. 3(a)). Scenario II is a $100\text{ m} \times 100\text{ m}$ public square with 15 static cameras deployed to monitor road traffic and access to buildings (Fig. 3(b)). For both scenarios the view angle is $\phi = 0.5\pi$. In scenario I, we set the view range $r_v = 15\text{ m}$ and the communication range $r_c = 20\text{ m}$. In scenario II, we set $r_v = 30\text{ m}$ and $r_c = 40\text{ m}$. Mobile cameras move freely with $\mathbf{u}_{\max} = [3\text{ m/s}, 0.20\pi\text{ rad/s}, 0.25\pi\text{ rad/s}]$ and $\mathbf{a}_{\max} = [1.5\text{ m/s}^2, 0.20\pi\text{ rad/s}^2, 0.25\pi\text{ rad/s}^2]$. Point targets move more slowly than mobile cameras and have an acceleration that follow a zero-mean bivariate Gaussian distribution (covariance matrix: $\text{diag}([0.3\ 0.3])$). We initialize the target priority randomly with a uniform distribution.

We quantify tracking accuracy as *mean tracking error*, ϵ , i.e. the difference between the estimated target state $s_j(t)$ within $C_j(t)$, for $j = 1, \dots, N$, and the corresponding ground-truth state $\hat{s}_j(t)$ averaged over the number of targets, N , and the whole experiment duration, T :

$$\epsilon = \frac{1}{TN} \sum_{t=1}^T \sum_{j=1}^N \|\mathbf{s}_j(t) - \hat{\mathbf{s}}_j(t)\|. \quad (16)$$

We also quantify the *observation ratio*, η , i.e. the ratio of the temporal interval during which a target is observed, weighted by the target priority:

$$\eta = \frac{1}{TW} \sum_{t=1}^T \sum_{j=1}^N B_j(t) w_j, \quad (17)$$

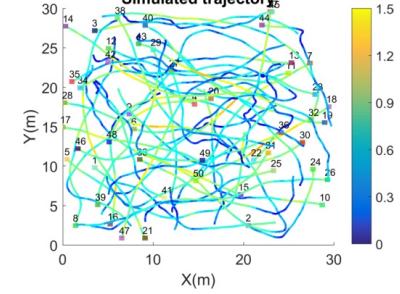


Figure 4. Simulated target trajectories. A starting point is indicated with a filled square. Colors indicate the speed of the target (m/s).

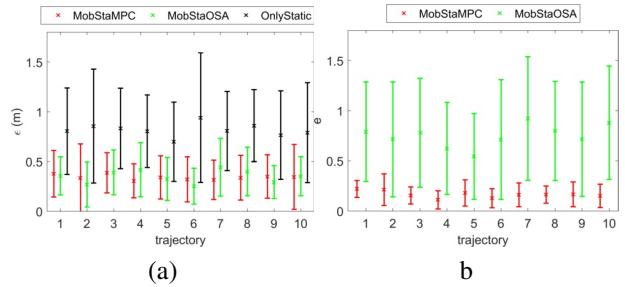


Figure 5. Comparison between MobStaMPC (the proposed framework), OnlyStatic (using only static cameras) and MobStaOSA (a motion strategy with a one-step-ahead optimal controller): (a) mean tracking error, ϵ , and (b) energy cost, e for mobile cameras.

where $W = \sum_{j=1}^N w_j$ and $B_j(t)$ is a binary function indicating whether o_j is observed at t . Specifically, we define the observation ratio of mobile cameras as $\eta^M = \frac{1}{TW} \sum_{t=1}^T \sum_{j=1}^N B_j^M(t)w_j$, where $B_j^M(t)$ indicates whether o_j is observed by a mobile camera at t . Targets with higher priority and observed for longer will lead to a larger η .

Finally, we quantify the *normalized energy consumption*, e , i.e. the total normalized energy of all mobile cameras averaged over N and T :

$$e = \frac{1}{TN} \sum_{t=1}^T \sum_{i=1}^{N_M} E_{N,i}(t), \quad (18)$$

where N_M is the number of mobile cameras.

We first validate the improvements on tracking accuracy and energy efficiency brought by the proposed motion strategy when targets are initialized at the center of the FoV of mobile cameras. We set the objective function as $J = 0.5J_1 + 0.5J_2$ with time horizon $T_h = 3$. We test 50 trajectories under scenario I. Each trajectory is independently generated and lasts 50 time steps (Fig. 4). Fig. 5(a) shows the mean tracking error and corresponding standard deviation for the first 10 tested trajectories (see Fig. 3(b)). The result with only static cameras is plotted in black as a baseline for comparison. The mean tracking error of

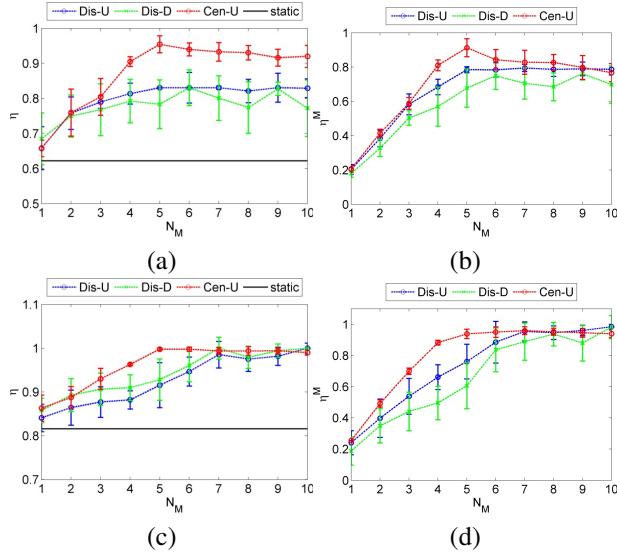


Figure 6. Comparison between Dis-U (the proposed local selection utility), Cen-U (the centralized assignment), Dis-D (the distance-based selection) in terms of observation ratios, η and η^M , with an increasing number of mobile cameras in (a, b) scenario I and (c, d) scenario II. The result of ‘static’ (using static cameras only) is shown as reference.

all tested trajectories is 0.73 m with OnlyStatic and 0.34 m with the proposed MobStaMPC, indicating $> 50\%$ improvement in tracking accuracy. Compared to the one-step-ahead optimization (MobStaOSA), MobStaMPC achieves similar tracking accuracy with much less energy due to the multiple-step-ahead planning.

We further validate the improvements of target observation time and energy reduction brought by the proposed robot-target assignment strategy (Fig. 6 and Fig. 7). The performance achieved with the centralized assignment serves as a performance upper bound as it uses the (ideal) knowledge of all robots and targets. The centralized strategy updates the assignment at each time step using the generic form of the selection utility. We also analyze the influence of the target selection criteria on the target observation ratio and the energy cost by testing the criterion using the proposed utility and the target-robot distance only. In the following experiments, we test the performance with an increasing number of mobile cameras, N_M , for tracking 5 targets with different priorities. We initialize the location of mobile cameras using a uniform distribution. Results are averaged over 10 independent runs.

Fig. 6 compares the observation ratios when the number of mobile cameras increases. The improvement in the observation ratio of mobile cameras, η^M , saturates when $N_M = N$ in scenario I and when $N_M > N$ in scenario II. This is because in a smaller area (scenario I) each target is more likely to be tracked by a mobile camera, while more

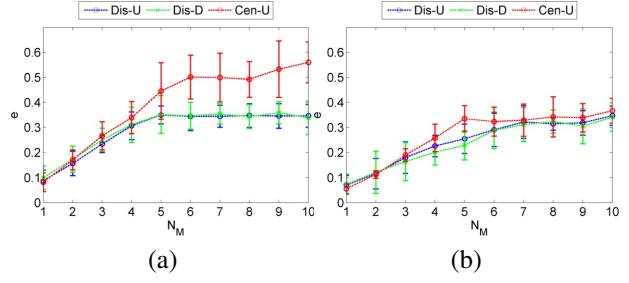


Figure 7. Comparison between Dis-U (the proposed local selection utility), Cen-U (the centralized assignment) and Dis-D (the distance-based selection) in terms of energy cost, e , with an increasing number of mobile cameras in (a) scenario I and (b) scenario II.

mobile cameras are required in a larger area due to their limited communication range. The η^M achieved with Dis-U approaches but cannot exceed that of the centralized assignment (Cen-U) when $N_M \leq N$ due to the limited available knowledge on targets and robots.

The η^M achieved with Cen-U saturates around $N_M = N$ in both scenarios and η^M may slightly deteriorate when $N_M > N$ (Fig. 6 (b)). This is because Cen-U updates the assignment at each time step, which may cause switching between mobile cameras for tracking a target when there are redundant mobile cameras. The switching can lead to temporary target loss, which reduces η^M and increases the energy cost, e . Such switching occurs more frequently in a smaller area and this explains the larger deterioration on η^M (Fig. 6 (b)) and energy cost with Cen-U (Fig. 7 (a)) in scenario I, compared to that in scenario II.

The proposed Dis-U achieves a higher η^M than the distance-based assignment (Dis-D) in both scenarios with almost the same energy consumption (Fig. 7). Dis-D aims to cover targets in an energy-efficient manner only based on distance, which makes the robot-target assignment sensitive to scene dynamics without guaranteeing longer observation on targets with higher priority.

6. Conclusion

We presented a collaborative framework for target tracking with a mixed camera network composed of static and mobile cameras. Mobile cameras move on demand based on an individual target selection strategy based on local knowledge only and using energy-efficient motion planning. The proposed strategy reduces energy consumption compared to a one-step-ahead optimal controller and achieves higher prioritized observation time compared to a traditional distance-based assignment.

Future work includes introducing collision avoidance constraints and modeling communication errors and delays.

References

- [1] N. Anjum and A. Cavallaro. Single camera calibration for trajectory-based behavior analysis. In *IEEE Conf. on Advanced Video and Signal Based Surveillance*, pages 147–152, London, Sept 2007.
- [2] J. Banfi, J. Guzzi, A. Giusti, L. Gambardella, and G.A. Di Caro. Fair multi-target tracking in cooperative multi-robot systems. In *Proc. of IEEE Int'l Conf. on Robotics and Automation*, pages 5411–5418, Seattle, WA, May 2015.
- [3] F. Chenavier and J.L. Crowley. Position estimation for a mobile robot using vision and odometry. In *Proc. of IEEE Int'l Conf. on Robotics and Automation*, pages 2588–2593, Nice, May 1992.
- [4] P. Dollar, R. Appel, S. Belongie, and P. Perona. Fast feature pyramids for object detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 36(8):1532–1545, Aug 2014.
- [5] S. Giannini, A. Petitti, D. Di Paola, and A. Rizzo. Asynchronous consensus-based distributed target tracking. In *Proc. of IEEE Conference on Decision and Control*, pages 2006–2011, Firenze, Dec 2013.
- [6] O. Hlinka, O. Sluciak, F. Hlawatsch, and M. Rupp. Distributed data fusion using iterative covariance intersection. In *Proc. of IEEE Int'l Conf. on Acoustics, Speech and Signal Processing*, pages 1861–1865, Florence, May 2014.
- [7] A.T. Kamal, J.A. Farrell, and A.K. Roy-Chowdhury. Information consensus for distributed multi-target tracking. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 2403–2410, Portland, OR, June 2013.
- [8] S. Katragadda and A. Cavallaro. Neighbour consensus for distributed visual tracking. In *Proc. of IEEE Int'l Conf. on Intelligent Sensors, Sensor Networks and Information Processing*, pages 1–6, Singapore, April 2015.
- [9] A. Kolling and S. Carpin. Multirobot cooperation for surveillance of multiple moving targets - a new behavioral approach. In *Proc. of IEEE Int'l Conf. on Robotics and Automation*, pages 1311–1316, May 2006.
- [10] H. W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, 1955.
- [11] T. P. Lambrou. Optimized cooperative dynamic coverage in mixed sensor networks. *ACM Trans. on Sensor Networks*, 11(3):46:1–46:35, Feb 2015.
- [12] S. Liu and D. Sun. Minimizing energy consumption of wheeled mobile robots via optimal motion planning. *IEEE/ASME Trans. on Mechatronics*, 19(2):401–411, April 2014.
- [13] D. Panagou, M. Turpin, and V. Kumar. Decentralized goal assignment and trajectory generation in multi-robot networks: A multiple Lyapunov functions approach. In *Proc. of IEEE Int'l Conf. on Robotics and Automation*, pages 6757–6762, Hong Kong, May 2014.
- [14] J. J. Park and B. Kuipers. Autonomous person pacing and following with model predictive equilibrium point control. In *Proc. of IEEE Int'l Conf. on Robotics and Automation*, pages 1060–1067, Karsruhe, Germany, May 2013.
- [15] L. E. Parker. Distributed algorithms for multi-robot observation of multiple moving targets. *Autonomous Robots*, 12(3):231–255, May 2002.
- [16] M. Schwager, B. J. Julian, M. Angermann, and D. Rus. Eyes in the sky: Decentralized control for the deployment of robotic camera networks. *Proc. of the IEEE*, 99(9):1541–1561, Sept 2011.
- [17] W. Sheng, Q. Yang, J. Tan, and N. Xi. Distributed multi-robot coordination in area exploration. *Robotics and Autonomous Systems*, 54(12):945 – 955, August 2006.
- [18] W. Sun, Z. Yang, X. Zhang, and Y. Liu. Energy-efficient neighbor discovery in mobile ad hoc and wireless sensor networks: A survey. *Communications Surveys Tutorials*, 16(3):1448–1459, March 2014.
- [19] Y. Wang and A. Cavallaro. Coalition formation with wireless camera networks for distributed tracking. In *Proc. of IEEE Int'l Conf. on Advanced Video and Signal-based Surveillance*, Karsruhe, Germany, Aug 2015.
- [20] H. Wei, W. Lu, P. Zhu, G. Huang, J. Leonard, and S. Ferrari. Optimized visibility motion planning for target tracking and localization. In *Proc. of IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems*, pages 76–82, Chicago, IL, Sept 2014.