



Repositorio Institucional de la Universidad Autónoma de Madrid

<https://repositorio.uam.es>

Esta es la **versión de autor** del artículo publicado en:

This is an **author produced version** of a paper published in:

IEEE Sensors Journal 15.5 (2015): 2657 – 2668

DOI: <http://dx.doi.org/10.1109/JSEN.2014.2367015>

Copyright: © 2015 IEEE

El acceso a la versión del editor puede requerir la suscripción del recurso

Access to the published version may require subscription

Cost-aware coalitions for collaborative tracking in resource-constrained camera networks

Juan C. SanMiguel and Andrea Cavallaro

Abstract—We propose an approach to create camera coalitions in resource-constrained camera networks and demonstrate it for collaborative target tracking. We cast coalition formation as a decentralized resource allocation process where the best cameras among those viewing a target are assigned to a coalition based on marginal utility theory. A manager is dynamically selected to negotiate with cameras whether they will join the coalition and to coordinate the tracking task. This negotiation is based not only on the utility brought by each camera to the coalition but also on the associated cost (i.e. additional processing and communication). Experimental results and comparisons using simulations and real data show that the proposed approach outperforms related state-of-the-art methods by improving tracking accuracy in cost-free settings. Moreover, under resource limitations the proposed approach controls the tradeoff between accuracy and cost, and achieves energy savings with only a minor reduction in accuracy.

Index Terms—camera network, coalition, utility, cost, tracking.

I. INTRODUCTION

Networks of wireless cameras with embedded capabilities for sensing, processing and communication enable emerging pervasive and mobile applications in many areas such as environment monitoring and human-computer interaction [1][2]. Fixed cameras can cooperate wirelessly with mobile battery-powered cameras to capture closer views of selected objects [3]. Cameras in large networks may interact locally forming coalitions to provide coordinated decisions across neighboring cameras. Coalitions can benefit target tracking [4], a key task in camera networks, where sets of cameras may operate jointly using decentralized tracking [5].

Current solutions to manage multiple cameras employ centralized [6][7][8][9][10] or decentralized [11][12][13][14] architectures. Centralized approaches have limited scalability and often cast the problem as a global resource allocation. Decentralized proposals rely on camera ranking [11][13][14] or local interaction [12]. Cameras are grouped according to criteria such as camera-target distance [12][14], Field of View (FOV) overlap [8], inter-view target matching [11][13],

occlusion-free views [9] and user preferences [7]. Cameras are assumed to provide accurate data and to contribute equally to the coalition, which may not be true due to errors and view variability across cameras. Coalition formation requires identifying the best cameras over time. Moreover, collaboration costs of forming and operating camera coalitions are generally not defined [11][13][7][9][14], thus limiting the use of existing approaches in resource-constrained camera networks.

In this paper, we propose a framework for camera-coalition formation that simultaneously accounts for benefits and costs of camera collaboration. We cast coalition formation as a resource allocation problem where cameras are available network resources. A greedy decentralized approach quantifies the expected benefit of cameras joining the coalition via their utility for tracking a target. Coalition costs are defined based on the energy required for communication and processing. Using marginal utility theory [15], we estimate the contribution of each camera for the tracking task for their efficient allocation to a coalition. A coalition manager negotiates with cameras when their utility justifies the associated cost of joining a coalition. Cameras participate in a coalition until exhaustion of their capabilities, measured as load and remaining energy. The coalition is formed and updated as the target moves, thus adapting to resource and environment changes. Simulations and real-data experiments show that the proposed approach improves related works for coalition formation and tracking.

The paper is organized as follows: Section II discusses the related work and Section III states the addressed problem. Section IV overviews the proposed framework for coalition-based tracking and Section V describes the proposed approach for coalition formation. Section VI presents the experimental results and Section VII concludes the paper.

II. RELATED WORK

Coalitions in camera networks can be formed statically (at system initialization) or dynamically (at predefined intervals or driven by external events) [11]. We are interested in dynamic approaches to create event-driven coalitions for target tracking. These approaches can be based on centralized optimization, decentralized rankings or local interaction (Table I).

Centralized optimization maximizes functions that quantify target location accuracy [6][17][9] or visibility [10]. Such optimization is often restricted to small-size networks as it explores all the combinations of cameras to form a coalition [6][17]. Greedy approaches reduce such complexity by sequentially expanding the coalition only with cameras maximizing certain criteria [9][10]. Centralized approaches need

This work was partially supported by the EU Crowded Environments monitoring for Activity Understanding and Recognition (CENTAUR, FP7-PEOPLE-2012-IAPP) project under GA number 324359, by the Artemis JU and UK Technology Strategy Board as part of the Cognitive & Perceptive Cameras (COPCAMS) project under GA number 332913 and by the Spanish Government (TEC2011-25995). Juan C. SanMiguel and Andrea Cavallaro are with Centre for Intelligent Sensing, Queen Mary University of London, Mile End Road, E1 4NS, London (UK), e-mail: juan.carlos.sanmiguel@qmul.ac.uk, a.cavallaro@qmul.ac.uk

Copyright (c) 2014 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Table I

SUMMARY OF MAIN APPROACHES FOR COALITION FORMATION IN TARGET TRACKING. KEY: DIS: DISTRIBUTED; DEC: DECENTRALIZED; CEN: CENTRALIZED; O: OPTIMIZATION; R: RANKING; LI: LOCAL INTERACTION.

Ref.	Scheme	Type	Approach	Multiple cameras	Accuracy indicators	Collaborative processing	Costs		Resources	
							Comms	Process	Battery	Load
[11]	DEC	R	Distance grouping	✓		✓				
[12]	DEC	LI	Auction	✓		✓				
[13]	DEC	R	Feature grouping	✓	✓	✓				
[6]	CEN	O	Full optimization	✓	✓	✓			✓	
[16]	CEN	O	Full optimization							✓
[7]	CEN	LI	Game-theory	✓	✓					
[17]	CEN	O	Full optimization	✓		✓				
[9]	CEN	O	Greedy optimization	✓	✓	✓				
[14]	DEC	LI	Auction	✓						
[10]	CEN	O	Greedy optimization	✓	✓	✓				
Proposed	DEC	LI	Greedy optimization	✓	✓	✓	✓	✓	✓	✓

a global network view, including for example, the knowledge of the number of targets and available resources [9].

Decentralized rankings select top-ranked cameras based on camera-target distance [11] or feature matching such as orientation [13]. Distance-based criteria require accurate target state estimations to create coalitions whereas feature-based criteria rely on matching accuracy. Each camera sends target-related data to neighboring cameras to decide whether to join the coalition. These approaches replicate the computational cost needed to transmit data for distributed decision making.

Local (pair-wise) interaction enables scalability at network and target level. Such interaction can be decentralized or distributed if some or none of the cameras act as coordinators, respectively. For example, the contract net protocol is used to create coalitions via decentralized auction schemes (announce, bid and select) [12]. Optimal allocation of shared resources is solved via centralized Constraint Satisfaction that has the limitations of centralized optimization. Auctions are also used to handover targets between cameras based on camera-target distance and only the closest camera is selected to track [14]. Centralized interaction can also rely on game theory such as the bargaining of cameras in a multi-player game to maximize the network utility aligned to user-defined criteria over the target [7]. The highest-utility camera is used for tracking without collaboration among cameras in the coalition [7][14].

The interaction between mobile and fixed cameras implicitly imposes resource constraints as mobile cameras are battery-powered and communicate wirelessly [18]. The resources required to collaborate should be used to create the coalition as collaboration may imply costs not justified by the associated performance improvements. Resource constraints are used to select cameras that are idle [12], have high remaining battery level [6] or have available resources [8]. These approaches do not consider collaboration costs and coalitions are based on individual camera resources. Other approaches define computational [16] and transmission [19] costs without collaborative processing, thus decreasing the benefits of coalitions.

Similar approaches also exist for coalitions in *wireless sensor networks* [20][21]. These methods assume known sensor accuracy and allocate sensors to targets by considering load [20] and energy [21] constraints. However, such detailed sensor knowledge may not be available in camera networks. Moreover, coalition assignments use distributed optimization

based on global network information. This approach may limit scalability and requires high computational costs. Instead, the proposed approach only estimates camera utilities and employs local interaction without requiring any global data. Unlike the homogeneous collaborative processing [11][13], we weight coalition collaboration based on cameras' utilities. We propose a framework to model utilities and costs of camera coalitions, whereas related works in camera networks only describe cost-free coalitions [12][9][10][14]. While camera battery levels are used [6], collaboration costs are not modeled. Finally, unlike centralized approaches [6][8][9][10], we define a decentralized scheme for fast allocation of cameras to the coalition.

III. PROBLEM FORMULATION

Let $\Omega = \{c_1, \dots, c_N\}$ be a network of N cameras. We consider networks of calibrated and resource-limited wireless cameras. We suppose synchronized and delay-free communication with a range twice the Field of View (FOV) range and that cameras viewing the same target (i.e. overlapping FOVs) can directly communicate. Cameras with non-overlapping FOVs may communicate via look-up tables [22] or adaptive transmission ranges [23]. We use the index k to define the time steps when target measurements are captured. Let \mathbf{x}^k be the state of a moving target defined as $\mathbf{x}^k = [x, y, \dot{x}, \dot{y}]$, where (x, y) is the target center and (\dot{x}, \dot{y}) is the target velocity. The measurements obtained by each c_i are $\mathbf{z}_i^k = [z_{i,x}, z_{i,y}]$.

Let \mathcal{L}^k be a subset of cameras c_i viewing the target at k :

$$\mathcal{L}^k = \{c_i : c_i \in \Omega, 1 \leq i \leq N\}, 0 \leq |\mathcal{L}^k| \leq N, \quad (1)$$

where $|\cdot|$ is the set cardinality (size). \mathcal{L}^k can be obtained using external calibration data [22] or by matching distinguishable targets across camera views using for example color histograms [11]. Target-to-measurement association is considered given for each camera (no clutter).

For tracking a target, sub-sets of cameras are grouped into a time-varying coalition \mathcal{C}^k at each k :

$$\mathcal{C}^k = \{c_j : c_j \in \mathcal{L}^k, 1 \leq j \leq N\}, 0 \leq |\mathcal{C}^k| \leq |\mathcal{L}^k|, \quad (2)$$

where $|\mathcal{C}^k|$ is the coalition size. The goal is to automatically identify $\mathcal{C}^k \subseteq \mathcal{L}^k$ over time, i.e. the best set of cameras for the coalition that can increase performance via collaboration.

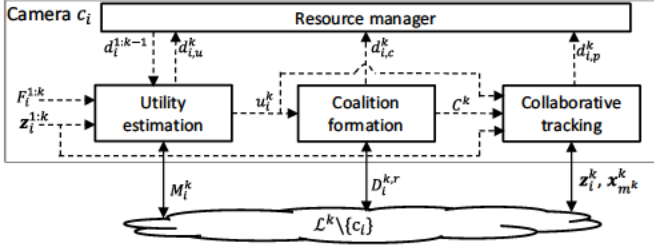


Figure 1. The proposed coalition-based tracking framework. Utility is estimated by the cameras viewing the target. Coalition formation and tracking are performed by the manager c_m^k and selected cameras, respectively. Dotted and solid arrows represent internal (stages in each camera) and external (cameras in the network) connections, respectively.

IV. COALITION-BASED COLLABORATIVE TRACKING FRAMEWORK

The proposed decentralized multi-camera framework for coalition-based target tracking is divided into four stages (Figure 1): resource management, utility estimation, coalition formation and collaborative tracking. The key symbols we use in this paper are given in Table II.

A. Resource management and utility estimation

Each camera c_i consists of an image sensor, a processor, a resource manager and a wireless communication module. The resource manager receives consumption-related information from the three other modules and monitors the currently available resources. The camera resources are described as the current, l_i^k , and maximum, l_{max} , processing load, the current, b_i^k , and initial, b_{max} , battery level and the energy consumptions or demands $d_i^{1:k-1}$ up to time $k-1$.

We create coalitions by estimating the utilities of the cameras to track the target. A utility function measures the goodness of actions in specific problems and helps the decision making process [24]. In camera networks, utilities are employed to express preference for occlusion-free views [9], representative camera views [7] or FOV configurations [25].

Let \succsim be a pairwise preference relation $\forall c_i \in \Omega$, where $c_1 \succsim c_2$ means that c_1 is preferred or indifferent to c_2 . In our case, a utility function $g: \Omega \rightarrow \mathbb{R}^+$ represents \succsim if and only if, $\forall c_1, c_2 \in \Omega$, $c_1 \succsim c_2 \Leftrightarrow g(c_1) \geq g(c_2)$. We define the utility of each camera $c_i \in \mathcal{L}^k$ to track the target using the current and past frames $F_i^{1:k}$, the target measurements $z_i^{1:k}$ and the available resources (via the resource manager) as:

$$u_i^k = g(\mathbf{x}^k | F_i^{1:k}, z_i^{1:k}, c_i), \quad (3)$$

where $g(\cdot)$ depends on the estimated target state \mathbf{x}^k (e.g. target size [7] and location [25]) and the configuration of camera c_i (e.g. FOV distortion and focus [26]).

In order to make the utility and the current camera resources available to other cameras in \mathcal{L}^k , messages $M_i^k = (u_i^k, o_i^k, b_i^k, l_i^k)$ are broadcast to the network where o_i^k is the camera priority. For example, an heterogeneous camera network may use different priorities depending on each camera capabilities. Then, after selecting a coalition manager c_m^k and forming a coalition \mathcal{C}^k for a target (see Section V), we proceed with collaborative tracking within the coalition.

Table II
KEY SYMBOLS AND NOTATIONS

Symbol	Notation
k	Time index for obtaining measurements
K	Total number of measurements
r	Index for negotiations
c_i	Camera i^{th}
z_i^k	Target measurement vector obtained by c_i
u_i^k	Camera utility for c_i
b_i^k	Battery level for c_i (energy)
l_i^k	Load level for c_i (number of targets)
d_i^k	Energy task consumption for c_i
$g(\cdot)$	Function to compute camera utility
M_i^k	Message broadcast by c_i (utility and resources)
$D_i^{k,r}$	Message from c_i to c_m^k (reply to join proposals)
\mathcal{L}^k	Set of cameras simultaneously viewing the target
c_m^k	Coalition manager
\mathbf{x}^k	Target state vector estimated by c_m^k
\mathbf{P}^k	Target state covariance estimated by c_m^k
$h(\cdot)$	Function to select the camera manager
$\mathcal{C}^{k,r}$	Set of cameras for the coalition under formation
\mathcal{C}^k	Set of cameras for the coalition
$\Psi(\mathcal{C}^{k,r})$	Coalition utility
$\Theta_s(\mathcal{C}^{k,r})$	Coalition s^{th} cost
$\Delta U(c_i, \mathcal{C}^{k,r})$	Marginal utility of c_i to join the coalition
$\Delta C_s(c_i, \mathcal{C}^{k,r})$	Marginal s^{th} cost of c_i to join the coalition
$\Upsilon(c_i, \mathcal{C}^{k,r})$	Combination of marginal utility and costs
λ_s	Factor to weight the s^{th} marginal cost

B. Collaborative tracking

We perform decentralized tracking via the coalition manager c_m^k that estimates the target state over time. Each camera in the coalition $c_j \in \mathcal{C}^k$ generates its target measurement z_j^k , which is sent to c_m^k . We use the Information Filter (IF) fusion scheme [27] where the target state is first predicted by c_m^k and then it is updated with the received z_j^k from coalition cameras.

For each time step k , the manager performs prediction as:

$$\mathbf{x}^{k-} = \phi \mathbf{x}^{k-1}, \quad (4)$$

$$\mathbf{P}^{k-} = \phi \mathbf{P}^{k-1} \phi^T + \mathbf{Q}^k, \quad (5)$$

where \mathbf{x}^{k-} and \mathbf{x}^{k-1} are the predicted and previous target states at c_m^k , respectively; \mathbf{P}^{k-} and \mathbf{P}^{k-1} are the predicted and previous error covariance at c_m^k , respectively; $\phi \in \mathbb{R}^{\rho \times \rho}$ is the state prediction matrix; $\rho = 4$ is the target state size and \mathbf{Q}^k is the noise covariance defined as a zero-mean Gaussian process. Such prediction is then updated with the received z_j^k measurements using the IF scheme via the Fisher information vector $\mathbf{v}^k \in \mathbb{R}^\rho$ and matrix $\mathbf{V}^k \in \mathbb{R}^{\rho \times \rho}$ as:

$$\mathbf{v}^k = (\mathbf{P}^{k-})^{-1} \mathbf{x}^{k-} + \sum_{\forall j | c_j \in \mathcal{C}^k} u_j^k (\mathbf{H}_j^k)^T (\mathbf{W}_j^k)^{-1} z_j^k, \quad (6)$$

$$\mathbf{V}^k = (\mathbf{P}^{k-})^{-1} + \sum_{\forall j | c_j \in \mathcal{C}^k} u_j^k (\mathbf{H}_j^k)^T (\mathbf{W}_j^k)^{-1} \mathbf{H}_j^k, \quad (7)$$

where \mathbf{H}_j^k and \mathbf{W}_j^k are the measurement and associated covariance matrix for c_j , respectively; and \mathcal{C}^k is the coalition

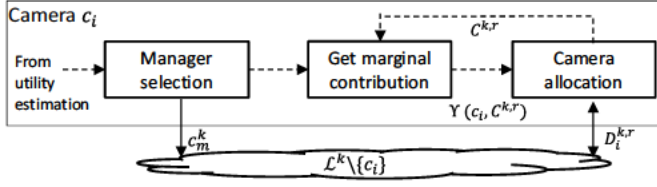


Figure 2. Block diagram of the proposed approach for coalition formation. The highest-energy camera in \mathcal{L}^k selects the manager, which estimates the contributions of cameras and iteratively allocates them to the coalition. Dotted and solid arrows represent internal (stages in each camera) and external (cameras in the network) connections, respectively.

determined by c_m^k . Note that the contribution of c_j is weighted according to its utility u_j^k . Finally, the target state and its covariance are estimated in the coalition manager as:

$$\mathbf{x}^k = (\mathbf{V}^k)^{-1} \mathbf{v}^k, \quad (8)$$

$$\mathbf{P}^k = (\mathbf{V}^k)^{-1}. \quad (9)$$

As the coalition manager may change over time, the results obtained (\mathbf{x}^k and \mathbf{P}^k) are sent to the next manager c_m^{k+1} . Finally, the demanded energy by all processes is updated in the resource manager of each c_j : $d_{j,u}^k$ for utility estimation, $d_{j,c}^k$ for coalition formation and $d_{j,p}^k$ for collaborative tracking.

V. COALITION FORMATION

The proposed decentralized approach for camera coalition formation models cameras as network resources and allocates them to a coalition considering their utilities as well as the associated collaboration costs (Figure 2).

A. Manager selection

Dynamic coalition formation requires cameras to take time-varying roles for enabling decentralized coordination and determining optimal network-wide decisions. A coalition manager is selected at every time step k , thus distributing the needed resources for coordination. The coalition manager communicates with the cameras in the network, determines the coalition members and performs tracking.

Existing approaches perform *centralized manager selection* randomly [28] or using energy-based criteria [8]. These approaches have low scalability and often require high computational costs. We instead cast such selection problem as an unconstrained local optimization based only on the available resources in \mathcal{L}^k , to determine the coalition manager c_m^k as:

$$c_m^k = \underset{\forall c_i \in \mathcal{L}^k}{\operatorname{argmax}} h(c_i), \quad (10)$$

where $h(c_i)$ measures the *adequacy* of c_i to be the coalition manager. This manager selection is only performed by the c_i with the highest battery level in \mathcal{L}^k . The *adequacy* of c_i is a weighted combination of its current energy level and load:

$$h(c_i) = \beta h_b(b_i^k) + (1 - \beta) h_l(l_i^k), \quad (11)$$

where $h_b(b_i^k) = b_i^k / b_{max}^k$; $h_l(l_i^k) = l_{max}^k / l_i^k$; $\beta \in [0, 1]$ weights the resource priority; $b_{max}^k = \max \{b_i^k\}$ and $l_{max}^k = \max \{l_i^k\}$

are the maximum battery and load in \mathcal{L}^k ($\forall c_i \in \mathcal{L}^k$). Then, c_m^k creates the coalition with at most $|\mathcal{L}^k|$ steps (negotiations) between consecutive measurements at k and $k + 1$.

B. Marginal camera contribution

The formation of effective coalitions depends on the utility of cameras joining the coalition under formation $\mathcal{C}^{k,r}$ subject to their associated costs. The index r represents each round of negotiation between c_i and c_m^k where cameras are sequentially added to $\mathcal{C}^{k,r}$. When adding new cameras is no longer useful, $\mathcal{C}^{k,r}$ becomes the final coalition \mathcal{C}^k . Assessing the contribution of a camera (utility and costs) is needed to expand $\mathcal{C}^{k,r}$.

We use marginal utility theory [15] to measure the local contribution of a camera. We define the marginal utility of adding c_i to the coalition under formation as:

$$\Delta U(c_i, \mathcal{C}^{k,r}) = \Psi(\mathcal{C}^{k,r}) - \Psi(\mathcal{C}^{k,r} \setminus \{c_i\}), \quad (12)$$

where $\Psi(\mathcal{C}^{k,r})$ and $\Psi(\mathcal{C}^{k,r} \setminus \{c_i\})$ are the utilities of the coalition after and before adding c_i (without considering costs). The coalition utility $\Psi: \mathcal{C}^{k,r} \rightarrow \mathbb{R}^+$ is defined as:

$$\Psi(\mathcal{C}^{k,r}) = 1 - \prod_{\forall i | c_i \in \mathcal{C}^{k,r}} (1 - o_i^k u_i^k), \quad (13)$$

where u_i^k and o_i^k are the utility and priority of c_i . This type of combination of utilities has been employed in cooperative communications [29], video surveillance [30] and wireless sensor networks [21]. Note that although o_i^k may be adapted over time to the network goals, in this work we assume equal priority for all cameras ($o_i^k = 1, \forall c_i \in \mathcal{L}^k$), thus coalition formation is only driven by utility and available resources.

We define the marginal cost of adding c_i to the coalition as:

$$\Delta C_s(c_i, \mathcal{C}^{k,r}) = \Theta_s(\mathcal{C}^{k,r}) - \Theta_s(\mathcal{C}^{k,r} \setminus \{c_i\}), \quad (14)$$

where $\Theta_s: \mathcal{C}^{k,r} \rightarrow \mathbb{R}^+$ is the s^{th} coalition cost. We define two costs: one for processing, Θ_p , and one for communication, Θ_c .

The *processing* cost $\Theta_p(\mathcal{C}^{k,r})$ measures the energy required by the cameras to operate in the coalition:

$$\Theta_p(\mathcal{C}^{k,r}) = |\mathcal{C}^{k,r}| \sum_{n=1}^A e_p(a_n^c) + \sum_{n=1}^A e_p(a_n^m), \quad (15)$$

where A is the number of task types; a_n^c and a_n^m are the number of type n tasks performed by all $c_i \in \mathcal{C}^k$ and c_m^k , respectively; $e_p(\cdot)$ is the energy required to complete a task.

The *communication* cost $\Theta_c(\mathcal{C}^{k,r})$ is associated to the energy needed to exchange data within the coalition:

$$\Theta_c(\mathcal{C}^{k,r}) = (E_{tx} + |\mathcal{C}^{k,r}| E_{rx}) \sum_{n=1}^P (p_n^c + p_n^m) q_n, \quad (16)$$

where P is the number of available packet types; p_n^c and p_n^m are the number of type n packets exchanged by any camera c_i and c_m^k , respectively; q_n is the number of bits of type n packet; E_{tx} and E_{rx} are constants to define the energy/bit to transmit and to receive. Standard consumption models [31]

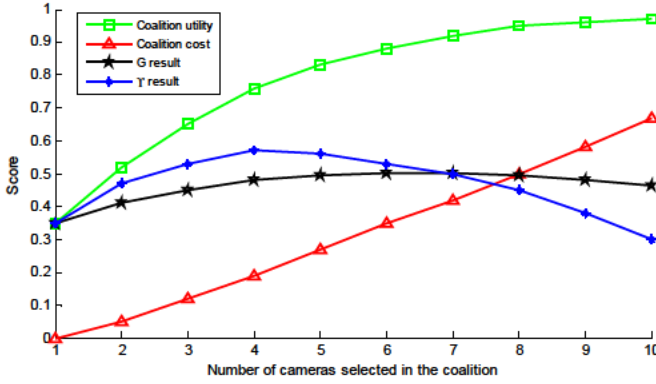


Figure 3. Illustrative example to determine the coalition using the proposed marginal contribution, Υ , and global optimization [15], G . The score (y-axis) shows the coalition values (utility and cost) obtained by the approaches. The maximum value for Υ and G is the optimal set size (4 and 6, respectively).

can be used as energy models in Eqs. (15) and (16). We define only the communication for the coalition collaboration and consider two packet types to exchange measurements (p_1^c , from the coalition cameras to the manager) and estimations (p_2^m , from the manager to the coalition cameras).

Finally, the marginal contribution of a camera to the coalition is determined by jointly considering the marginal utility and cost, which are combined using $\Upsilon : \mathcal{C}^{k,r} \rightarrow \mathbb{R}$ as:

$$\Upsilon(c_i, \mathcal{C}^{k,r}) = \Delta U(c_i, \mathcal{C}^{k,r}) - \sum_{s=1}^{N_c} \lambda_s \Delta C_s(c_i, \mathcal{C}^{k,r}), \quad (17)$$

where $N_c = 2$ is the number of costs and λ_s weights the relative importance between unitary variations of the ΔC_s cost and ΔU utility. λ_s allows us to define the constraints related to the limited resources.

Figure 3 compares the proposed use of marginal camera contributions, Υ , and the traditional global optimization [15], G , for sequential expansion of $\mathcal{C}^{k,r}$ using a set \mathcal{L}^k with 10 cameras. Υ requires a local decision to select the best camera at each r expansion step (x-axis) whereas G focuses on the global difference between utility and cost. Υ and G obtain two optimal coalitions with 4 and 6 cameras, respectively. For Υ , the cost of including any 5th camera is higher than the improvement in utility. For G , the optimization is influenced by the previously included cameras and the difference is higher than the previous value for a particular camera.

C. Camera allocation

For allocating cameras to the coalition, the search space contains all sub-sets $\mathcal{C}^{k,r} \subseteq \mathcal{L}^k$ of any size. This subset-selection is an NP-hard problem in general, requiring the evaluation of $2^{|\mathcal{L}^k|}$ sub-sets. Centralized optimization schemes are not feasible to generate the optimal coalition due their high computational complexity and their requirement of global knowledge of the network resources. A distributed search scheme relying on local information is desired for fast coalition formation and robustness to failures.

We cast the coalition formation as a greedy decentralized resource allocation via rounds of local negotiations. We use the

Contract Net Protocol [32] and formulate such resource allocation as bidding processes. Greedy approaches have been also applied to small-size multi-camera networks for measurement selection [9][10]. Both approaches differ from the proposed one as they are centralized, do not consider costs and do not scale well for large camera networks.

For each round r , c_m^k determines the best camera $\bar{c}_i \in \mathcal{L}^k$ to join the coalition $\mathcal{C}^{k,r}$ based on its marginal contribution:

$$\bar{c}_i = \underset{\forall c_i \in \mathcal{L}^k \setminus \mathcal{C}^{k,r-1}, \Upsilon(c_i, \mathcal{C}^{k,r}) > 0}{\operatorname{argmax}} \Upsilon(c_i, \mathcal{C}^{k,r}), \quad (18)$$

where the constraint $\Upsilon(c_i, \mathcal{C}^{k,r}) > 0$ imposes that the utility is greater than the cost of camera \bar{c}_i joining the coalition.

Then, c_m^k sends to \bar{c}_i a bid with the value $\Upsilon(\bar{c}_i, \mathcal{C}^{k,r})$ for the r^{th} negotiation round. We use the individual rationality principle [33] and \bar{c}_i accepts the proposed allocation by c_m^k until depletion of its resources, according to:

$$D_i^{k,r} = \begin{cases} 1 & \text{if } b_i^k - e_p > 0 \wedge l_i^k + l_p < l_{\max} \\ 0 & \text{otherwise} \end{cases}, \quad (19)$$

where 1 indicates the acceptance and 0 indicates the rejection by \bar{c}_i at round r of the manager proposal; e_p and l_p are, respectively, the energy and load cost associated to the resources that \bar{c}_i allocates to participate in the coalition. The remaining \bar{c}_i resources, battery b_i^k and load l_i^k , are updated after this allocation to be ready for the next bid, if any, to join another coalition. Finally, \bar{c}_i informs c_m^k with the decision $D_i^{k,r}$.

After each negotiation round, c_m^k confirms the acceptance of camera joining the coalition (i.e. $D_i^{k,r} = 1$) and iteratively repeats this round-based negotiation until the marginal cost of adding a new camera is higher than the associated marginal utility ($\Upsilon(c_i, \mathcal{C}^{k,r}) < 0$). Thus, c_m^k is able to refine the coalition, achieving efficient allocation of cameras and adapting dynamically to resource changes. Algorithm 1 summarizes the proposed approach for coalition formation.

VI. EXPERIMENTAL RESULTS

We evaluate the proposed approach, CAWC (Cost-Aware Coalition), for target tracking via simulated and real camera networks. Simulations allow testing for variable-size networks without deployment limitations. Real data examines the impact of real trackers and utilities on coalition performance.

A. Setup

The coalition manager tracks the target in a common ground-plane by fusing the measurements of cameras. The motion model of targets is $\mathbf{x}^{k+1} = \phi \mathbf{x}^k + \zeta^k$, where ϕ is a first order model [34] and $\zeta^k \sim \mathcal{N}(0, \mathbf{Y})$ is a zero-mean Gaussian noise with covariance $\mathbf{Y} = \operatorname{diag}([10 \ 10 \ 1 \ 1])$. As performance measure, we use the tracking error defined as the L2 norm [34] between the true state and the state estimated by the coalition manager. We compute the mean value for all the targets and report this value as performance indicator.

We analyze cost-free and cost-aware camera networks. For cost-free, the proposed approach uses utilities to select best

Algorithm 1 Cost-Aware Coalition (CAWC) formation.**Phase 1: Manager selection**Input: $c_i \in \mathcal{L}^k$ (cameras viewing the target)Output: c_m^k (coalition manager)

```

1:  $v_{max} = 0$ 
2: for  $i = 1$  to  $|\mathcal{L}^k|$  do
3:   compute  $h(c_i)$  (Eq. (10)).
4:   if  $h(c_i) > v_{max}$  then
5:      $c_m^k \leftarrow c_i$ 
6:      $v_{max} = h(c_m^k)$ 
7:   end
8: end

```

Phase 2: Camera allocationInput: $c_i \in \mathcal{L}^k$ and c_m^k Output: \mathcal{C}^k (selected coalition)

```

1:  $r = 0$ 
2:  $\mathcal{C}^{k,r} \leftarrow \emptyset$  (camera set for negotiating)
3: While  $v_{max} > 0$  do
4:    $r = r + 1$ 
5:    $v_{max} = 0$ 
6:   for  $i = 1$  to  $|\mathcal{L}^k|$  do
7:      $\mathcal{C}^{k,r} \leftarrow \mathcal{C}^{k,r} \cup \{c_i\}$ ,  $c_i \in \mathcal{L}^k$ 
8:      $c_m^k$  computes  $\Delta U(c_i, \mathcal{C}^{k,r})$ ,  $\Delta C_s(c_i, \mathcal{C}^{k,r})$  and  $\Upsilon(c_i, \mathcal{C}^{k,r})$ 
       (as Eqs. (12), (14) and (17), respectively).
9:     if  $\Upsilon(c_i, \mathcal{C}^{k,r}) > v_{max}$  then
10:       $\bar{c}_i \leftarrow c_i$ 
11:       $v_{max} = \Upsilon(c_i, \mathcal{C}^{k,r})$ 
12:    end
13:  end
14:   $D_i^{k,r} = L(\bar{c}_i, c_m^k)$ 
15:  if  $D_i^{k,r} = 1$  then
16:     $\mathcal{C}^{k,r} \leftarrow \bar{c}_i$ 
17:  end
18: end
19:  $\mathcal{C}^k = \mathcal{C}^{k,r}$ 

```

Function L (local negotiation for resource allocation)Input: \bar{c}_i (selected camera) and c_m^k Output: $D_i^{k,r}$ (reply message)

```

1:  $c_m^k$  sends bid ( $v_{max}$ ) to  $\bar{c}_i$  to join coalition
2:  $\bar{c}_i$  calculates  $D_i^{k,r}$ 
3:  $\bar{c}_i$  sends  $D_i^{k,r}$  to  $c_m^k$  (accept bid if enough resources)
4: if  $D_i^{k,r} = 1$  then
5:    $\bar{c}_i$  updates resource usage ( $b_i^k$  and  $l_i^k$ )
6: end

```

subsets with the 25%, 50%, 75% and 100% cameras in \mathcal{L}^k (respectively CAWC-25, CAWC-50, CAWC-75 and CAWC-100). We compare against the relevant work: best camera based on utility (BU) [7] and target distance (BD) [14]; fusion of all cameras in \mathcal{L}^k assuming homogeneity (ALL) [11] and ideal weighting with known measurement accuracy (IDEAL).

We evaluate cost-aware settings by considering processing (CAWC-P), communication (CAWC-C) and both (CAWC-ALL). The marginal costs λ_s define the resource constraints (see Eq. 17). For comparison, we use brute-force coalition creation (BF) and all $c_i \in \mathcal{L}^k$ (ALL) as, respectively, the optimal

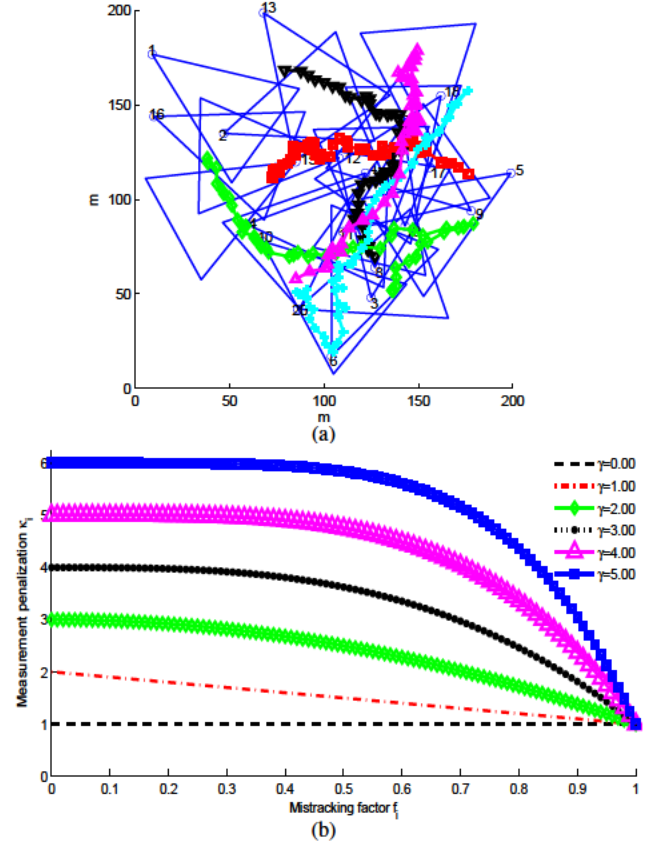


Figure 4. Setup for the experiments with simulated data. (a) Network of 20 cameras randomly deployed ($200m \times 200m$). FOVs are depicted as blue triangles and target trajectories (5) as colored lines. (b) Measurement weight κ_i for mistracking values of i^{th} camera and heterogeneity factors γ .

formation strategy and the cost-free optimal performance.

For the coalition costs, we use standard consumption models for communication and processing with default values [31]: $E_{tx} = E_{rx} = 50nJ/bit$, $q_1 = 768bits$ and $q_2 = 25bits$. We do not consider the costs for sensing and utility estimation ($d_{iu}^k = 0 \forall c_i$) as they are common for all cameras and the drawn energy is only for the coalition creation $d_{ic}^k = 0 \forall c_i \neq c_m^k$; and coalition processing $d_{ip}^k = \Theta_c(c_i) + \Theta_p(c_i)$, $\forall c_i \in \mathcal{C}^k$. We use $\beta = 0.5$ for selecting c_m^k (Eq. 11).

B. Simulations

1) *Data*: Targets move in a $200m \times 200m$ area observed by randomly placed cameras with overlapping FOVs. A 2D model defines each camera FOV [35] as its spatial location $s_i = (x, y)$, orientation angle (θ_i), covering distance ($d_i = 80$) and viewing angle ($\varpi_i = \pi/4$). The parameters s_i and θ_i are randomly selected from a uniform distribution. Figure 4(a) shows a generated network and sample trajectories of targets.

Each camera provides measurements when the target is inside its FOV as $z_i^t = H_i x^t + \eta_i^t$, where H_i is the measurement matrix [34] and $\eta_i^t \sim \mathcal{N}(0, \kappa_i W)$ is the measurement noise defined as a zero-mean Gaussian noise with variable covariance $\kappa_i W = \kappa_i \text{diag}([5 \ 5])$. The term κ_i determines the measurement accuracy of each camera:

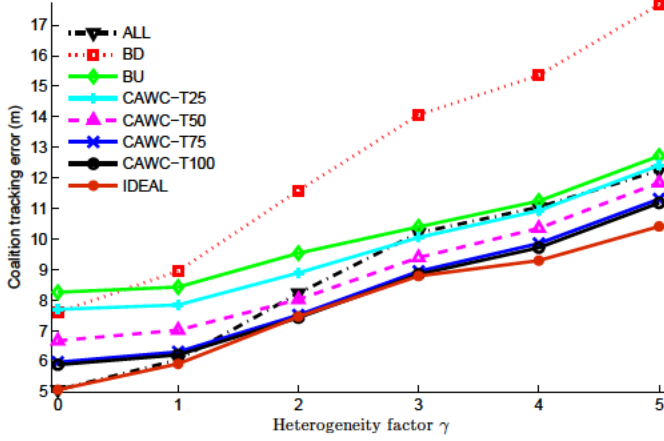


Figure 5. Comparison of tracking error versus camera heterogeneity factor using simulated data ($N = 30$). Homogeneous cameras correspond to $\gamma = 0$. KEY. ALL: all $c_i \in \mathcal{L}^k$ [11]; BD: closest-to-target $c_i \in \mathcal{L}^k$ [14]; BU: highest-utility $c_i \in \mathcal{L}^k$ [7]; CAWC-X: proposed Cost-Aware Coalition using subsets of \mathcal{L}^k ($X=\{25,50,75,100\}$); IDEAL: all $c_i \in \mathcal{L}^k$ with known covariance.

$$\kappa_i = 1 + \gamma(1 - f_i)^\gamma \quad (20)$$

where γ is the penalization (i.e. heterogeneity) of the mis-tracking factor $f_i \in [0, 1]$ for each camera view. f_i controls the likelihood to provide unreliable tracking data that simulates problems such as low target-background contrast in each view. We randomly set f_i for each c_i and control the measurement accuracy via γ (i.e. multiplier of the covariance \mathbf{W}). Figure 4(b) shows the measurement penalization κ_i for the camera heterogeneity factors γ used in the simulations. We approximate the camera utility in Eq. 3 via the mistracking factor as $u_i = 1 - f_i$. For the simulations, we use $T = 1$ (one target) and $K = 50$ (time steps). The results are the mean of 50 runs.

2) *Cost-free*: Figure 5 shows the tracking error dependence with the measurement heterogeneity γ for a network of $N=30$ cameras. Under homogeneity ($\gamma = 0$), all cameras generate measurements with equal accuracy and tracking performance depends on the coalition size (the higher the better). ALL and IDEAL outperform other approaches. For $\gamma > 0$, cameras provide measurements with variable accuracy and selecting the best ones is crucial. ALL accumulates errors of low-accuracy measurements as all cameras are equally fused. BD shows that the camera closest to the target does not guarantee accurate measurements. BU uses the best camera providing good performance but it does not benefit from coalition collaboration. CAWC-25, CAWC-50, CAWC-75 and CAWC-100 (CAWC-25:100) consider coalitions where each c_i result is weighted according to its utility. As we increase γ , utility weighting becomes more relevant as shown in the comparison. CAWC-25:100 outperform the state-of-the-art approaches for medium and high heterogeneity ($\gamma > 2$) showing the advantages of the proposed approach to reduce errors. IDEAL is the best, closely followed by the CAWC-25:100, but IDEAL requires the knowledge of each camera measurement covariance $\kappa_i \mathbf{W}$; whereas CAWC-25:100 only use the camera utility as indicator.

Figure 6 shows the relation between tracking error and network density for measurement heterogeneity $\gamma = 2$. By

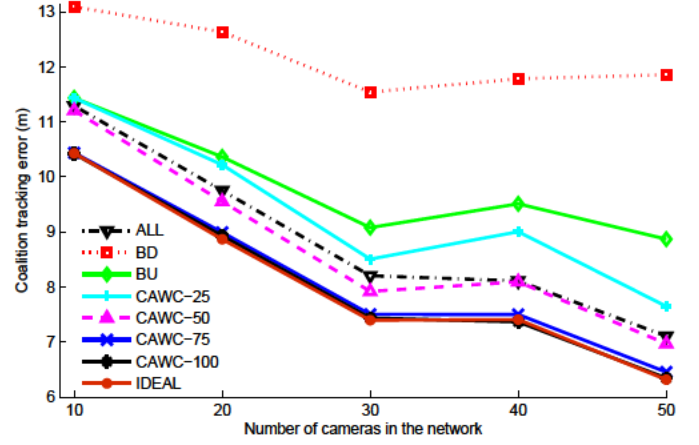


Figure 6. Comparison of tracking error versus the number of cameras in the network ($N=10-50$) for $\gamma = 2$ (camera heterogeneity). KEY. ALL: all $c_i \in \mathcal{L}^k$ [11]; BD: closest-to-target $c_i \in \mathcal{L}^k$ [14]; BU: highest-utility $c_i \in \mathcal{L}^k$ [7]; CAWC-X: proposed Cost-Aware Coalition using subsets of \mathcal{L}^k ($X=\{25,50,75,100\}$); IDEAL: all $c_i \in \mathcal{L}^k$ with known covariance.

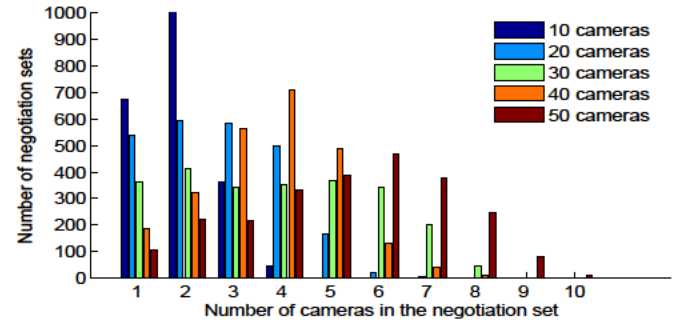


Figure 7. Size of negotiation set \mathcal{L}^k for target tracking in networks with varying numbers of cameras.

increasing the density, targets are viewed by more cameras and coalitions tend to increase their size. BD is the worst demonstrating the limitations of the distance criterion for coalitions. BU shows that the highest-utility camera improves performance for low densities ($N=10-20$) with a stabilization for high densities ($N=30-50$). ALL, CAWC-50:100 and IDEAL reduce error for an increasing number of cameras. CAWC-50:100 error is always below ALL, showing that utility weighting is useful for all camera densities, being a good estimator of the real measurement covariance $\kappa_i \mathbf{W}$. IDEAL and CAWC-100 results are similar as measurement heterogeneity is not high ($\gamma = 2$).

Figure 7 depicts the number of cameras simultaneously viewing targets (i.e. the negotiation set \mathcal{L}^k) for various network densities. Targets in low-density networks ($N=10-20$) are often viewed by sets of 1-4 cameras, where two-camera-sets are frequent. Thus, the benefit of using coalitions is less noticeable (see Figure 6) as the same camera is always employed regardless the applied approach. Medium-density networks ($N=30$) provide sets of 2-5 cameras where coalitions have more choices for camera selection. In high-density networks ($N=40-50$), many cameras observe each target and coalitions decrease the tracking error (see Figure 6). However, such high density may correspond to less common network deployments.

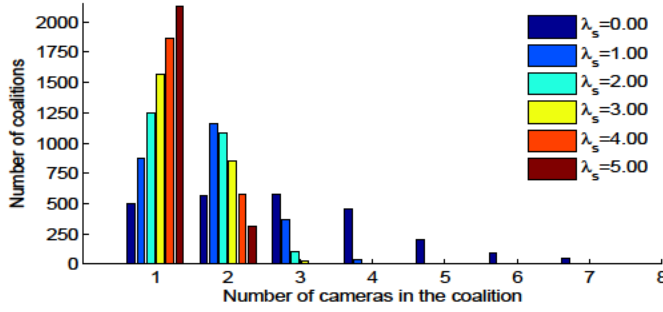


Figure 8. Size of the coalition created by the proposed approach for Cost-Aware Coalitions (CAWC) for various cost factors λ_s ($N = 30$ and $\gamma = 2$).

3) *Cost-aware*: Figure 8 shows the coalition size for different marginal costs. Smaller coalitions are created for increasing λ_s as the cost associated to resources of additional cameras is more important than their improvements. The maximum coalition utility needs only few cameras. Extreme cases are for $\lambda_s = 0$ (cost-free) and $\lambda_s = 5$ (high-cost) where all cameras and the best one are in the coalition, respectively.

Figure 9 compares the mean coalition size, performance (tracking error) and resources (energy saved) of the selected approaches. In Figure 9(a), BF and CAWC approximately generate equal-size coalitions, decreasing their size as resources are more limited. Between the two coalition costs, processing predominates over communication as the energy required is considerable higher for the former. Coalitions based on communication costs do not significantly differ from the cost-free situation (ALL) with a maximum difference of 0.4 cameras for $\lambda_s = 5$. For Figures 9(b) and 9(c), ALL always employs the same number of cameras, being not affected by the available resources. Hence, ALL is used as baseline result for performance and energy. Figure 9(b) shows that the tracking error of CAWC and BF gradually grows as marginal costs λ_s increase. The tracking error has a maximum deviation of $\sim 20\%$ for $\lambda_s = 5$ where the energy savings are $\sim 61\%$, as compared with the baseline (ALL). For $\lambda_s = 1$, both BF and CAWC generate coalitions that increase the tracking error $\sim 7\%$ but require $\sim 38\%$ less energy. Coalitions aware of communication costs slightly reduce performance by $\sim 3\%$ and increase energy savings by $\sim 10\%$. The processing cost dominates over the communication one to estimate the coalition cost during the creation process. The performance drop can be modified by changing λ_s , which depends on the application considered and the available resources (e.g. energy budget for communication and processing).

Although BF and the CAWC-ALL have similar tracking error and cost, their computational complexity is different. Figure 10 shows the needed iterations (negotiations) for various network sizes. There is no significant difference for smaller networks ($N=10-20$). However, BF cost exponentially grows for larger networks ($N=30-50$). Instead, CAWC-ALL increases linearly the number of iterations and requires less computations.

These results suggest a performance-energy tradeoff exhibited by varying the marginal cost weight λ_s . Figure 11 shows this relation for different network densities and marginal costs λ_s . The results depict the difference between the proposed approach CAWC (cost-aware) and ALL (cost-free, using all

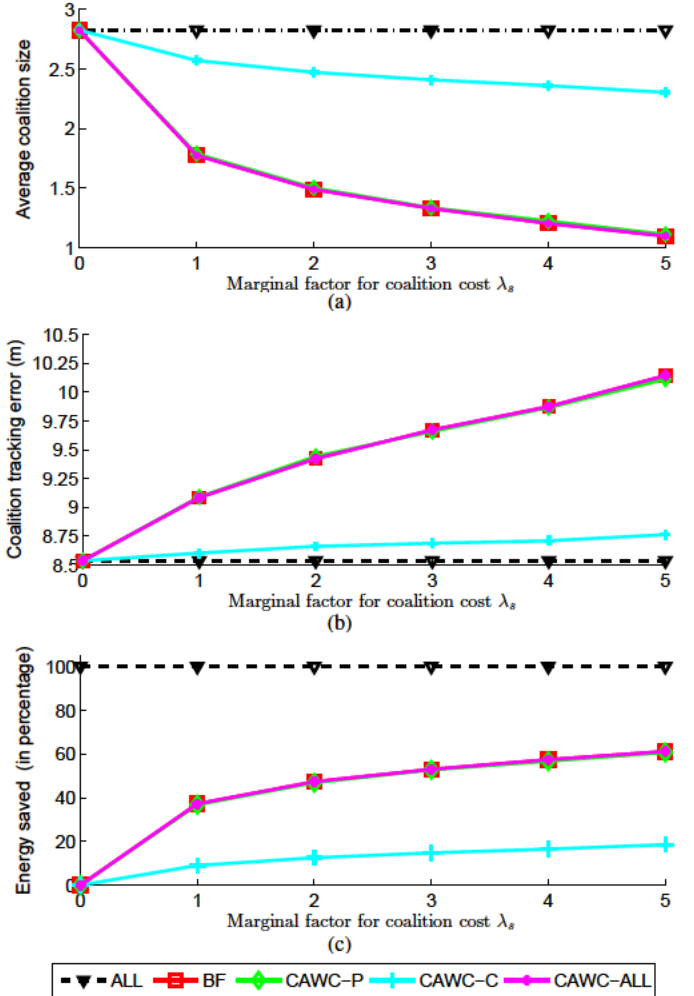


Figure 9. Comparison of coalition (a) sizes, (b) tracking error and (c) saved energy (ALL is used as baseline) for varying coalition costs λ_s ($N = 30$). KEY: ALL: all $c_i \in \mathcal{L}^k$ (cost-free); BF: Brute-Force search; CAWC-P: proposed Cost-Aware Coalition using Processing cost only; CAWC-C: CAWC using Communication cost only; CAWC-ALL: CAWC using both costs.

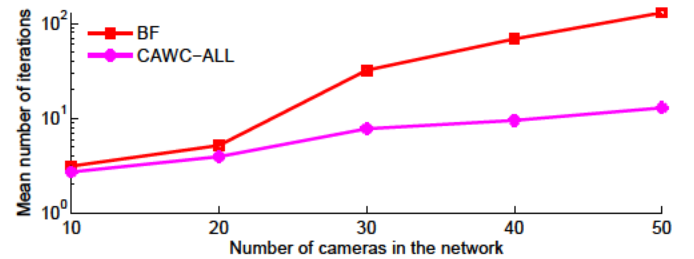


Figure 10. Comparison of the number of iterations (negotiations) to create coalitions. KEY: BF: Brute-Force search; CAWC-ALL: proposed Cost-Aware Coalition creation using processing and communication costs.

cameras) in terms of increased tracking error and saved energy. Increasing the number of cameras leads to higher energy savings. The energy is approximately reduced tenfold when considering 10 and 50 camera networks as coalitions only employ the most useful cameras. Conversely, the tracking error grows when the marginal costs increase. Less cameras are used to employ less resources and, therefore, the coalition is more sensitive to errors in camera measurements. By decreasing marginal costs, more cameras are used and low-accuracy

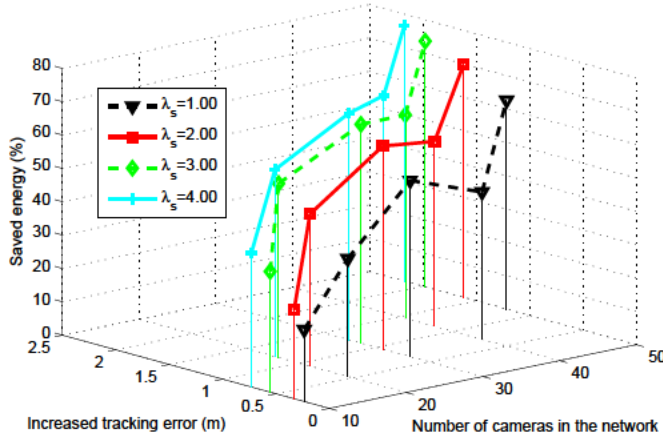


Figure 11. Performance comparison (energy saved and increased tracking error) for different network sizes and marginal costs λ_s . Results compare the proposed approach against all negotiating cameras (i.e. cost-free coalitions).

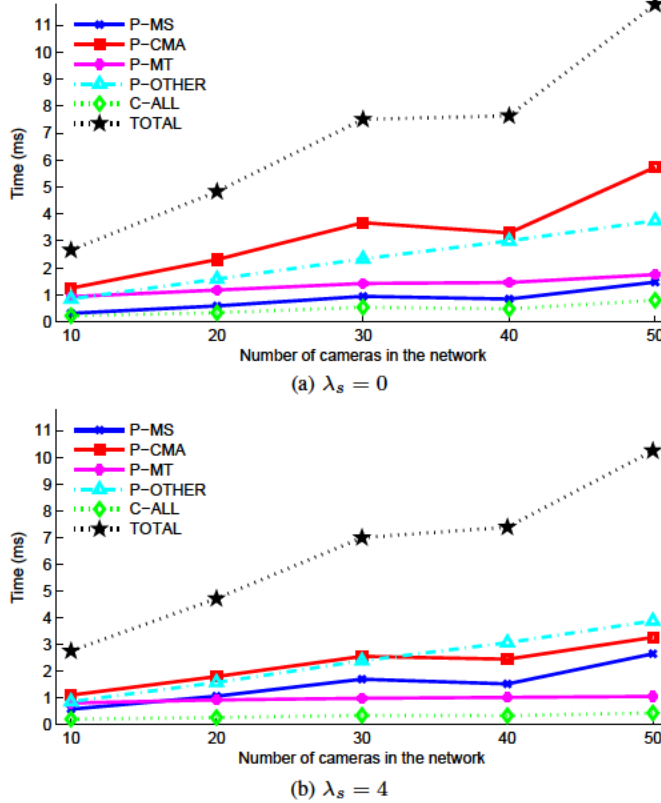


Figure 12. Execution time for the tasks of the proposed approach in (a) cost-free and (b) cost-aware networks of varying size. KEY. P-MS: Processing for Manager Selection; P-CMA: Processing for Camera Allocation; P-MT: Processing for Manager Tracking; P-OTHER: Other Processing tasks; C-ALL: Communication tasks; TOTAL: accumulated time.

measurements are better filtered in the coalition. Note that the increased tracking error for $N=10$ is less than for $N > 10$ due to the reduced possibilities of creating coalitions as negotiation sets often contain one camera only (see Figure 7).

Figure 12 shows the computational cost of CAWC for varying network size and cost settings. The processing time is measured via a Matlab implementation and a standard PC (P-IV 2Ghz 4GB-RAM). For the communication time, we consider standard wireless protocols such as 802.11a (11

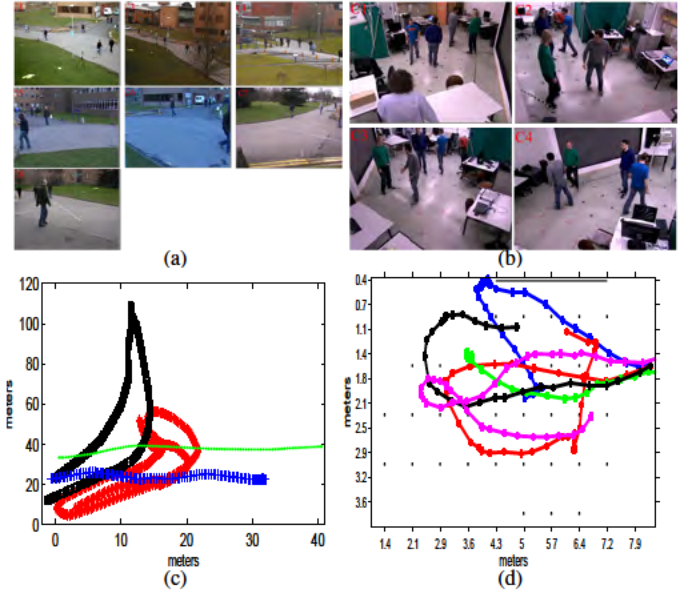


Figure 13. Setup for real data experiments. Visual samples of available cameras for (a) frame 198 of *S2_L1* (PETS2009) and (b) frame 300 of *chap* (ICGLab6). Top view of selected trajectories are represented for (c) PETS2009 (four targets) and (d) ICGLab6 (five targets) datasets.

Mbit/s) without overhead. The amount of exchanged data depends on $|\mathcal{L}^k|$ as each c_i sends a 4-double packet M_i^k . For each allocation round, c_m^k sends a 3-double packet (type, destination and bid) and the receiver c_i replies with a 2-double packet. Other packets include a 2-double packet to announce c_m^k and a 12-double packet to distribute the estimates [11]. Results show that the time increases with growing network sizes and decreases with increasing λ_s . The total processing time range is 2.66-11.77 ms, which is suitable for real-time.

C. Real data

1) *Dataset*: We use nine targets from two multi-camera datasets: *S2_L1* from PETS2009¹ (seven cameras) and *chap* from ICGLab6²[36] (four cameras). This data represents many tracking challenges such as abrupt motion, occlusions, varying illumination and scale changes. We only analyze targets visible in (at least) three cameras for more than 100 frames. Figure 13 shows visual examples and the trajectories of the used targets.

To get the target measurements for each frame (k step), each camera runs a multi-hypothesis tracker based on sparse appearance modeling [37] and an online utility estimator based on tracking performance [38], both tuned to each camera view. As other approaches [7][11], we manually initialize each target when it appears in the camera FOV. Targets are tracked until leaving the FOV of all cameras or until the failure of all the associated trackers. The length of the sequences ranges from 200 to 500 frames (making a total of ~ 9500 frames analyzed). For comparison, manual annotations of the ideal result in the ground-plane are used every frame (PETS2009) or every ten frames (ICGLab6). The results are the mean of 5 runs.

¹<http://pets2009.net/>

²<http://irs.icg.tugraz.at/download.php>

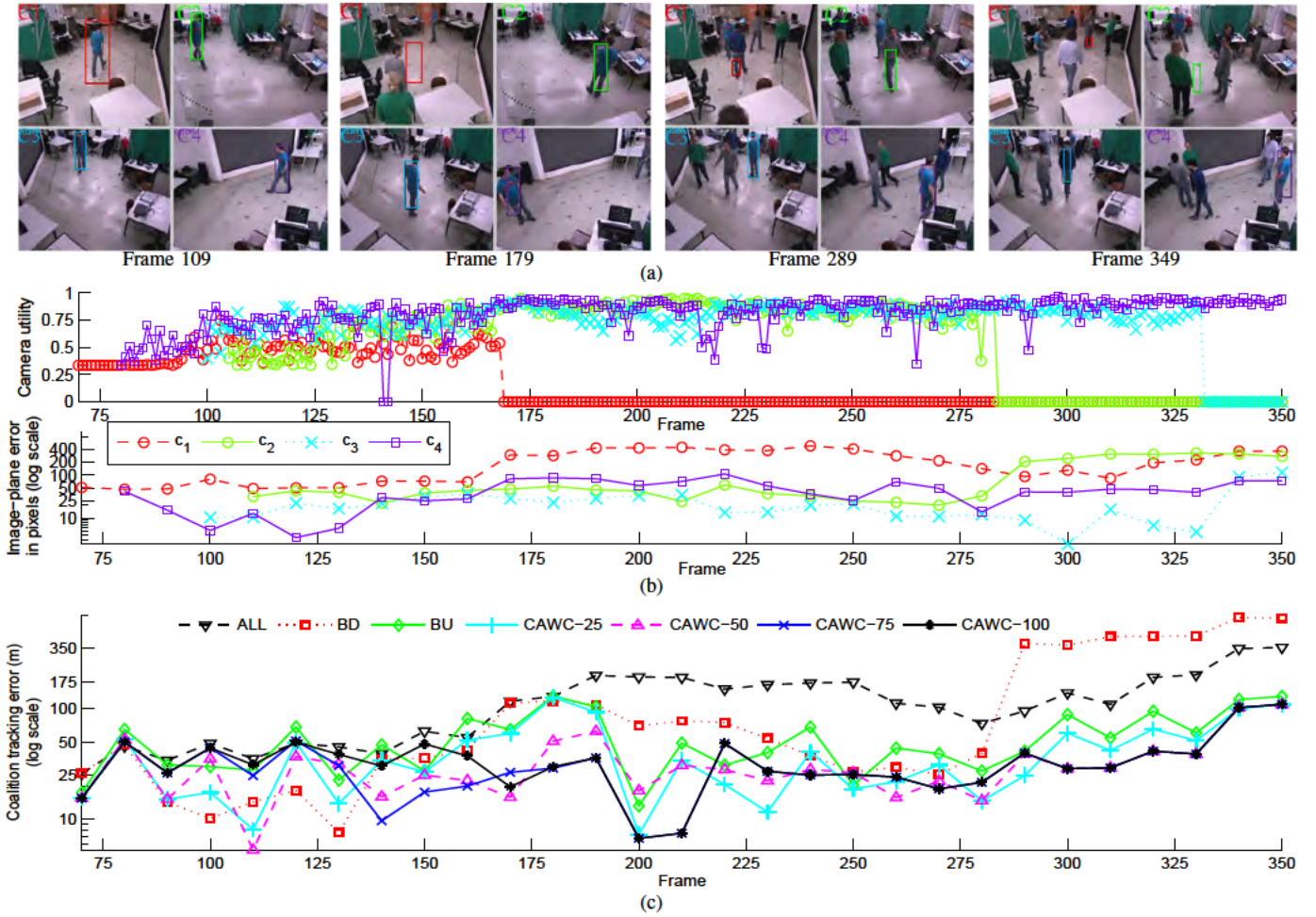


Figure 15. Results for the T0 target of *chap* (ICGLab6). (a) Visual examples; (b) camera results: utility (top) and tracking error (bottom); (c) coalition tracking error. KEY. ALL: all $c_i \in \mathcal{L}^k$ [11]; BD: closest-to-target $c_i \in \mathcal{L}^k$ [14]; BU: highest-utility $c_i \in \mathcal{L}^k$ [7]; CAWC-X: proposed Cost-Aware Coalition using subsets of \mathcal{L}^k ($X=\{25,50,75,100\}$).

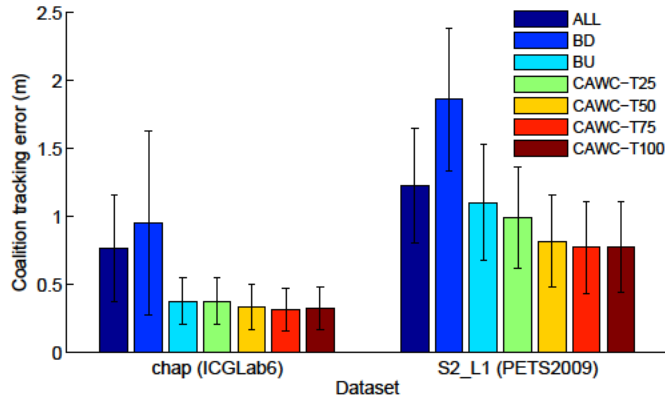


Figure 14. Performance comparison for real data in cost-free settings. Data corresponds to the mean and variance of the selected targets for each dataset. KEY. ALL: all $c_i \in \mathcal{L}^k$ [11]; BD: closest-to-target $c_i \in \mathcal{L}^k$ [14]; BU: highest-utility $c_i \in \mathcal{L}^k$ [7]; CAWC-X: proposed Cost-Aware Coalition using subsets of \mathcal{L}^k ($X=\{25,50,75,100\}$).

2) *Cost-free*: Figure 14 shows the tracking performance for each dataset. *The conclusions are similar to those obtained with the simulated data.* For both datasets, CAWC gets lower error than competitive approaches as each camera contribution is weighted by utilities. The use of multiple cameras increases

the robustness against isolated tracking failures by providing different viewpoints. BD presents the worst results as they rely on cameras that may appear to be close to the target due to errors. BU has different results for the two datasets. In ICGLab6, the error is slightly higher than the proposed approaches as all trackers tend to provide good results until failure. However, trackers are less accurate for PETS2009 and therefore their utility has a larger variability. ALL exhibits a dependence with the number of cameras. The more cameras are combined, the better the results as the chance of integrating cameras with good performance is higher. Note that IDEAL is not applied to real data since we do not know in advance the effect of the tracking problems on the final performance.

Figure 15 shows selected visual results for the T0 target of the ICGLab6 dataset and the results in terms of estimated utility and tracking errors (camera and coalition in the image and ground planes, respectively). At the beginning (frames 70-170), the trackers in the four cameras show good performance and the utility estimator is able to weight their accuracy. For example, c_1 and c_2 have the lowest utility (highest error). Then, the tracker in c_1 loses the target around frame 170 and its utility drops to zero, which prevents its use by the proposed approaches (CAWC-25:100) and BU. Such failure

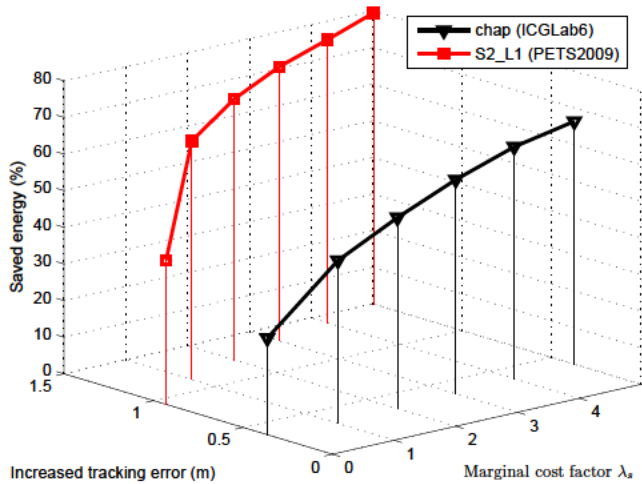


Figure 16. Performance comparison (energy saved and increased tracking error) for different marginal costs λ_s using real data. Results compare the proposed approach against all negotiating cameras (i.e. cost-free coalitions).

influences the performance of ALL. BD is also distracted by a wrong estimation in c_1 , being frequently selected as the closest camera to the target. Similarly occurs for frames 280 and 330 for c_2 and c_3 , respectively. BD and ALL accumulate the camera failures and significantly increase the error. BU and CAWC-25:100 present the lowest error with a slightly reduction as more cameras are employed. *This example shows the benefits of utilities for coalition-based tracking.*

3) *Cost-aware:* Figure 16 shows the tradeoff between tracking error and resource consumption using real data. For both datasets, a significant reduction is achieved by using high-utility cameras. The energy saved ranges from 20% to 80%, which implies stopping the trackers if their utility indicates that they are no longer useful in the coalition (e.g. in case of failures). However, such savings imply a penalization in tracking performance, noticeable for the PETS2009 dataset. The error increases around 50% (from 0.92 to 1.5 meters) and depending on the application, this error might be critical. On the other hand, the tracking error remains constant for the ICGLab6 dataset. As cameras are close enough to targets and provide high-quality views, the coalition gets similar performance with small (high λ_s) or large (low λ_s) coalition sizes. These results are expected as CAWC-25:100 obtained similar performance for cost-free real data (see Figure 14). Thus, cost-aware multi-camera networks may consider the camera-target distance to select views, where utility estimation of close views can lead to good performance at a reduced cost. For coalitions combining different viewing distances, a cost reduction can be achieved that decreases tracking performance.

D. Discussion

The experimental results showed the importance of camera accuracy indicators, such as the camera utility, for collaborative processing via coalitions that improves tracking performance. In fact, cameras with heterogeneous views may provide data with variable accuracy whose weighted integration is key to increase tracking performance. High-density camera

networks reduce tracking error as multiple views are available for each target. Results under limited resources exhibit a performance-cost tradeoff. By increasing the cost of the coalition (i.e. including more cameras), tracking performance increases. However, network resources may be quickly depleted, thus reducing network lifetime. Results also show that the energy consumption can be significantly reduced at the cost of a small performance decrease. Finally, the processing cost dominates over the communication cost as few data packets are exchanged between cameras in the proposed approach.

VII. CONCLUSIONS

We presented a framework to create camera coalitions that considers task-performance improvement and costs for camera cooperation. A dynamically selected coalition manager enables decentralized coordination for cooperative tracking where coalition members communicate only with their local fusion centers. We defined a decentralized negotiation scheme to allocate resources (cameras) to coalitions over time. The proposed approach was demonstrated on simulations and real data for target tracking in cost-free and resource-constrained camera networks, outperforming state-of-the-art approaches. The proposed approach is not restricted to decentralized tracking and it can be applied to other collaborative tasks in camera networks and to alternatives schemes such as token-passing using Particle Filters [39].

As future work, we will explore distributed approaches for simultaneous resource allocation in multi-target scenarios.

REFERENCES

- [1] H. Aghajan and A. Cavallaro, *Multi-camera networks: principles and applications*. Academic press, 2009.
- [2] Y. Charfi, N. Wakamiya, and M. Murata, "Challenging issues in visual sensor networks," *IEEE Wireless Commun.*, vol. 16, no. 2, pp. 44–49, Apr. 2009.
- [3] C. Micheloni, B. Rinner, and G. L. Foresti, "Video analysis in Pan-Tilt-Zoom camera networks," *IEEE Signal Processing Mag.*, vol. 27, no. 5, pp. 78–90, Sep. 2010.
- [4] A. Roy-Chowdhury and B. Song, *Camera networks: the acquisition and analysis of videos over wide areas*. Morgan & Claypool Publishers, 2012.
- [5] M. Taj and A. Cavallaro, "Distributed and decentralized camera tracking," *IEEE Signal Processing Mag.*, vol. 28, no. 3, pp. 46–58, May 2011.
- [6] L. Liu, X. Zhang, and H. Ma, "Optimal node selection for target localization in wireless camera sensor networks," *IEEE Trans. on Vehicular Technology*, vol. 59, no. 7, pp. 3562–3576, Sep. 2010.
- [7] Y. Li and B. Bhanu, "Utility-based camera assignment in a video network: A game theoretic framework," *IEEE Sensors J.*, vol. 11, no. 3, pp. 676–687, Mar. 2011.
- [8] M. Alaei and J. Barcelo-Ordinas, "A collaborative node management scheme for energy-efficient monitoring in wireless multimedia sensor networks," *Wireless Networks*, vol. 19, no. 5, pp. 639–659, Aug. 2013.
- [9] A. Ercan, A. El Gamal, and L. Guibas, "Object tracking in the presence of occlusions using multiple cameras: A sensor network approach," *ACM Trans. Sens. Netw.*, vol. 9, no. 2, p. 36, Mar. 2013.
- [10] L. Tessens, M. Morbée, H. Aghajan, and W. Philips, "Camera selection for tracking in distributed smart camera networks," *ACM Trans. Sens. Netw.*, vol. 10, no. 2, pp. 1–36, Jan. 2014.
- [11] H. Medeiros, J. Park, and A. Kak, "Distributed object tracking using a cluster-based Kalman filter in wireless camera networks," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 4, pp. 448–463, Aug. 2008.
- [12] F. Qureshi and D. Terzopoulos, "Multi-camera control through constraint satisfaction for persistent surveill," in *IEEE Int. Conf. on Advanced Video and Signal-based Surveill.*, Sep. 2008, pp. 211–218.

- [13] J. Yoder, H. Medeiros, J. Park, and A. Kak, "Cluster-based distributed face tracking in camera networks," *IEEE Trans. Image Process.*, vol. 19, no. 10, pp. 2551–2563, Oct. 2010.
- [14] L. Esterle, P. Lewis, X. Yao, and B. Rinner, "Socio-economic vision graph generation and handoff in distributed smart camera networks," *ACM Trans. Sens. Netw.*, vol. 10, no. 2, pp. 1–24, Jan. 2014.
- [15] C. Zopounidis and P. M. Pardalos, *Handbook of multicriteria analysis*. Springer-Verlag, 2010.
- [16] Y. Yao, C.-H. Chen, A. Koschan, and M. Abidi, "Adaptive online camera coordination for multi-camera multi-target Surveillance," *Comput. Vis. Image Understand.*, vol. 114, no. 4, pp. 463–474, Apr. 2010.
- [17] W. Li and W. Zhang, "Sensor selection for improving accuracy of target localisation in wireless visual sensor networks," *IET Wireless Sensor Systems*, vol. 2, no. 4, pp. 293–301, Dec. 2012.
- [18] M. Magno, F. Tombari, D. Brunelli, L. Di Stefano, and L. Benini, "Multimodal video analysis on self-powered resource-limited wireless smart camera," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 3, no. 2, pp. 223–235, Jun. 2013.
- [19] S. Aziz and D. Pham, "Energy efficient image transmission in wireless multimedia sensor networks," *IEEE Commun. Lett.*, vol. 17, no. 6, pp. 1084–1087, Jun. 2013.
- [20] Y. Fu, Q. Ling, and Z. Tian, "Distributed sensor allocation for multi-target tracking in wireless sensor networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 4, pp. 3538–3553, Oct. 2012.
- [21] J. Lin, F. Lewis, W. Xiao, and L. Xie, "Energy-efficient distributed adaptive multisensor scheduling for target tracking in wireless sensor networks," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 6, pp. 1886–1896, Jun. 2009.
- [22] J. Park, P. Bhat, and A. Kak, "A look-up table based approach for solving the camera selection problem in large camera networks," in *Int. Workshop on Distributed Smart Cameras (DCS)*, Oct. 2006, pp. 1–5.
- [23] S. Lin, J. Zhang, G. Zhou, L. Gu, J. Stankovic, and T. He, "ATPC: Adaptive transmission power control for wireless sensor networks," in *Int. Conf. on Embedded Networked Sensor Syst.*, 2006, pp. 223–236.
- [24] P. Fishburn, *Utility Theory for Decision Making*. Huntington, NY: Robert E. Krieger, 1970.
- [25] C. Ding, B. Song, A. Morye, J. Farrell, and A. Roy-Chowdhury, "Collaborative sensing in a distributed PTZ camera network," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3282–3295, Jul. 2012.
- [26] A. Mavrinac and X. Chen, "Modeling Coverage in Camera Networks: A Survey," *Int. Journal of Computer Vision*, vol. 101, no. 1, pp. 205–226, Jan. 2013.
- [27] A. Mutambara, *Decentralized estimation and control for multisensor systems*. CRC press, 1998.
- [28] M. Baidas and A. Mackenzie, "Altruistic coalition formation in cooperative wireless networks," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4678–4689, Nov. 2013.
- [29] N. Karamchandani, P. Minero, and M. Franceschetti, "Agile broadcast services: Addressing the wireless spectrum crunch via coalitional game theory," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 794–808, Feb. 2014.
- [30] M. Ebdan and S. Roberts, "Graph marginalization for rapid assignment in wide-area Surveillance," *Ad Hoc Networks*, vol. 9, no. 2, pp. 180–188, Mar. 2011.
- [31] A. Wang and A. Chandrakasan, "Energy-efficient DSPs for wireless sensor networks," *IEEE Signal Processing Mag.*, vol. 19, no. 4, pp. 68–78, Jul. 2002.
- [32] T. Sandholm, "Agents in electronic commerce: Component technologies for automated negotiation and coalition formation," *Auton. Agents and Multi-Ag. Syst.*, vol. 3, no. 1, pp. 73–96, Mar. 2000.
- [33] —, "An implementation of the contract net protocol based on marginal cost calculations," in *National Conf. on Artificial Intelligence (AAAI)*, 1993, pp. 256–264.
- [34] A. Kamal, J. Farrell, and A. Roy-Chowdhury, "Information weighted consensus filters and their application in distributed camera networks," *IEEE Trans. Autom. Control*, vol. 58, no. 12, pp. 3112–3125, Dec. 2013.
- [35] B. Dieber, C. Micheloni, and B. Rinner, "Resource-aware coverage and task assignment in visual sensor networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 10, pp. 1424–1437, Oct. 2011.
- [36] H. Possegger, S. Sternig, T. Mauthner, P. Roth, and H. Bischof, "Robust real-time tracking of multiple objects by volumetric mass densities," in *IEEE Int. Conf. on Computer Vision and Pattern Recog. (CVPR)*, 2013, pp. 2395–2402.
- [37] D. Wang, H. Lu, and M.-H. Yang, "Online object tracking with sparse prototypes," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 314–325, Jan. 2013.
- [38] J. SanMiguel, A. Cavallaro, and J. Martinez, "Adaptive online performance evaluation of video trackers," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2812–2823, May 2012.
- [39] O. Hlinka, F. Hlawatsch, and P. Djuric, "Distributed particle filtering in agent networks: A survey, classification and comparison," *IEEE Signal Processing Mag.*, vol. 30, no. 1, pp. 61–81, Jan. 2013.



Juan C. SanMiguel is assistant professor at Universidad Autónoma de Madrid (Spain). He received the M.S. degree in Electrical Engineering (Ingeniero de Telecomunicación degree) in 2006 and the Ph.D. in Computer Science and Telecommunication in 2011, both at Universidad Autónoma de Madrid. From June 2013 to June 2014, he was a postdoctoral researcher at Queen Mary University of London (UK) under a Marie Curie IAPP fellowship. He has participated in several national and international projects dealing with multimedia content trans-

mission, human-object interactions recognition, visual tracking and video surveillance. He also serves as a reviewer for several international Journals and Conferences. His current research interests are focused on probabilistic object tracking, performance evaluation and multi-camera activity understanding.



Andrea Cavallaro is Professor of Multimedia Signal Processing and Director of the Centre for Intelligent Sensing at Queen Mary University of London, UK. He received his Ph.D. in Electrical Engineering from the Swiss Federal Institute of Technology (EPFL), Lausanne, in 2002 and the Laurea (Summa cum Laude) in Electrical Engineering from the University of Trieste in 1996. He was a Research Fellow with British Telecommunications (BT) in 2004/2005 and was awarded the Royal Academy of Engineering teaching Prize in 2007; three student paper awards

on target tracking and perceptually sensitive coding at IEEE ICASSP in 2005, 2007 and 2009; and the best paper award at IEEE AVSS 2009. Prof. Cavallaro is Area Editor for the IEEE Signal Processing Magazine; and Associate Editor for the IEEE Transactions on Image Processing. He is an elected member of the IEEE Signal Processing Society, Image, Video, and Multidimensional Signal Processing Technical Committee. Prof. Cavallaro was General Chair for IEEE/ACM ICDSC 2009, BMVC 2009, M2SFA2 2008, SSPE 2007, and IEEE AVSS 2007; and Technical Program chair of IEEE AVSS 2011; the European Signal Processing Conference (EUSIPCO 2008) and of WIAMIS 2010. He has published more than 130 journal and conference papers, and four books: Multi-camera networks (2009), Elsevier; Video tracking (2011), Wiley; Analysis, retrieval and delivery of multimedia content (2012), Springer and Intelligent multimedia surveillance (2013), Springer.