Research Article

Game-Theoretic Camera Selection Using Inference Tree Method for a Wireless Visual Sensor Network

Yeong-Jae Choi,[1] Go-Wun Jeong,[1] Yong-Ho Seo,[2] and Hyun S. Yang[1]

[1] *Department of Computer Science, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea*
[2] *Department of Intelligent Robot Engineering, Mokwon University, 88 Doanbuk-ro, Seo-gu, Daejeon 302-729, Republic of Korea*

Correspondence should be addressed to Yong-Ho Seo; yhseo@mokwon.ac.kr

In a wireless visual sensor network consisting of wireless, battery-powered, and field-of-view (FoV) overlapping and stationary visual sensors, trade-offs exist between extending network lifetime and enhancing its sensing accuracy. Moreover, aggregating individual inferences from each sensor is essential to generate a globally consistent inference, because these individual inferences can be biased by noise or other unexpected conditions. Those challenges can be addressed by reducing the amount of data transmission among the sensors and by activating, in a timely manner, only a desirable camera subset for given targets. In this paper, we initialize an optimal data transmission path among visual sensors using the inference tree method, which is vital for collecting individual inferences and building a global inference. Based on the optimal data transmission path, we model the camera selection problem in a cooperative bargaining game. In this game, based on the serial dictatorial rule, camera sensors cooperatively attempt to raise the overall sensing accuracy by sequentially deciding their own mode between "sleep" and "active" in descending order of their bargaining power. Simulated results demonstrate that our proposed approach outperforms other alternatives, resulting in reduced resource overhead and improved network lifetime and sensing accuracy.

1. Introduction

There has been an increasing necessity to extract relevant information for multiple targets moving around inside wide areas for surveillance purposes. Moreover, these requirements must be fulfilled in a cost-efficient manner. A visual sensor is primarily equipped with an image sensing device, several processing units, communication facilities, and a set of batteries. This composition is very suitable for surveillance, because of the advantageous characteristics such as a wide monitoring area, rich visual information, and human-friendly data. Following the development of these inexpensive, powerful, and easily-deployable visual sensors, wireless visual sensor networks (WVSN) consisting of wireless, battery-powered, and field-of-view (FoV) overlapping and stationary visual sensors have been widely employed for surveillance in public places [1, 2]. Compared with other types of wireless sensors, visual sensors are impacted more by their limited bandwidth, lifespan, computation, and storage capabilities, because they contend with high-dimension data sets containing rich information generated from images [3]. Thus, it will be necessary to initialize an optimal data transmission path to reduce the amount of data transmission among the sensors for a global inference and to efficiently activate only selected cameras, which optimizes their collective coverage of given targets in a timely manner. The latter is referred to as *Camera Selection* (CS).

In this paper, we initialize an optimal data transmission path among visual sensors utilizing the inference tree method, which is a key component in aggregating individual inferences and building a global inference with minimized transmissions [2]. Based on the optimal transmission path, every visual sensor can exchange data with other sensors. Additionally each sensor can autonomously switch its mode between "sleep" (in *sleep* mode, the sensor stops capturing data; it will continue to transmit data) and "active" only with local knowledge, during advanced target analysis beyond basic tracking. This local rationale can be feasible under the practical assumption that FoV overlapping cameras can directly communicate with each other; additionally, the view

(a) Configuration   (b) Physical topology   (c) Logical topology
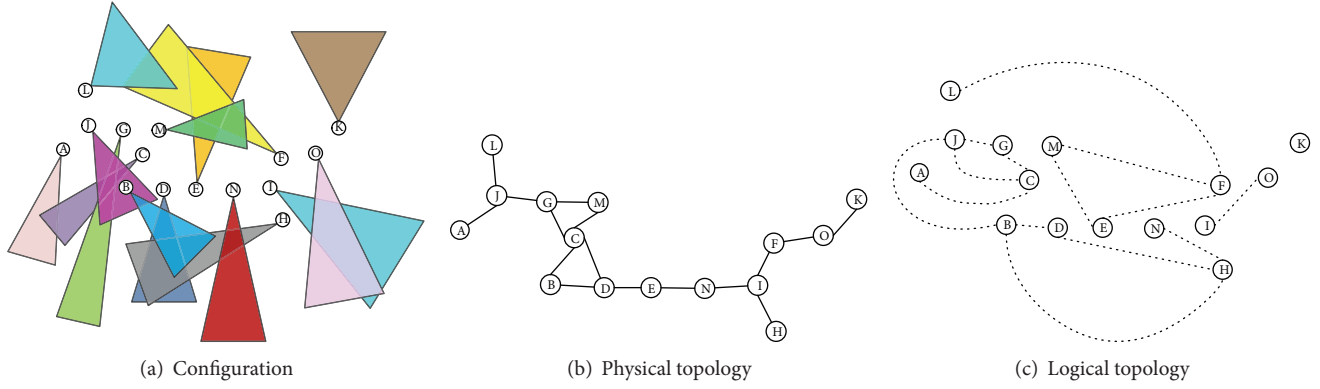
Figure 1: A wireless visual sensor network.

of a target is shared only between neighboring cameras. Thus, the camera's local knowledge is sufficient to measure its sensing contribution towards the global sensing accuracy, considering its neighbors' contributions. As discussed in [4, 5], each of the multiple cameras can cooperatively bargain for an optimal collective coverage. Our proposed approach utilizes the serial dictatorial rule, in which preferred cameras are prioritized to select their mode in earlier steps to achieve efficient computation.

The remainder of this paper is organized as follows. Section 2 introduces and discusses several advanced CS solutions; Section 3 describes the inference tree method used to initialize the data transmission path; Section 4 models a CS problem in a cooperative game; Section 5 describes our proposed serial dictatorial rule-based solution to force every camera to select its optimal mode using its local knowledge; Section 6 simulates and analyzes our approach in several network performance metrics and resource overhead; Section 7 compares our solution with its representative alternatives employing our concerned metrics; and Section 8 concludes the paper.

2. Related Work

Because of the lack of CS-related studies at this time, we note that several considerable research efforts have proposed comparable solutions, mainly based on greedy selection (GS) or potential game (PG).

The GS-based approaches select a camera best satisfying the criteria of interest at the time until l cameras are selected for a certain l that has been heuristically determined in advance. For example, a candidate camera may be selected based on the extent to which it improves the current visual hull and thus reduces occlusion [6, 7]; candidate cameras may also be selected based on the degree to which their images are different from images of already-selected cameras, with a goal of producing varied images that reduce redundancies [8]. Although this selection process may provide richer information about targets, it does not adaptively cope with the dynamics in the targets' locations because of the stationary l.

For seamless tracking, the PG studies assign a camera to every target based on the maximum utility between the

camera and the associated target. A camera's utility for a target is quantified by how large the camera records the target and its face in stationary camera networks [4] or the degree to which the camera can closely observe the target in active camera networks [5]. Until a set of probabilities describing how effectively a target is tracked by a camera for every camera and every target is converged, a set of these utilities and the probability set are alternatingly updated; additionally, they both influence each other. This one-for-one selection could efficiently produce results for tracking or locating given targets; however, it may be insufficient to produce advanced target information, such as multitarget interaction analysis.

All of the discussed techniques have approached a CS solution in a centralized manner, with global knowledge of every camera and every target. To obtain such global knowledge, high bandwidth consumption will necessarily occur at a central operator in centrally controlled networks or at every camera sensor in distributed controlled networks. To avoid such dissipation, our approach employs only limited knowledge; however, it aims to perform similarly to, or improve, the previously mentioned alternatives.

3. Inference Tree Method for Initializing Transmission Path

A WVSN can be described as displayed in Figure 1. Figure 1(a) displays the configuration of a WVSN, Figure 1(b) illustrates a physical topology based on wireless connection reachability (or geometric proximity), and Figure 1(c) represents a logical topology based on the FoV overlapping constraint. In this paper, however, we assume that FoV overlapping cameras can directly communicate with each other. Thus, the physical topology graph can be ignored. This assumption is feasible in practical applications and enables us to focus on the data transmission and camera selection challenges.

As displayed in Table 1 [9], power consumption costs are higher for a broadcasting network compared with a unicasting network. In order to minimize data transmissions for building a global inference in a WVSN, we need to convert the logical topology from a broadcasting/multicasting network to a unicasting network. To initialize a transmission path as

Network	Data	Energy cost per bit (uWs/byte)
Broadcast	Send	2.1
	Receive	0.26
Unicast	Send	0.48
	Receive	0.12

a unicasting network, we use the inference tree method [2]; the process is illustrated in Figure 2. When a logical topology is provided as an input, a weight value is calculated for each edge based on the amount of FoV overlapping between two visual sensors. The result of this initial step is a weighted graph. From the weighted graph, a maximum spanning tree is produced, and the center node of the maximum length path is selected as a root node. Utilizing the root node selected in the second step, the weighted graph is converted into a minimum depth tree employing a breadth-first search algorithm. The result of this step is an initial inference tree. Lastly, it is optimized with up and down actions necessary to build a balanced tree. We utilized the final resulting tree as an optimal data transmission path for the WVSN.

The results from our implementation are displayed in Figure 3. Figure 3(a) illustrates that there is a significant reduction in the number of data transmissions for both the leaf and internal nodes. Figure 3(b) illustrates that energy consumption for data transmissions has decreased for all nodes. Figures 3(a) and 3(b) successfully support that the inference tree method is effective for initializing an optimal data transmission path in WVSN.

4. Cooperative Game for CS

Consider c cameras, indexed by i, statically deployed, and t targets, indexed by j, randomly moving inside a geographical area. As previously stated, we also assume that any two neighboring cameras are able to communicate with each other if their FoVs overlap. The locations of cameras are initially calibrated and stationarily fixed. The locations of targets are updated based on any object localization algorithm of [10], utilizing the most recently recorded images at each time instant, provided to their associated cameras. Whenever new locations are provided, the expected target locations at the next time instant are also estimated by the extended Kalman filter as in [11]. At this point, every camera i is aware of the set of its observable targets to move into its FoV, termed T_i.

Illustration of parameter notations in camera i's FoV with targets j and j' can be described as displayed in Figure 4. Safe region sr_i is the set of every 2D point inside the dotted-line square, where a target is seen observed safely enough. Unsafe distance l_j of j not located in sr_i is the distance from the center of sr_i to j's location. Distinction angle $\theta_{jj'}$ of j and j' is the included angle between the i-to-j vector and the i-to-j' vector.

Declaring that m_i represents the mode of camera i between 0 for sleep (in *sleep* mode, the sensor stops capturing data; it will continue to transmit data) and 1 for active, we modeled our defined CS problem in the form of a classic, normal game given as $\langle player, action, utility \rangle$ in [12]. If function *utility* is identical for every player, the players are encouraged to be cooperative to maximize their shared utility [13]. These conditions are analogous to a CS where the objective, from the point of view of a game designer, is to obtain a minimal set of active cameras that can achieve a high sensing accuracy for given targets, equal to the accuracy provided by an entire network [14]. To enable every camera to autonomously select its mode for the objective, we replace the given game form by $\langle \{i, j\}, m_i, U_g \rangle$, where U_g is the global utility equally shared within the entire network and can be quantified as follows:

$$U_g\left(\{m_i\}\right) = \sum_{j=1}^{t} U_j\left(\{m_i\}_j\right). \tag{1}$$

For $\{m_i\}_j$, the mode set of the cameras, which are able to observe target j, (1) quantifies the global sensing accuracy given t targets by summing the extent to which each target j is well observed by its associated cameras in global target utility U_j. This value generally becomes larger as more cameras are active, but not necessarily. A small number of images may omnidirectionally cover a target; conversely, images that are too similar, produced by closely located and similarly oriented cameras, redundantly dissipate resources to transmit and process these images. By setting the upper bound for global target utilities to 1 to restrict the redundancies as in (3), we observe that the global target utility of target j is constructed by individual target utilities $\{tu_j\}$ obtained by each of the associated cameras as in the following equation:

$$U_j\left(\{m_i\}_j\right) = \sum_{\{m_i\}_j} tu_j\left(m_i\right), \tag{2}$$

such that

$$U_j\left(\{m_i\}_j\right) \leq 1, \tag{3}$$

$$tu_j\left(m_i\right) = m_i \, safe_i\left(j\right) \begin{cases} 1 & \text{if } |T_i| = 1 \\ \min_{j' \in T_i \setminus j} dist_i\left(\theta_{jj'}\right) & \text{if } |T_i| \geq 2, \end{cases} \tag{4}$$

$$safe_i\left(j\right) = \begin{cases} 1 & \text{if } j \in sr_i \\ \dfrac{1}{l_j} & \text{if } j \notin sr_i, \end{cases} \tag{5}$$

$$dist_i\left(\theta_{jj'}\right) = \begin{cases} 1 & \text{if } \theta_{jj'} \geq \dfrac{90}{A_i} \\ \sin\left(A_i \theta_{jj'}\right) & \text{if } \theta_{jj'} < \dfrac{90}{A_i}. \end{cases} \tag{6}$$

The target utility of target j by camera i represents the smallest likelihood that j is sufficiently observed without any occlusion in i's FoV according to i's mode. This likelihood is determined utilizing three values: m_i of whether or not i is
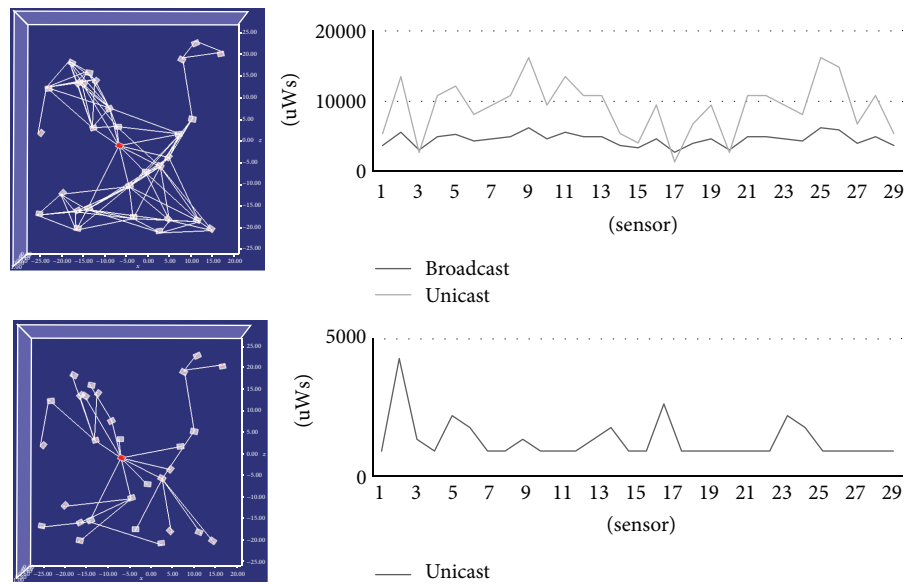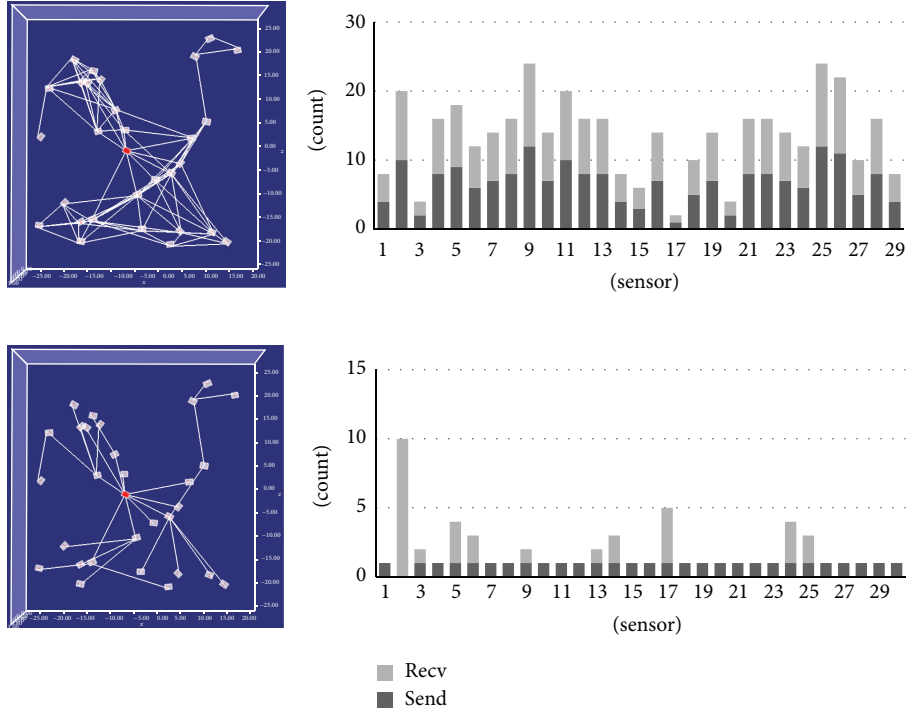
Figure 2: The process of inference tree method.

(a) Message count for 30 visual sensors

(b) Energy consumption using 1 Kbyte packet for 30 visual sensors
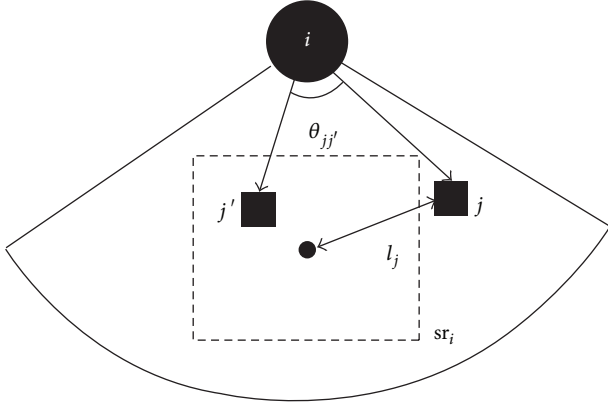
Figure 3: The results of the inference tree method.

FIGURE 4: Illustration of parameter notations in camera i's FoV with targets j and j'. Safe region sr_i is the set of every 2D point inside the dotted-line square, where a target is observed safely enough. Unsafe distance l_j of j not located in sr_i is the distance from the center of sr_i to j's location. Distinction angle $\theta_{jj'}$ of j and j' is the included angle between the i-to-j vector and the i-to-j' vector.

active, $safe_i$ of the degree to which j can be safely observed by i, and $dist_i$ of the degree to which j is occluded by other targets in i's FoV, as in (4). As stated in Figure 4, j can be *safely* observed if it is located in i's safe region sr_i; otherwise, the safety value for j inversely decreases as j's unsafe distance l_j becomes greater as in (5). To measure the occluded degree of j in i's FoV, we define the distinction probability between images simultaneously taken by j and j' as in (6), where θ_{jj}, is the distinction angle of j and j' defined in Figure 4 and A_i is the scaling factor of i, because a greater $\theta_{jj'}$ will result in greater reduction of the occlusion between j and j'. Utilizing these definitions, both U_j and tu_j can have a value between 0 and 1, indicating that even a single camera can provide a thorough observation of a target if the camera safely monitors the target without occlusion, as we have assumed.

Forthwith, we discuss the extent to which an individual camera contributes to a certain global utility. Camera utility of camera i evaluates its contribution degree to observe the set of its observable targets T_i according to m_i by summing the target utilities of every j in T_i as in the following equation:

$$cu_i\left(m_i\right) = \sum_{j \in T_i} tu_j\left(m_i\right). \qquad (7)$$

The equations from (4) to (7) reflect the following characteristics that only camera sensors can possess.

(i) Every camera can be assumed to have, in its FoV, its own safe region where any target is observable in sufficient detail to give desirable information with minimal distortion.

(ii) Owing to the 3D-to-2D projection of imaging, the occlusion among multiple targets in a camera's FoV obstructs the extraction of the targets' information [3, 4].

(iii) The interaction analysis among multiple targets provides more relevant information about the targets beyond their locations [15].

While (i) and (ii) are, respectively, represented by (5) and (6) for (4), more relevant information is provided by cameras with a higher camera utility, which becomes greater as it observes more targets or its associated target utilities are greater, as in (7).

Subsequently, every camera selects its mode to cooperatively maximize their payoff and the global utility U_g, for the least number of active cameras while considering the condition (3). The mode selection process, to be discussed in the following section, is greatly enhanced by taking the camera utilities into account.

5. Serial Dictatorial Rule-Based Bargaining Solution

According to [13], the serial dictatorial rule is a sequence of dictatorial rules conducted by individual players whose exercising order is statically arranged by their bargaining powers. By evaluating camera utilities provided in (7), we consider that a camera has greater bargaining power if it observes a greater number of targets, in a less occluded manner, in the corresponding safe region. Given that all cameras possessing greater bargaining power have already determined their modes and a camera must presently select a mode, it will select the mode maximizing its payoff, U_g, according to the dictatorial rule. The camera is under the assumption that other cameras that have not determined their modes are in sleep mode [14]. This bargaining process is serially performed until every camera determines its mode while communicating with neighboring cameras as follows.

5.1. Order Cameras by Their Camera Utilities. Given estimated locations of T_i, every camera i computes the target utilities of every target in T_i and its own camera utility assuming $m_i = 1$ and subsequently transmits them to its neighboring cameras. Thereafter, i obtains its position in the dictatorial ordering list of it and its neighbors in descending order of their camera utilities, while initializing the mode set for the list, $M_i = \{0\}$.

5.2. Select the Current Mode. Prior to mode selection, every i waits for all the modes of its more bargaining-powerful neighbors to be announced while updating M_i if it is not the first on the list. Otherwise, it instantly assumes its mode by (8) for $M_{i,j}^{m_i}$ which is the target j's associated subset of M_i where only the mode of i is replaced by m_i. Consider

$$m_i = \begin{cases} 1 & \text{if } \sum_{j \in T_i} \left(U_j\left(M_{i,j}^1\right) - U_j\left(M_{i,j}^0\right)\right) > 0 \\ 0 & \text{otherwise.} \end{cases} \qquad (8)$$

A camera i will decide to be active only if it can improve the total of the global target utilities for every target of T_i by its contribution. Specifically, *a camera covers its FoV only*

if at least one target in its FoV is not sufficiently observed at the moment, indicating that it will not be responsible for the targets already well covered by other cameras. Subsequent to this mode decision, i announces its selected mode followed by i; accordingly every neighbor, termed i', updates $M_{i'}$ with i's new mode and drops every j such that $U_j(M_{i',j}) = 1$ obtained by the new $M_{i'}$ from $T_{i'}$ for rapid computation.

This local reasoning with limited knowledge soundly and completely extends maximizing U_g for the following theorem.

Theorem 1. *For every camera i with its mode m_i, it always holds that*

$$U_g\left(M^*\right) - U_g\left(M^{*,-i}\right) \geq \sum_{j \in T_i}\left(U_g\left(M_{i,j}^{m_i}\right) - U_g\left(M_{i,j}^{m_{-i}}\right)\right), \tag{9}$$

where M^* is the bargained mode set for every camera, m_{-i} is the opposite mode of m_i, $M^{*,-i}$ is the mode set where only m_i is replaced by m_{-i} in M^*, and $M_{i,j}$ is i's assumed mode set of i and i's neighbors associated with target j for the bargaining process.

Proof. The following equation (10) holds by the global utility definition and (11) is derived because the change of camera i's mode affects only the global target utilities of every j in T_i. As M_j^* and $M_j^{*,-i}$ are, respectively, restated as $M_{i,j}^{m_i}$ and $M_{i,j}^{m_{-i}}$, we must demonstrate that (12) always holds for every i with T_i for our claim. Consider that

$$U_g\left(M^*\right) - U_g\left(M^{*,-i}\right) = \sum_{j=1}^{t}\left(U_j\left(M_j^*\right) - U_j\left(M_j^{*,-i}\right)\right) \tag{10}$$

$$U_g\left(M^*\right) - U_g\left(M^{*,-i}\right) = \sum_{j \in T_i}\left(U_j\left(M_j^*\right) - U_j\left(M_j^{*,-i}\right)\right), \tag{11}$$

$$\sum_{j \in T_i}\left(U_j\left(M_j^*\right) - U_j\left(M_j^{*,-i}\right)\right) \geq \sum_{j \in T_i}\left(U_j\left(M_{i,j}^{m_i}\right) - U_j\left(M_{i,j}^{m_{-i}}\right)\right). \tag{12}$$

To more easily understand the claim, we restate the target argument as the following if-then rule.

By the nature of our bargaining process, $U_g(M^*) \geq U_g(M^{*,-i})$ always holds for every camera i. Given the condition, (12) also always holds for every i.

Subsequently, we verify this claim for both its soundness and completeness.

Soundness. We demonstrate that (12) and $U_g(M^*) \geq U_g(M^{*,-i})$, respectively, hold in the following two cases.

(a) case of $m_i = 1$ and $m_{-i} = 0$.

When i decides its mode as active while assuming that every mode for the less bargaining-powerful neighboring cameras is sleep, it is believed that it can improve the global target utility of any in T_i. Let us say

that $\{j'\}$ is the target set each global target utility of which is actually raised by i. The difference resulting from i's mode change is given as follows:

$$\sum_{\{j'\}}\sum_{m_{i'} \in M_{i,j'}\setminus m_i} \mathrm{tu}_{j'}\left(m_{i'}\right). \tag{13}$$

Because any of less bargaining-powerful neighbors could be active to contribute to the improvement of the concerned global target utilities, $M_{j'}^*$ is likely to contain an equal or greater number of active cameras than $M_{i,j'}$. Thus, (12) is valid. In addition, (13) is always greater than or equal to 0 by the definition of the target utility, which eventually leads to the conclusion that $U_g(M^*) \geq U_g(M^{*,-i})$ is valid.

(b) Case of $m_i = 0$ and $m_{-i} = 1$.

When i determines its mode as sleep, it believes that every j in T_i is sufficiently covered by more bargaining-powerful neighbors, which derives $\sum_{j \in T_i}(U_j(M_{i,j}^1) - U_j(M_{i,j}^0)) = 0$ by (8). Similarly, it holds that $\sum_{j \in T_i}(U_j(M_j^*) - U_j(M_j^{*,-i})) = 0$ regardless of the other modes in M_j^*. Thus, (12) is valid in this case and $U_g(M^*) = U_g(M^{*,-i})$ is valid, too.

Completeness. We demonstrate that M^* exists, bargained by our bargaining process satisfying (12).

Consider any M^* where every mode is deterministically decided by its associated camera at its order, without possessing actual knowledge of the entire M^*. For every i, regardless of its decided mode, the entire mode set M_j^* for any target j will have more or the same number of active cameras compared to the virtual $M_{i,j}$ because i assumes that its less bargaining-powerful cameras adopt the sleep mode, as we discussed above. Therefore, the difference resulting from its mode change in M_j^* would always be equal to or greater than in $M_{i,j}$. In summary, M^* can be efficiently and deterministically obtained by (8) and any M^* satisfies (12) for every camera. □

Therefore, the selected mode by (8) with limited knowledge for every camera optimizes the global utility by Theorem 1.

6. Simulations and Analyses of Our Approach

In this section, we evaluate our proposed serial dictatorial rule-based game for CS utilizing several network performance metrics in different simulations and quantitatively analyze the complexity required to achieve each step of our approach's design.

6.1. Simulated Performance Analysis. Our proposed approach has been simulated in the following multicamera and multitarget environment.

(a) Dense deployment of $d = 7$ (b) Sparse deployment of $d = 3$
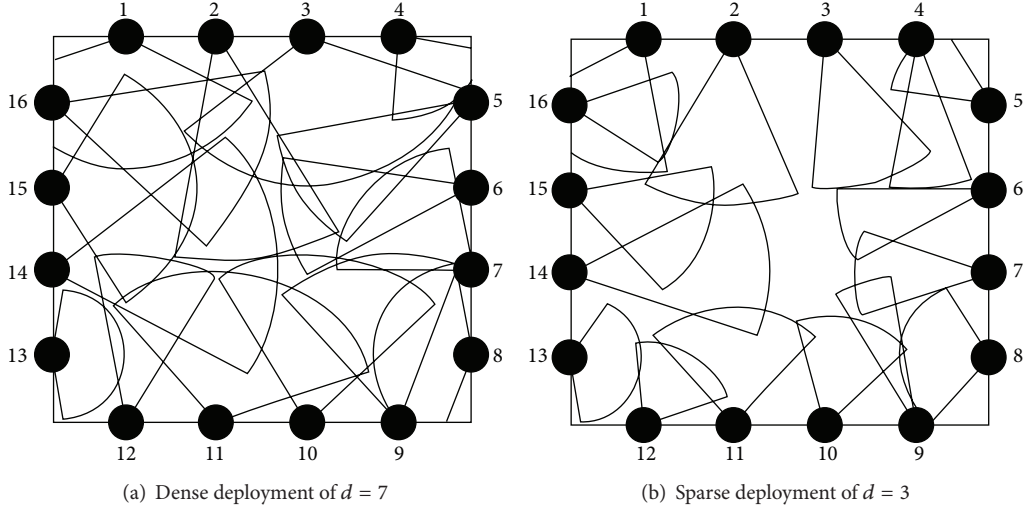
FIGURE 5: Two different camera deployment environments with $d = 7$ of the left and $d = 3$ of the right.

(a) Distantly located targets   (b) Closely located targets   (c) Error case
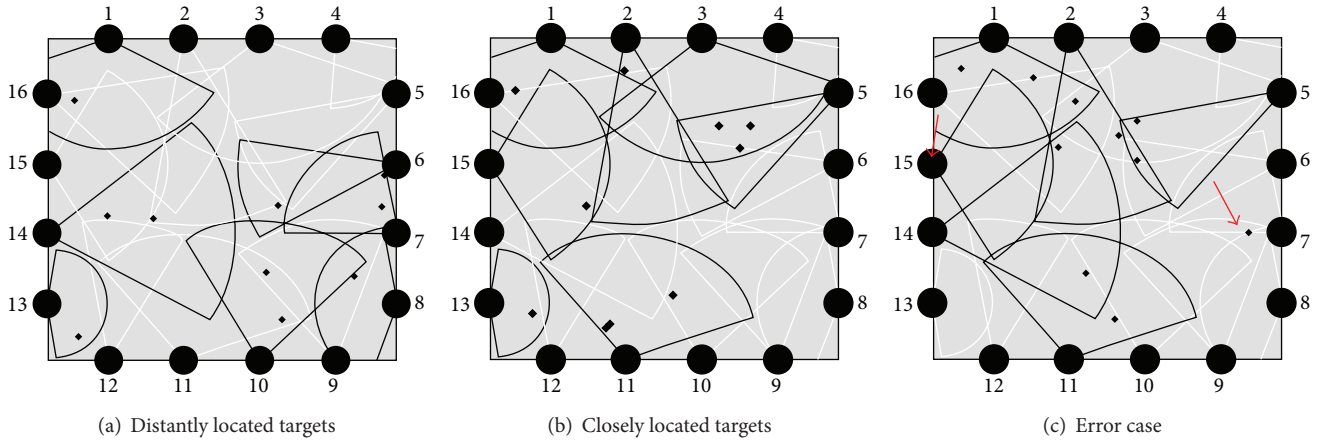
FIGURE 6: Two successful and a single-failed examples of $d = 7$ and 10 targets.

(i) 16 stationary cameras of different FoVs, labeled 1 through 16, are deployed and calibrated in a square-shape area of 230×230 cells as in Figure 5.

(ii) Every camera is provided with its own safe region and FoV center point in advance.

(iii) The scaling factor A_i for every camera i is fixed to 2, because we assume that any two targets having a distinction angle of 45 or more degrees in the camera's FoV can be perfectly and separately monitored.

(iv) Every camera is initially charged with 1000 power bars and only dissipates a bar per time instance if it is active.

(v) The maximum neighboring density for every camera, d, is set to 7 as in Figure 5(a) or 3 as in Figure 5(b).

(vi) In a scenario, 10 or 20 targets freely move inside the area, under a speed of 15 cells per time instance, until the network lifetime has expired.

(vii) The target locations at the previous time instance obtained by any image-based localization of [10] are always announced to their associated cameras. After receiving this location information, every camera computes the expected locations of its observable targets at the current time instance by the extended Kalman filter as in [11]. To add authenticity to the simulation, we implement imperfect localization by adding a small amount of noise to our target localization.

To permit greater focus on the network performance of our interest achieved by our approach, we vary only the simulation situations with different neighboring densities and different target cardinalities, whereas the energy and scaling specifications of the cameras are not changed. In this environment, our approach is evaluated by observing the number of cameras that are active on average, #Active, the number of cameras that are redundantly active, #RActive, the number of

TABLE 2: Simulation results in the four cases of two densities and two target cardinalities.

(d, t)	#Active	#RActive	#Missing	Lifetime
(7, 10)	6.2	0.85	0.32	1364
(3, 10)	5.7	1.2	0.78	1481
(7, 20)	8.9	0.79	0.72	1163
(3, 20)	9.1	0.93	1.5	1249

TABLE 3: Analyzed complexity by each operation.

	Operation	Complexity
Computation	(1) Sensing and processing images	α
	(2) Estimating the current target locations	$O(t^2)$
	(3) Computing utilities	$O(t^2 + td)$
	(4) Selecting the mode	$O(t)$
Communication	(5) Transmitting processed data	β
	(6) Exchanging the target locations	$O(td)$
	(7) Exchanging utilities	$O(td)$
	(8) Exchanging selected modes	$O(td)$

TABLE 4: Simulation results of the five approaches.

(d, t)		(7, 10)			
App.	Ours	5-GS	6-GS	7-GS	PG
#Active	6.2	5	6	7	6.7
#RActive	0.85	0.21	0.75	1.29	0.84
#Missing	0.32	1.5	0.74	0.40	0.32
Lifetime	1364	1482	1363	1287	1280
(d, t)		(3, 10)			
App.	Ours	5-GS	6-GS	7-GS	PG
#Active	5.7	5	6	7	5.8
#RActive	1.2	2.3	1.7	0.93	1.5
#Missing	0.78	1.6	1.1	0.90	0.80
Lifetime	1481	1912	1732	1489	1474
(d, t)		(7, 20)			
App.	Ours	8-GS	9-GS	10-GS	PG
#Active	8.9	8	9	10	9.9
#RActive	0.73	0.14	0.59	0.91	0.69
#Missing	0.72	1.7	1.1	0.82	0.72
Lifetime	1163	1201	1139	1107	1125
(d, t)		(3, 20)			
App.	Ours	8-GS	9-GS	10-GS	PG
#Active	9.1	8	9	10	9.2
#RActive	0.92	2.1	1.6	0.88	1.2
#Missing	1.5	2.6	2.1	1.8	1.5
Lifetime	1249	1446	1335	1304	1243

targets that are missed, *#Missing*, at a time instance, and the amount of time that a camera network survives, *Lifetime*. To assist in understanding the three metrics, *#Active*, *#RActive*, and *#Missing*, we present two successful and a single-failed examples of $d = 7$ and 10 targets as in Figure 6. Generally, distantly located targets are well covered as in Figure 6(a); however, Cameras 3 and 5 of Figure 6(b) collectively monitor their observable targets closely located, because they could be occluded in the FoV of Camera 5. For both cases, *#Active* is 7. As in Figure 6(c), cameras may miss targets as Cameras 7 and 9, *#Missing* = 1, or be redundantly active as Camera 15, *#RActive* = 1, because of wrongly given previous, or differently estimated current, target locations.

Table 2 illustrates the average simulated values of our four primary performance metrics (in two or more significant figures) over 100 random scenario tests. The following claims can be derived from the results.

Claim 1. On average, six cameras for 10 targets and nine cameras for 20 targets are active. Accordingly, Lifetime for 10 targets is longer than that for 20 targets.

Claim 2. As d is larger, *#RActive* and *#Missing* are smaller. It would be more advantageous for more neighboring cameras to synthesize more accurate target locations by exchanging different information and the full network coverage of $d = 7$ is wider than that of $d = 3$.

Claim 3. A smaller numbers of active cameras in all cases somewhat extend the network lifetime from 1000 to 1481.

Claim 4. Because of the low *#RActives* and *#Missings* in all cases, our approach might be able to work over some localization errors, which necessarily occur in any existing localization techniques.

6.2. Complexity Analysis. Because we consider wireless cameras, we must discuss the resource overhead required by our proposed method. For each time instance in our design, energy consumption occurs according to the operations listed in Table 3.

The complexities for (1) and (5) depend on the models or algorithms camera sensors employ, and we leave them as α and β. We strongly emphasize that the two operations are conducted only by active cameras, and our proposed approach, on average, activates 0.57 (=9.1/16) times fewer cameras, including the worst case. The computational complexity for (2) is referred to as $O(t^2)$ [16]. Each target utility, each camera utility, and each global target utility respectively consume $O(t)$, $O(t^2)$, and $O(td)$ computations, which leads to $O(t^2 + td)$ computation for (3). Given such utilities, a camera determines its mode by searching in the $O(t)$ space for (4). Conversely, the communication complexity for (6) to (8) is equal to $O(td)$ because a camera maximally exchanges t pieces of information with d neighbors.

The energy consumption of camera sensors is dominated by (1) and (5) because of the significant size of image data [3, 17]. This supports our simple assumption about power consumption that only active cameras can monotonously
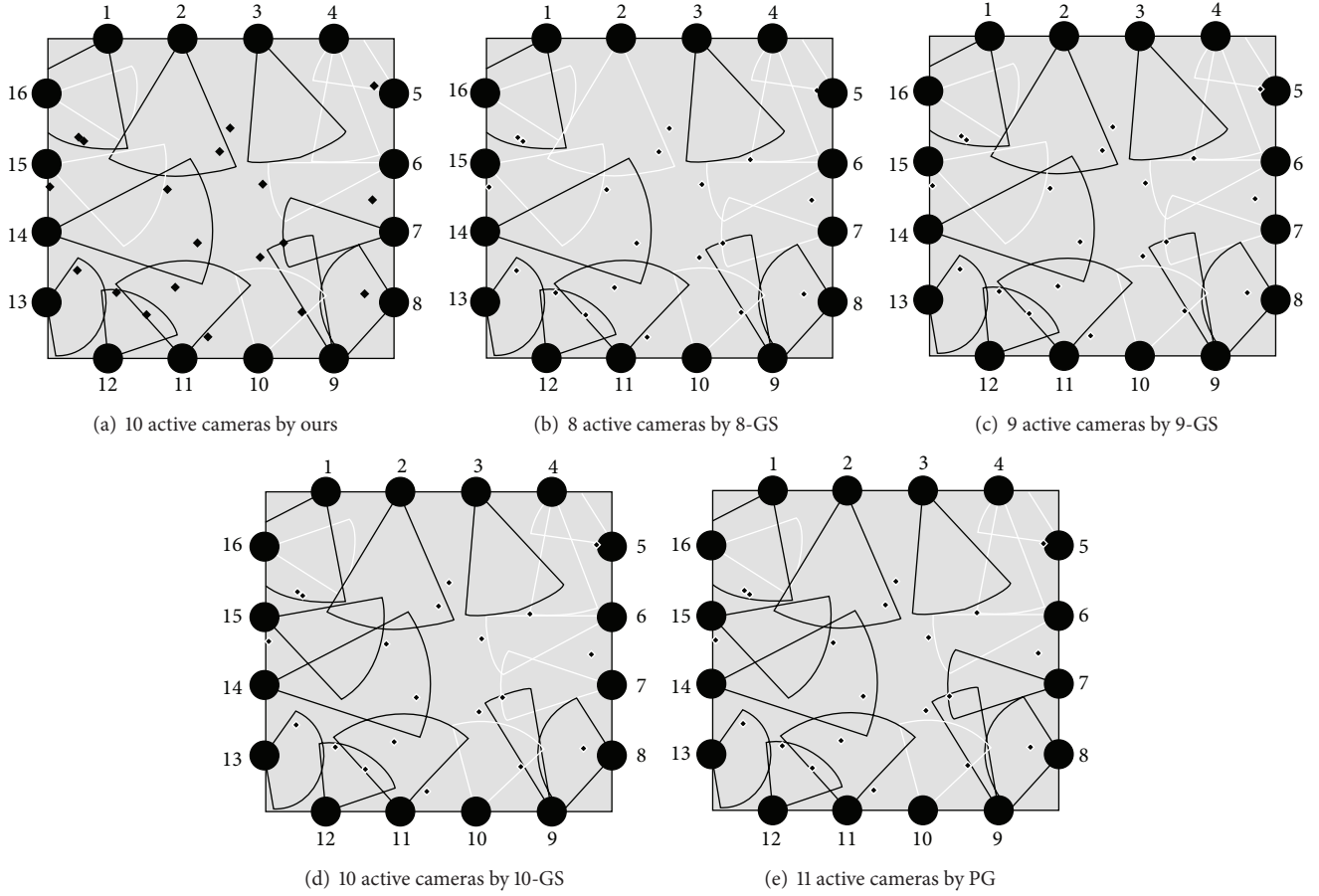
(a) 10 active cameras by ours

(b) 8 active cameras by 8-GS

(c) 9 active cameras by 9-GS

(d) 10 active cameras by 10-GS

(e) 11 active cameras by PG

FIGURE 7: Screen shots of the five approaches in the simulation of $(d, t) = (3, 20)$.

consume a single power bar per time instance and indicates that our approach can be considered as fairly competitive if targets are not highly crowded in any of the FoVs.

7. Comparison Work

Utilizing the same environments employed in our simulations, we simulated the representative alternatives, PG of [8] and GS of [5], while measuring the four network performance metrics, *#Active*, *#RActive*, *#Missing*, and *Lifetime*. For GS in particular, we have heuristically assumed that the optimal number of active cameras for 10 targets and 20 targets are, respectively, 6 and 9 by *#Actives* of Table 2. Because the numbers are not consistently optimal, we examined three GS tests for the two different target sets as {5-GS, 6-GS, 7-GS} for 10 targets and {8-GS, 9-GS, 10-GS} for 20 targets. Table 4 lists each of the network performance results for the five different approaches, on average, after 100 tests. A smaller number of active cameras typically result in longer network lifetime; correspondingly, the GSes with smaller numbers of active cameras provided the network with longer life. However, their performance observed for *#RActive* and *#Missing* was lower than analogous results from our approach and PG. Compared with our approach, PG produces similar results for

every factor. However, PG requires each camera to compute required utilities for every camera and to communicate with every other camera. This consumes greater resources compared with our approach. Therefore, we emphasize that our approach generally provides more advantageous trade-offs between {*#Active*, *Lifetime*} and *#Missing* and between *Lifetime* and resource overheads compared with the alternate approaches.

As representative instances of this comparison process, we provide five simulation screenshots for each approach in one simulation of $(d, t) = (3, 20)$ as in Figure 7. Aside from unobservable targets, our approach covers every target by 10 active cameras, whereas PG activates one redundant camera, Camera 15. As previously stated, because GS cannot adaptively select the number of active cameras, it misses targets as in Figures 7(b) and 7(c) or it additionally activates redundant cameras as in Figure 7(d).

8. Conclusion

In this paper, we addressed trade-offs between extending network lifetime and enhancing its sensing accuracy. To minimize the energy consumption necessary for data transmission while aggregating individual inferences to build

a global inference, we utilized the inference tree method to initialize an optimal data transmission path, and we demonstrated that it is very effective for reducing the number of data transmissions and energy consumption. We modeled a CS in the context of a cooperative bargaining game, where every participating camera serially optimizes the global utility, employing only local knowledge based on the serial dictatorial rule. The simulated results demonstrated that our approach extends network lifetime and performs accurately over limitedly accurate target locations. Moreover, our approach is energy-efficient for uncrowded targets, compared with the alternative representative conventional studies.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Computer Networks*, vol. 51, no. 4, pp. 921–960, 2007.

[2] Y.-I. Cho, *Collaborative data aggregation using inference tree for occupancy reasoning in visual sensor networks [Ph.D. thesis]*, KAIST, Daejeon, South Korea, 2011.

[3] S. Soro and W. Heinzelman, "A survey of visual sensor networks," *Advances in Multimedia*, vol. 2009, Article ID 640386, 21 pages, 2009.

[4] Y. Li and B. Bhanu, "Utility-based dynamic camera assignment and hand-off in a video network," in *Proceedings of the 2nd ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC '08)*, pp. 1–9, Stanford, Calif, USA, September 2008.

[5] C. Soto, B. Song, and A. K. Roy-Chowdhury, "Distributed multi-target tracking in a self-configuring camera network," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 1486–1493, Miami, Fla, USA, June 2009.

[6] D. B. Yang, J. W. Shin, A. O. Ercan, and L. J. Guibas, "Sensor tasking for occupancy reasoning in a network of cameras," in *Proceedings of the 1st International Workshop on Broadband Advanced Sensor Networks*, 2004.

[7] H. Lee, L. Tessens, M. Morbee, H. Aghajan, and W. Philips, "Sub-optimal camera selection in practical vision networks through shape approximation," in *Advanced Concepts For Intelligent Vision Systems*, vol. 5259 of *Lecture Notes in Computer Science*, pp. 266–277, 2008.

[8] R. Dai and I. F. Akyildiz, "A spatial correlation model for visual information in wireless multimedia sensor networks," *IEEE Transactions on Multimedia*, vol. 11, no. 6, pp. 1148–1159, 2009.

[9] L. M. Feeney and M. Nilsson, "Investigating the energy consumption of a wireless network interface in an ad hoc networking environment," in *Proceedings of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 3, pp. 1548–1557, Anchorage, AK, USA, April 2001.

[10] C. H. Lampert, M. B. Blaschko, and T. Hofmann, "Efficient subwindow search: a branch and bound framework for object localization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2129–2142, 2009.

[11] J. Lin, W. Xiao, F. L. Lewis, and L. Xie, "Energy-efficient distributed adaptive multisensor scheduling for target tracking in wireless sensor networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 6, pp. 1886–1896, 2009.

[12] Y. Shoham and K. Leyton-Brown, "3. Introduction to noncooperative game theory: games in normal form," in *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundation*, pp. 47–88, Cambridge University Press, New York, NY, USA, 2009.

[13] O. Kibris, "Cooperative game theory approaches to negotiations," in *Handbook of Group Decision and Negotiation*, Springer, New York, NY, USA, 2010.

[14] G. Jeong, Y.-H. Seo, S.-S. Yeo, and H. S. Yang, "Serial dictatorial rule-based games for camera selection," in *Proceedings of the 4th FTRA International Conference on Mobile, Ubiquitous, and Intelligent Computing*, 2013.

[15] J. Candamo, M. Shreve, D. B. Goldgof, D. B. Sapper, and R. Kasturi, "Understanding transit scenes: a survey on human behavior-recognition algorithms," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 1, pp. 206–224, 2010.

[16] R. Triebel, "6. Online estimation: the kalman filter," in *Information Processing in Robotics*, ETH, Zürich, Switzerland, 2009.

[17] C. B. Margi, R. Manduchi, and K. Obraczka, "Energy consumption tradeoffs in visual sensor networks," in *Proceedings of the 24th Brazilian Symposium on Computer Networks (SBRC '06)*, Curitiba, Brazil, May 2006.