

## Research Article

# Camera Sensor Activation Scheme for Target Tracking in Wireless Visual Sensor Networks

**Wei Li**

*Research Institute of Electronic Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China*

Correspondence should be addressed to Wei Li; [aegeanwei@gmail.com](mailto:aegeanwei@gmail.com)

Received 25 January 2013; Revised 21 March 2013; Accepted 25 March 2013

Academic Editor: Shuai Li

Copyright © 2013 Wei Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Target tracking has become an elementary task in wireless visual sensor networks (WVSNs). In this paper, we propose a camera sensor activation scheme for target tracking in WVSN. Our objective is to balance the tradeoff between the accuracy of target tracking and the resources of networks. By studying the sensing model and deployments of cameras, an observation correlation coefficient is derived to describe the correlation characteristics of visual information observed by cameras with overlapped field of views. According to the observation correlation coefficient, a correlation-based camera activation scheme is designed. Experimental results show that the proposed observation correlation coefficient can model the correlation characteristics of visual information in WVSN. Further simulations show that the correlation-based camera activation scheme has satisfied performance in target tracking compared with other sensor selection methods.

## 1. Introduction

Wireless visual sensor networks (WVSNs) include cameras to capture visual data from the environment and processing components to process the data locally in a desired way [1]. Compared with the wireless scalar sensor networks, WVSN can provide information-rich descriptions of captured events, which is adopted in various security and surveillance applications, such as remote and distributed video-based surveillance, environmental monitoring, and ambient-assisted living and personal care [2]. In these surveillance applications, target tracking is one of the major issues, which is to accurately determine the location of possible moving target within the least amount of time. For example, an accurate and timely determination of vehicle location is required for battlefield situational awareness [3].

Target tracking in WVSN involves a number of significant challenges. Firstly, WVSN is a resource-limited network. The capabilities of processing and transmission in WVSN are not powerful enough. There may be constraints of bandwidth for transmitting all the raw data from camera sensors to the central node and difficulties in analyzing a huge amount of visual data. The existing algorithms of tracking cannot solve

the target tracking in WVSN. Secondly, WVSN is an energy-limited network. Sensor nodes are typically battery powered, and it may not be feasible to replace or recharge the batteries of sensors in many remote sensing applications. As a result, an important characteristic of sensor networks is the limited energy. In particular, compared with the scalar sensors, image processing and transmission consume remarkable energy, in spite of the fact that the content of interest in each frame capture by the visual node is very high. Therefore, it is crucial for energy efficient use in WVSN. Thirdly, WVSN is a low-cost network. Visual nodes are usually equipped with low-resolution cameras due to the cost limitation [4]. Pixels on the target in WVSN are relatively less than that in other camera networks, which will increase the difficulties in target extraction, localization, and tracking.

To address the above challenges, the key is to balance the tradeoff between the value of information in the measurements and the resources of WVSN [5]. In this paper, we follow the information-driven sensor querying framework in which camera sensors are selectively activated based on their correlations. Because of constraints on resources, energy consumption, and low resolutions of cameras, we consider a group-based target tracking scheme, where there is a small

group of sensors active, while the rest of the network is idle. A camera which has detected the target will call up other cameras with the most correlations to compose a group of active cameras for locating and tracking the target. If any active camera loses the view of the target, it goes back to sleep, and a new camera will be activated to renew the active group based on the correlation computation. The crux of the group-based target tracking scheme is the approach of sensor activation which must be carried out with accuracy and in real time as important considerations. The scheme of activating cameras should provide the most accurate location information of the target, and it should also assure that the WWSN could execute in real time.

In this paper, we address the problem of cameras sensor activation for tracking a target in WWSN. Motivated by the fact that the image observed by a camera is directly related to its field of view, we designed an observation correlation coefficient to evaluate the correlations between two cameras based on the sensing model rather than analyzing the specific images of cameras. In cooperation with the observation correlation coefficient, the most correlated cameras are selected to be activated for locating the target in the world frame. Our main contributions include the following:

- (1) we design the observation correlation coefficient to describe the correlation from images observed by cameras with overlapped field of views. By this method, a large bulk of computation for image processing to find the correlated camera is avoided;
- (2) a small group of cameras are involved in the target tracking rather than all the cameras viewing the same target, which is beneficial for saving the cost, the computation, and communication.

The remainder of this paper is organized as follows: in Section 2, we briefly highlight the related works. Section 3 presents assumptions and preliminaries. We also present the sensing model of the cameras and the model of target localization by two cameras. Section 4 introduces the observation correlation coefficient. The proposed camera sensor activation scheme is introduced in Section 5. Section 6 conducts experiments to validate and evaluate the proposed scheme, and conclusions are given in Section 7.

## 2. Related Works

Strategies of sensor scheduling for optimizing network lifetime in wireless sensor networks have been previously considered in the literatures. Yu et al. develops a camera scheduling strategy to maximize the lifetime of the visual sensor networks [6]. Soro et al. provides a heuristic approach for camera scheduling by proposing a cost function associated with each camera depending upon the remaining energy of the camera and the coverage geometry [7]. In the above two works, they utilize the ceiling cameras, which are impractical to deploy. Cai et al. organizes the directions of sensors into a group of non-disjoint cover set, where one cover set in which the directions cover all the targets is activated at one time to extend the network lifetime [8]. Alaei et al.

provides a priority-based sensor scheduling strategy to apply coordination on cluster sensor for awakening the minimum number of sensors to monitor the interested area [9]. Given the noisy measurements and the object prior distribution, Ercan et al. uses the minimum mean square error (MSE) of the best linear estimate of the object location in 2D as a metric for sensor selection [10]. Besides, an entropy-based analytical framework is developed to measure the amount of visual information provided by multiple cameras in WWSN [11]. This method is heavily relied on a joint probability distribution of the two image sources [12], whose accurate estimation is difficult to get due to the complexity of image contents and the difficulty in image modeling [11]. The information-driven sensor scheduling technique is introduced in [13, 14], which activates the next sensor based on maximizing information utility and minimizing resource cost.

In context of sensor scheduling for target tracking, some approaches have been developed for optimizing tracking performance subject to constraints on sensor usage. Ying et al. develops a Monte Carlo solution method to address the problem of sensor scheduling for target tracking, which is formulated as a partially observable Markov decision process [15]. Kreucher et al. adopts an active sensing approach to scheduling sensors for multiple target tracking applications that combines particle filtering, predictive density estimation, and relative entropy maximization [16]. Toh et al. develops a distributed target system implementing a novel competition-based distributed sensor scheduling scheme where a candidate sensor node with highest predicted tracking accuracy will be elected as the new tasking sensor with the highest probability [17]. The previous researches are mainly in the area of wireless sensor networks. However, these existing sensor scheduling algorithms cannot solve the issue in visual sensor networks, due to the significant differences in information acquiring and processing method from conventional sensor networks.

A distributed target tracking approach using a cluster-based Kalman filter was proposed in [18], where a camera is selected as a cluster head which aggregates all the measurements in the communication range to estimate its position using a Kalman filter and sends the estimation to the central unit for tracking the target. Song et al. [19] also proposed a consensus method to track the target in a camera network. In their work, if a camera detects a target, the information from all the other cameras viewing the same target is used for fusing. From the aforementioned related works, we can observe that the accuracy of target tracking is limited by the resource of sensor networks in existing sensor scheduling methods. Therefore, in this paper, we propose the sensor activation scheme to balance the tradeoff between the tracking accuracy and resources of WWSN.

## 3. Preliminaries

*3.1. Assumptions.* In this paper, all subsequent discussions are based on the following assumptions:

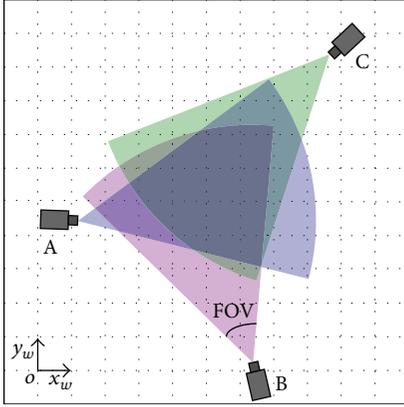


FIGURE 1: Field of view.

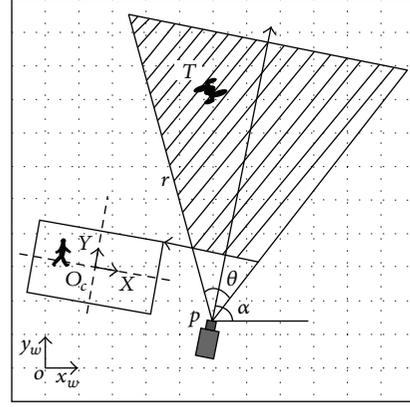


FIGURE 2: Simplified sensing model.

- (1) we consider a homogeneous WWSN, which only contains the camera sensors, in a two-dimensional field. For convenience, we assume that the camera sensors are placed in a square region. Actually, we are not concerned about the shape of the interested region, which can be circular or any other polygons;
- (2) we assume that all the cameras are placed horizontally around a space. The positions and bearings of the cameras have been already known in the world frame;
- (3) the WWSN is deployed with sufficient nodes such that the network is connected. Some of them are in the active mode during the period when no target is detected. Other cameras are in the sleep mode, which can be switched to the active mode immediately if they have received awakening signal;
- (4) the sleeping sensors can be woken up by external means. The time delay from the sleeping mode to the active mode is neglected in this paper.

**3.2. Sensing Model.** Different from scalar data sensors, cameras project a target from a 3D world to a 2D plane via a perspective point, which can only acquire the bearings of the target in the visual image. Specifically, as shown in Figure 1, a camera has a field of view (FOV) that represents the area on the  $x_w o y_w$  plane where a target captured by the camera is located. The FOV is represented as an isosceles triangle where both the equal sides join at a point representing the camera location. The angle between these equal sides is known as the FOV angle and is a factory specification defined for every camera.

It seems that cameras have unlimited sensing range; however, the target, which is described as pixels in the image, cannot be extracted from the background, if the target is located too far away from the camera, especially, when the resolution of camera is not good enough. Thus, for the reason of reducing the complexity, the sensing model of the camera in this paper is expressed by a triangle which is determined by four parameters  $(p, r, \theta, \alpha)$  [20], shown in Figure 2, where  $p$  means the position of the camera,  $r$  is the sensing radius,  $\alpha$  is the heading direction of the camera's projection (the angel

between the center line of the sight of the camera's FOV and the positive semiaxis of  $x_w$ , and rotating in counterclockwise is positive), and  $\theta$  is the FOV angle.

**3.3. The Model of Target Localization.** The purpose of target tracking is to generate the trajectory of a target over time by locating its position in every frame of the video [21]. The principle of target localization is to estimate the coordinates of intersection point when the bearings of target in different images are intersected. We set up a world coordinate system  $o x_w y_w z_w$  for the interested area. The  $x_w o y_w$  plane is the ground plane. The  $X_o_c Y$  plane is set as the image plane of cameras. For a camera sensor, its optical center can be denoted as  $o_c$ .

The coordinates of the intersection point of the target bearings in the world coordinate plane can be calculated as follows. We assume that simple background subtraction is performed locally at each camera node. The hull of the target can be extracted. A feature point is selected to represent the target. As shown in Figure 2,  $px$  and  $py$  are the pixels coordinates on the image plane. Based on the pixel information, we can compute the bearing of the target in the world coordinate frame. The bearings of the target in the image of camera can be computed according to the following formula:

$$k = \tan \left( \alpha - \arctan \left( \left( \frac{2px}{p_{\text{cons}}} \right) \cdot \tan \left( \frac{\theta}{2} \right) \right) \right), \quad (1)$$

where  $k$  is the bearing of the target,  $\alpha$  is the angle of camera rotating around  $z_w$  axis, when the direction of the camera is along with the  $x_w$ ,  $\alpha = 0$ , and rotating in counterclockwise is positive,  $\theta$  is the horizontal field of view, and  $p_{\text{cons}}$  is the number of pixels in the horizontal.  $px$  is the horizontal pixel coordinate in the image. In our localization scheme, only  $px$  is communicated to the base station.

If two cameras capture the same target at the same time, the target bearings generated from two cameras would be intersected. We can infer the coordinate of the target from the known positions of two cameras and the intersected point of

bearings of the target. The computation process is described by the following equation:

$$k_i = \frac{y - y_i^c}{x - x_i^c}, \quad (2)$$

where  $k_i$  is the bearing of the target in the  $i$ th camera's image.  $x_i^c$ ,  $y_i^c$  are the coordinates of the  $i$ th cameras in the world coordinate frame.  $x$  and  $y$  are the pending coordinates of the target in the world coordinate. If there is only one camera that could detect the target, the values  $x$  and  $y$  cannot be uniquely determined because there are two unknowns. Thus, at least two cameras that detect the camera are needed to determine the location of the target. The target position's computation matrices are shown as follows:

$$\begin{bmatrix} -k_1 & 1 \\ -k_2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y_1^c - k_1 x_1^c \\ y_2^c - k_2 x_2^c \end{bmatrix}. \quad (3)$$

Then,

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -k_1 & 1 \\ -k_2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} y_1^c - k_1 x_1^c \\ y_2^c - k_2 x_2^c \end{bmatrix}. \quad (4)$$

The position  $(x, y)$  of the target  $T$  in the world coordinate frame is obtained. Once there are more than two cameras that capture the target  $T$  simultaneously, we can utilize the statistical method to refine the estimated result for improving the accuracy [22].

**3.4. Multiple Targets Localization.** There may be multiple targets existing in the interested field. The biggest difference between the single target tracking and the multiple target tracking lies on the corresponding target matching, which can be addressed by the approaches incorporated with features of targets, such as the color, motion, contour, and boundary. From the perspective of statistics, the position of the same target estimated by a set of cameras would have more convergence in the world coordination. According to this concept, we apply a statistical method to match the corresponding target for locating targets in the world coordination.

*Step 1.* The visual hulls of targets are extracted, and the polygon centers of the target hulls are estimated. The polygon centre of the visual hull can be formulized as follows:

$$px = \sum_{i=1}^Q \frac{px_i}{m}, \quad (5)$$

$$py = \sum_{i=1}^Q \frac{py_i}{m}.$$

$px$  and  $py$  mean the center's coordinate in horizontal and vertical, respectively,  $px_i$  and  $py_i$  are the horizontal and vertical pixel positions of the visual hull of the target and  $Q$  is the amount of the hull's pixels.

*Step 2.* One target centre is selected from each camera that captures targets in order to make the correlated target pair.

Then, we use the target localization algorithm to generate possible coordinates of the target. If  $n$  cameras are deployed in the surveillance area, we could get  $C_n^2 C_a^1 C_b^1$  possible target positions.  $a$  and  $b$  are the numbers of visual hulls of targets captured by the two cameras, respectively.

*Step 3.* If we assume that one camera could capture  $m$  target at most, we could get  $m^n$  possible position pairs of targets at most. We compute the means of the target positions and the mean square errors (MSEs) between the estimates and the mean. The mean of pending coordinates of a target is formularized as follows:

$$X = \frac{x_1 + x_2 + \dots + x_n}{n}, \quad (6)$$

$$Y = \frac{y_1 + y_2 + \dots + y_n}{n}.$$

$X$  and  $Y$  are the mean of the pending position of the target;  $x_i$  and  $y_i$  are the horizontal and the vertical coordinates of the  $i$ th pending position of target, respectively. The MSE of a set of pending targets positions can be expressed as follows:

$$\text{MSE} = \frac{\sqrt{(\sum_{i=1}^n (x_i - X)^2 + \sum_{i=1}^n (y_i - Y)^2) / 2}}{n}. \quad (7)$$

*Step 4.* We make comparisons of the MSEs of the possible coordinates and find the minimum one. The mean with the little MSE is set as a possible target's position. The centers of visual hulls in the corresponding cameras' images are removed from the next iteration for finding the next correlated targets' image pair.

*Step 5.* The feasible set of centers of visual hulls of targets is updated, and Steps from 2 to 4 are repeated till all the centers are used. By this method, the corresponding target can be matched among the camera set.

## 4. Correlations between Images

In this section, we investigate how to measure the correlations of images between different cameras. Without loss of generality, we are given a set  $C = \{c_1, c_2, \dots, c_N\}$  of cameras sensors with fixed lens in a 2D interested area. For camera  $c_i$  and camera  $c_j$  in the set  $C$ , we will derive a correlation coefficient  $\lambda_{ij}$  to describe the degree of correlation between image of  $c_i$  and image of  $c_j$ .

Intuitively, the visual information provided by multiple cameras should have more correlation if the sensing regions, which are observed by cameras, are more close to each other in the world frame. If the images observed by cameras are less correlated, the sensing regions covered by cameras are definitely more distinct with each other. From this perspective, we can map the problem of estimating the correlation of images into the problem of computing the correlations of regions observed by cameras. Since the sensing area of the camera is modeled as a bounded triangular region in this paper, the correlations of observed regions can be reflected by the overlapping field of views.

We can project the overlap extent of cameras' FOVs to the overlap area of intersections of the sensing model of the cameras. This intersection is a series of triangle intersections; therefore, it will always be a convex polygon [23]. Utilizing the method of computing the union of two convex polygons, which is described in [24], we can distinguish the intersection region and order its vertices. Thus, according to the properties of the polygon, if the vertices are ordered counterclockwise, the area of the convex polygon can be computed by the following expression:

$$S = \frac{1}{2} \left( \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} + \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} + \cdots + \begin{vmatrix} x_n & y_n \\ x_1 & y_1 \end{vmatrix} \right) \quad (8)$$

or

$$S = \frac{1}{2} \sum_{i=1}^n (x_i y_{i+1} - x_{i+1} y_i), \quad (9)$$

where  $S$  is the overlap area;  $n$  is the number of vertices of the polygon.  $(x_i, y_i)$  are the coordinates of the  $i$ th vertex of the overlap region of the sensing region of cameras. Note that to close the polygon the first and the last vertex are the same; that is,  $(x_n, y_n) = (x_0, y_0)$ .

We design an observation correlation coefficient (OCC)  $\lambda$  to reveal the correlations between the images observed by cameras. Suppose that camera  $c_i$  and camera  $c_j$  are two arbitrary cameras on the ground plane that can observe the interested area. The observation correlation coefficient is defined as the ratio of the overlap area of intersection region of camera  $c_i$  and camera  $c_j$  to the entire area of sensing model of a single camera

$$\lambda_{ij} = \frac{|s_i \cap s_j|}{|s_i|} = \frac{s_{ij}}{s}, \quad (10)$$

where  $\lambda_{ij}$  is the observation correlation coefficient between camera  $c_i$  and camera  $c_j$ ,  $s_i$  and  $s_j$  mean the sensing region of the camera  $c_i$  and camera  $c_j$ , respectively,  $s_{ij}$  is the area of the intersection of  $s_i$  and  $s_j$ , and  $s$  is the sensing area, that is, the area being sensed by the camera. According to the aforementioned algorithm, we can determine the intersection of camera  $c_i$  and camera  $c_j$  and its area.  $s$  can be obtained based on the parameters of the sensing model.  $|\cdot|$  is the operator of computing the area

$$s = \frac{1}{2} r^2 \sin \theta. \quad (11)$$

According to the definition of  $\lambda$ , the maximum value is obtained when the sensing regions of camera  $c_i$  and camera  $c_j$  are identically overlapped, which equals to 1. If there is no intersection between the two cameras, the value of  $\lambda$  is zero, which is the minimum value. Thus, the value of  $\lambda$  ranges from 0 to 1.

The proposed observation correlation coefficient model is derived based on the sensing model and deployment information of cameras. To calculate the correlation between the camera  $c_i$  and the camera  $c_j$ , the camera  $c_i$  just needs to transmit its four parameters to the camera  $c_j$ :  $p, r, \theta, \alpha$ . If All

camera sensors follow the same sensing model, only  $p$  and  $\alpha$  are required to be transmitted. Once the camera  $c_j$  receives the parameters, it can calculate the observation correlation coefficient based on (10). It can be seen that the computation of observation correlation coefficient is independent of image processing.

We present a simulation to show how the observation correlation coefficient varies when the deployments of cameras change.

**4.1. OCC Influenced by Orientations of Cameras.** As shown in Figure 3(a), the camera with pink projection region is  $c_1$ , and the camera with the yellow projection region is  $c_2$ . We assume that camera  $c_1$  is fixed at  $(-L, 0)$ , and its projection bearing is  $0^\circ$ . We let the location of camera  $c_2$  change. The distance between the  $c_2$  and the original point  $(0, 0)$  maintains  $L$ . The projection bearing  $\alpha$  of camera  $c_2$  changes from  $-180^\circ$  to  $180^\circ$ . The difference of projection bearings between  $c_1$  and  $c_2$  is also  $\alpha$ . The observation correlation coefficient between  $c_1$  and  $c_2$  is illustrated in Figure 3(b).

From the results in Figure 3(b), we can see that with the increase of difference of projection bearings between the two cameras, the observation correlation coefficient decreases. When the two cameras entirely overlap with each other, the observation correlation coefficient equals 1. It means that the two cameras have the same view towards the environment. When the projection bearings of the two cameras are perpendicular, the OCC decreases to 0.32, which means that the correlations between the images of the two cameras become weak. If the projection bearings of the two cameras are opposite to each other, the OCC becomes 0, from which we can see that the images of the two cameras are totally independent of each other.

**4.2. OCC Influenced by the Positions of Cameras.** We assume that the camera  $c_1$  is fixed at the  $(5, 5)$  in the world frame. The orientation of  $c_1$  is  $60^\circ$ . Then, we fix the orientation of camera  $c_2$  and alter its position randomly in the region of  $10 \times 10$ . Based on the definition of OCC, we compute the value of OCC between  $c_1$  and  $c_2$ . Actually, it is not necessary to exhaustively search each camera in the set, since the OCC between the two cameras, whose distance from each other is greater than  $2r$ , will definitely be zero. In this case, we give the definition of neighboring sensors: the neighboring sensors are the sensors whose distances are less than  $2r$

$$c_j \text{ is } \begin{cases} \text{neighboring sensor of } c_i, & d_{ij} < 2r \\ \text{not neighboring sensor of } c_i, & \text{otherwise,} \end{cases} \quad (12)$$

where,  $d_{ij} = \|p_i - p_j\| = \sqrt{(x_i^c - x_j^c)^2 + (y_i^c - y_j^c)^2}$ ,  $p_i$  and  $p_j$  mean the positions of camera  $c_i$  and camera  $c_j$ , respectively, and  $(x_i^c, y_i^c)$ ,  $(x_j^c, y_j^c)$  are their corresponding coordinates in

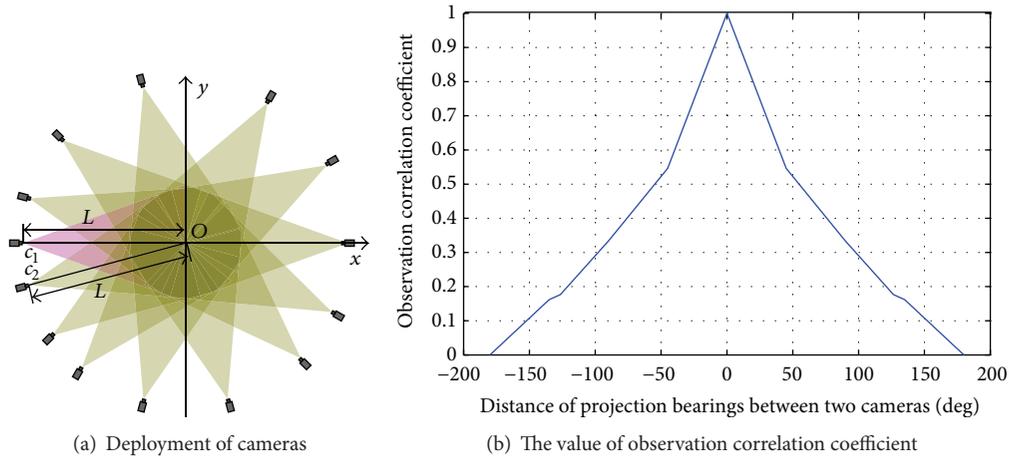
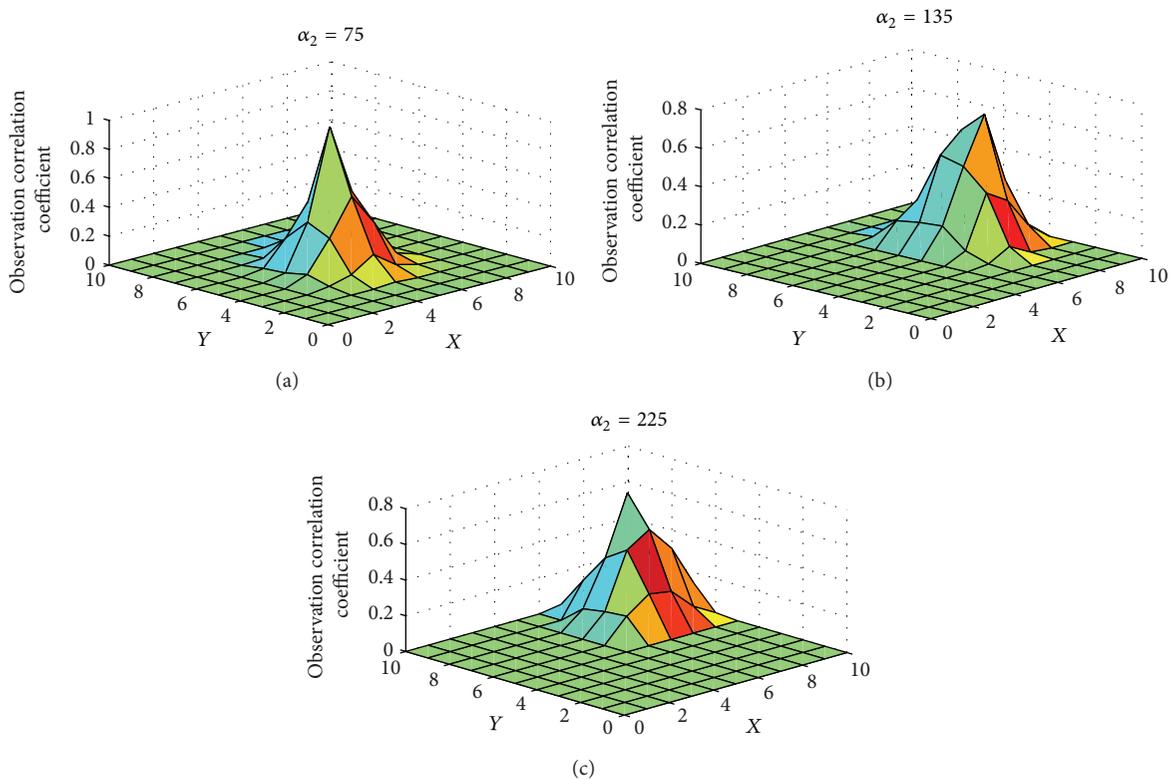


FIGURE 3: Illustration of observation correlation coefficient.

FIGURE 4: OCC value in different deployments of cameras: (a) orientation of  $c_2$  is  $75^\circ$ ; (b) orientation of  $c_2$  is  $135^\circ$ ; (c) orientation of  $c_2$  is  $225^\circ$ .

the world frame. Therefore, it is only required to compute the OCC when  $c_2$  is the neighbor sensor of  $c_1$ ; otherwise, the OCC is zero.

Figure 4 illustrates the estimated results of observation correlation coefficient between the two cameras when the orientations of  $c_2$  are  $75^\circ$ ,  $135^\circ$ , and  $225^\circ$ , respectively. We find that (1) different locations of cameras will generate different OCCs, in spite of the same orientations of cameras; (2) when the positions of cameras are identical, the OCC could be different due to the difference of orientation of cameras.

Therefore, positions and orientations of cameras are the two determining factors for the value of OCC.

In WWSN, as long as the interested region is specified, and the locations and sensing directions of cameras are estimated, the observation correlation coefficient between cameras' overlapped field of views can be obtained as introduced above. It is much easier to obtain the proposed observation correlation coefficient than to get the entropy correlation coefficient [11]. The more cameras  $c_1$  and  $c_2$  are correlated, the more observation correlation coefficient can be obtained by  $c_1$

and  $c_2$ . Therefore, the observation correlation coefficient can help find the most correlated cameras.

## 5. Camera Activation Scheme for Target Tracking

In this section, we extend our study to design a correlation-based scheme to activate cameras for tracking a target. In general, sensors are always deployed in large numbers with redundancy in sensor networks. WWSN is not exceptional. Therefore, not every camera node in the WWSN needs to be active for sensing and communication all the time. The principle of the scheme is that the camera nodes, whose observed images are more correlated with each other, should be activated for involving the target tracking. The correlation degree between the images depends on the observation correlation coefficient discussed in Section 4. As long as the OCC is given, the cameras that need to be activated can be determined. The cruxes of the camera activation scheme lie on the following: (1) if one camera detects a target, how to activate additional cameras to cooperatively locate it in the world frame? (2) If the target moves out of field of views of the activated cameras, how to renew the group of the activated cameras to keep the target in the view of wireless visual sensor networks?

*5.1. Buildup of the Group of Active Cameras.* If a camera detects a target in its image, it has to activate other cameras to locate the target. Note that only two cameras, which capture the target simultaneously, could obtain the coordinates of the target in the world frame in theory; however, the accuracy of the target localization and tracking can be gradually improved by involving more cameras. On the other hand, gaining measurements from multiple camera nodes and transmitting these measurements will consume more energy and occupy much bandwidth of wireless visual sensor networks. Therefore, in order to balance the tradeoff between the tracking quality and the amount of measurements, we select three camera sensors to be activated, because three cameras are the minimum amount of cameras for utilizing the statistical method discussed in Section 3.4. Then, we also use the statistical method, described in [22], to fuse the measurements to obtain the location of the target in the world frame.

Once the target has been detected by camera  $c_i$ , the next task is to activate more cameras to track the target. We select two additional cameras, which are most correlated with the camera that has detected the target, to be activated; that is, to activate the two cameras that have the maximum observation correlation coefficients with the camera  $c_i$

$$(c_m, c_n) = \arg \max_{c_m, c_n \in C} \{\lambda_{ni}, \lambda_{mi}\}. \quad (13)$$

We cannot assure that the camera with maximum OCC can definitely project target in its corresponding field of view. Therefore, a judgment has to be made: whether the two cameras with maximum OCC can successfully extract the target from their images. If so, it means that the target is located in

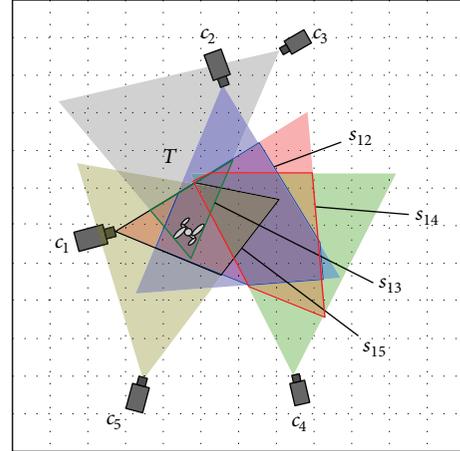


FIGURE 5: Overlaps between cameras.

the overlapping field of views of the selected cameras. By the use of epipolar geometry, discussed in Section 3, the target location in the world frame can be estimated. If not, we have to activate other cameras till all three cameras can extract the target from their images. The activation process also follows the principle of awaking the sensors with the maximum observation correlation coefficient. It will be illustrated by an example.

In Figure 5,  $c_1, c_2, c_3, c_4, c_5$  represent the five cameras, respectively. The triangle shadows with different color mean the sensing regions of the cameras. The colored polygons show the overlap area between the corresponding cameras' sensing region: the blue polygon is generated by  $c_1$  and  $c_2$ , the green one is by  $c_1$  and  $c_3$ , the red one is by  $c_1$  and  $c_4$ , and the black one is by  $c_1$  and  $c_5$ .

If camera  $c_1$  detects the target  $T$ , we will compute the OCC between  $c_1$  and its neighbor sensors  $\{c_2, c_3, c_4, c_5\}$ . According to the expression (10),  $\lambda_{12} = 0.83$ ,  $\lambda_{13} = 0.15$ ,  $\lambda_{14} = 0.45$ , and  $\lambda_{15} = 0.39$ . Thus,  $c_2$  and  $c_4$ , which have the maximum OCC with  $c_1$ , are to be selected to be activated. However, the target  $T$  is located out of the field of view of camera  $c_4$ . We have to activate another camera for involving the target localization. Among the other cameras except for  $c_2$  and  $c_4$ , the camera  $c_5$  has the maximum OCC with the camera  $c_1$  which can project the target in its image at the same time. Therefore, if  $c_1$  has detected the target, the cameras  $c_2$  and  $c_5$  will be activated to cooperate with  $c_1$  to locate the target in the world frame.

Note that when the camera nodes are sparsely deployed in the interested area, it is possible that there are less than 3 cameras that can capture the target simultaneously at some times. In the sparse deployment of cameras, the target is not always under surveillance by 3 cameras. For camera activation, therefore, we activate the available cameras that can capture the target rather than invariably activate up to the predefined number of cameras.

*5.2. Renew the Group of Active Cameras.* Once the target moves out of the field of view of the activated cameras,

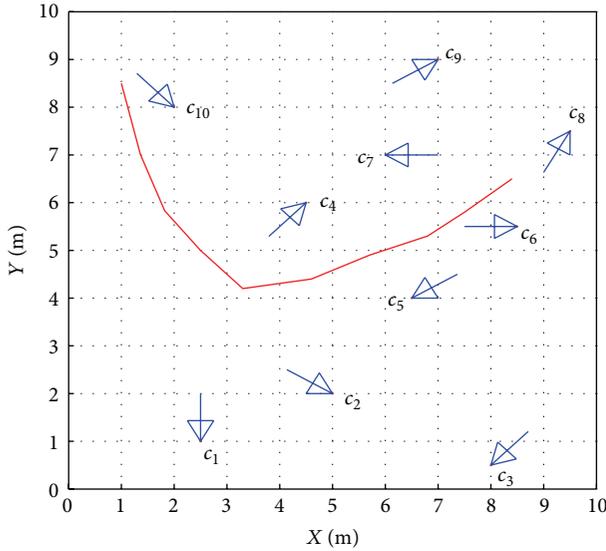


FIGURE 6: The deployment of cameras and the trajectory of the target.

we have to activate other cameras to keep the target under the incessant surveillance. In this section, we address the problem of renewing the group of active cameras for target tracking. In our proposed activation scheme mentioned above, three cameras should be activated. The camera sensor which detects the target initially is called the reference sensor in the active camera group. For example, in Figure 5, camera  $c_1$  is the reference sensor. The camera that has the biggest OCC with the reference sensor is called the primary sensor in the active group. The camera  $c_2$  is the primary sensor shown in Figure 5. The sensor which has the second biggest OCC with the reference camera is called the secondary sensor in the active group. The camera  $c_5$  is the secondary sensor in Figure 5.

If the target moves out of the field of view of the reference sensor, the sensor which has the biggest OCC with the primary sensor should be activated to join the active camera group for executing the target localization and tracking. If the primary sensor loses the view of the target, we should activate the sensor having the biggest OCC with the secondary sensor. If the secondary sensor cannot extract the target from its image, we have to activate the sensor that has the second maximum OCC with the primary sensor. When two cameras of the active camera group lose the view of the target simultaneously, the camera sensors that have the maximum OCC with the remaining active camera will be activated till there are three cameras in the active mode.

The details of the activation scheme are presented in the form of Pseudocode 1.

*5.3. Camera Activation Scheme for Multiple Target Tracking.* It is very possible that there is more than one target existing in the interested region in practical applications. The camera activation scheme has to be adjusted to satisfy requirements

of multiple targets tracking. In the case of multiple targets, we activate cameras for one target by one target.

If  $n$  targets are detected by  $n$  different cameras, which means that every reference node has only one target in its FOV. The buildup of active camera group is almost the same as that for single target tracking. The cameras that have the maximum observation correlation coefficients with the camera that captures the target are selected to be activated. Since maybe some cameras have multiple targets in its view, we utilized the statistical method, which is discussed in Section 3.4, to match the corresponding targets for locating. In the phase of renewing the active cameras, if the camera in the active group, which has only one target in its view, loses the target, the scheme of activating the additional cameras is no different from that discussed in Section 5.1. When the target loss occurs in the camera that has multiple targets in its view, it is required to judge which target moves out of the FOV. As discussed in Section 3.4, we can distinguish the corresponding active group for every target by the statistical method. Thus, if the camera that loses the target is the primary sensor, we should activate the camera having the biggest OCC with the secondary sensor for the corresponding target. If the camera that loses the target is the secondary sensor, the camera sensor which has the second maximum OCC with the primary sensor should be activated for the corresponding target.

There is also another possibility that some cameras have detected more than one target at the beginning, which means that some reference nodes have more than one target in its views. For this case, the phase of buildup of active camera group is different from the scheme for single target tracking. The selection of the primary sensor and the secondary sensor is divided into two steps: (1) find all the sensors that capture the same mount targets with the reference node from the neighboring sensors; (2) the two cameras that have the maximum OCC with the reference node are selected to be activated. The camera that has bigger OCC with the reference node in the two sensors is the primary sensor, and the other is the secondary sensor. Thus, the active camera group for the camera which has more than one target in its view is established. Furthermore, the target matching can also be obtained by using the statistical method shown in Section 3.4. In the phase of renewing the active camera group, since targets can be distinguished from each other, every active camera group for each target can be renewed by the method discussed in Section 5.2.

## 6. Performance Evaluations

In this section, we present the results of some simulations which were performed to examine the performance of the proposed camera activation scheme. For the reason of reducing the simulation complexity, we deploy ten camera sensors in a  $10\text{ m} \times 10\text{ m}$  2D field. The positions of the cameras are shown as the red points in the figure, and the blue lines show the bearing of cameras. All the cameras are with the FOV of  $57.4^\circ$  and the radius of  $10\text{ m}$ . One target is moving in the field, shown as the red line in Figure 6. All the ten cameras

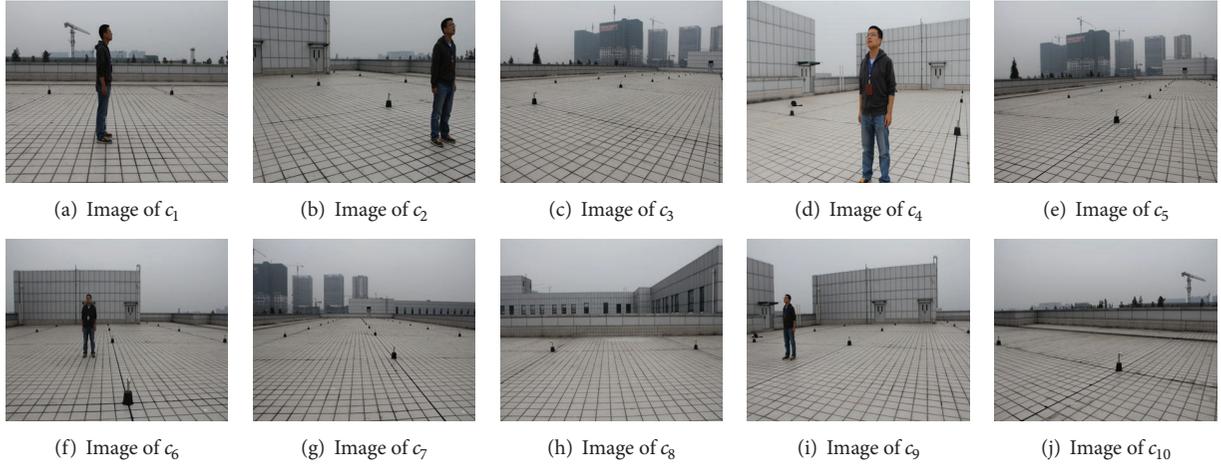


FIGURE 7: Images taken by the ten cameras at 4 s.

```

Begin  $t = 0$ , all the cameras are in the group  $C$ 
The camera sensor that has detected the target is set as the reference node  $c_r$ .
Find  $(c_m, c_n) = \arg \max_{c_m, c_n \in C} \{\lambda_{mr}, \lambda_{nr}\}$ , and activate them
If  $\lambda_{mr} \geq \lambda_{nr}$ 
     $c_m$  is the primary sensor, and  $c_n$  is the secondary sensor
Else
     $c_n$  is the primary sensor, and  $c_m$  is the secondary sensor
End if
Set updated camera group  $C' = \{C - c_r - c_m - c_n\}$ 
While ( $c_n$  or  $c_m$  cannot have the view of the target)
    {
        Find  $c' = \arg \max_{c' \in C'} \{\lambda(c_r, c')\}$ , and activate it
        Set  $C' = \{C - c_r - c_m - c_n - c'\}$ 
        Set  $c_m = c'$  or  $c_n = c'$ 
    }
Locate the target in the world frame  $(x_t, y_t)$  by the active camera group  $\{c_r, c_m, c_n\}$ 
For  $t = t + \text{time step}$  to  $t_{\max}$ 
While (any camera in the active group loses the view of the target)
    {
        Find  $c' = \arg \max_{c' \in C'} \{\lambda(c_*, c')\}$ , and activate it  $\{c_* = c_m \text{ or } c_* = c_n\}$ 
        Set  $C' = \{C - c_r - c_m - c_j - c'\}$ 
        If the camera that loses the view of the target is  $c_r$ 
             $c_r = c_m, c_m = c', c_n = c_n$ 
        Else if the camera that loses the target is  $c_m$ 
             $c_r = c_r, c_m = c_n, c_n = c'$ 
        Else the camera that loses the target is  $c_n$ 
             $c_r = c_r, c_m = c_m, c_n = c'$ 
        End if
    }
Update the position of the target  $(x_t, y_t)$  by active camera group  $\{c_r, c_m, c_n\}$ 
End for

```

## PSEUDOCODE 1

will record the images for the interested field continuously. Figure 7 shows the images of ten cameras at 4 s. We artificially fix the time step as 1 second, due to the fact that we simulate the process in the computer environment for this case. Note that the time period of processing is decided by the resources of camera node.

We take the utility-based sensor selection method [10] as a reference. In the method, the minimum MSE of the best linear estimate of the target location, which is a function of the cameras orientations, is used as a measurement for localization error. The optimal cameras are then activated by finding the camera orientations that minimize this metric.

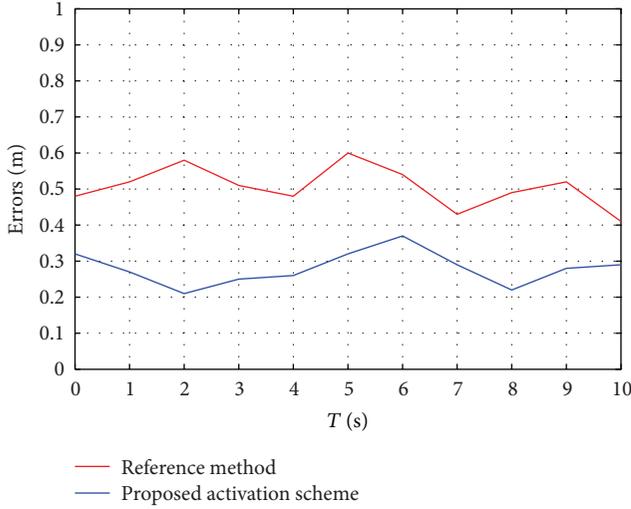


FIGURE 8: Tracking performances of the proposed activation scheme and the reference method.

Figure 8 shows the error performance of the reference method and our proposed activation scheme. The red line and the blue line show the errors of target tracking by the proposed activation scheme and the reference method, respectively. In each time step, the accuracy of target tracking in the proposed scheme is better than that in the reference method. For example, at 4 s, our proposed scheme has error of 0.26 m, by contrast, the error of target tracking in the reference method is 0.48 m. The reason behind this is that when  $c_1$  has captured the target,  $c_6$  and  $c_9$ , which have the biggest OCC with the  $c_1$ , are activated for executing the target tracking together with  $c_1$  in our proposed activation scheme. The reference method selects  $c_1$ ,  $c_2$ , and  $c_4$  to activate. Since the  $c_2$  and  $c_4$  are not very far from the object, the estimation of the target center is sensitive to the noise in the two cameras' images, and even a little change of the light condition might influence the final results. Therefore, the estimation of the target center contains more errors, and correspondingly, the accuracy of the result of target tracking decreases. From another perspective, it is not enough to only rely on the distance between the target and cameras to decide which camera should be activated for executing the target tracking.

Comparing the results of the proposed scheme and the results of the reference method, we find that in both cases, the errors of target tracking are irregular. This is in accordance with the fact that the images exhibit the state of the environment at the time of taking images, and the measurement errors cannot be predicted or predefined due to the changes of the environment.

We also make some comparisons between our proposed activation scheme and cluster-based scheme [19]. In cluster-based scheme, if a camera detects a target, all the measurements from all the other cameras that have the target in their field of view are used to estimate the position of the target. Figure 9 shows that the error performance of our proposed activation scheme is close to the cluster-based scheme.

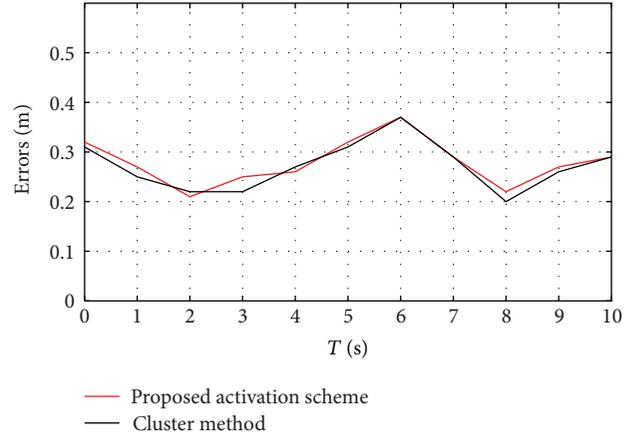


FIGURE 9: Tracking performances of the proposed activation scheme and the cluster-based method.

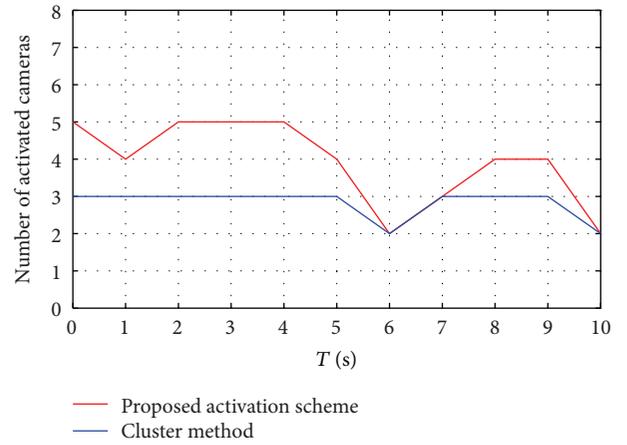


FIGURE 10: Amount of activated cameras at every time step.

Figure 10 shows the amounts of cameras that take part in target tracking in every time step by the two methods. It is apparent that cluster-based method involves more cameras to locate the position of target than our proposed activation method. In some time steps, such as at 6 s, 7 s, and 10 s, the cameras that have been activated are the same between the two methods. The reason behind this is that at such time steps, there are 3 cameras or less that can capture the target simultaneously. We activate the available cameras that view the target. From Figures 9 and 10, we can see that the proposed camera activation scheme has better performance in the tradeoff between the cameras usage and the quality of target tracking.

## 7. Conclusions

With the goal of accurately tracking a target, we have presented a camera sensor activation scheme in wireless visual sensor networks. By studying the sensing model and deployment of cameras in the network, we propose an observation correlation coefficient to describe the correlations

between images observed by cameras. The observation correlation coefficient is used for activating the most correlated cameras. In order to address the problem that more than one target exists in the interested area, we have provided a statistical method to match the targets for locating and tracking targets. Correspondingly, we also present the sensor activation scheme for multiple target tracking. With the help of simulations, we show that the proposed activation scheme could generate more accurate estimations of target tracking than the reference method. Therefore, the proposed scheme serves as a useful tool for activating cameras to track the target.

## Acknowledgments

This work was supported by “the Fundamental Research Funds for the Central Universities” and “Doctoral Fund of Ministry of Education of China.”

## References

- [1] W. Li and W. Zhang, “Sensor selection for improving accuracy of target localization in wireless visual sensor networks,” *IET Wireless Sensor Systems*, vol. 2, no. 4, pp. 293–301, 2012.
- [2] Y. Charfi, N. Wakamiya, and M. Murata, “Challenging issues in visual sensor networks,” *IEEE Wireless Communications*, vol. 16, no. 2, pp. 44–49, 2009.
- [3] J. Liu, J. Reich, and F. Zhao, “Collaborative in-network processing for target tracking,” *Eurasip Journal on Applied Signal Processing*, vol. 2003, no. 4, pp. 378–391, 2003.
- [4] I. F. Akyildiz, T. Melodia, and K. R. Chowdury, “Wireless multimedia sensor networks: a survey,” *IEEE Wireless Communications*, vol. 14, no. 6, pp. 32–39, 2007.
- [5] G. K. Atia, V. V. Veeravalli, and J. A. Fuemmeler, “Sensor Scheduling for energy-efficient Target Tracking in Sensor Networks,” *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4923–4937, 2011.
- [6] C. Yu and G. Sharma, “Camera scheduling and energy allocation for lifetime maximization in user-centric visual sensor networks,” *IEEE Transactions on Image Processing*, vol. 19, no. 8, pp. 2042–2055, 2010.
- [7] S. Soro and W. B. Heinzelman, “On the coverage problem in video-based wireless sensor networks,” in *Proceedings of the 2nd International Conference on Broadband Networks (BROAD-NETS '05)*, pp. 932–939, October 2005.
- [8] Y. Cai, W. Lou, M. Li, and X. Y. Li, “Energy efficient target-oriented scheduling in directional sensor networks,” *IEEE Transactions on Computers*, vol. 58, no. 9, pp. 1259–1274, 2009.
- [9] M. Alaei and J. M. Barcelo-Ordinas, “Priority-based node selection and scheduling for wireless multimedia sensor networks,” in *Proceedings of the 6th Annual IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob '2010)*, pp. 151–158, October 2010.
- [10] A. Ercan, D. Yang, A. E. Gamal, and L. Guibas, “Optimal placement and selection of camera network nodes for target localization,” in *Proceedings of the Conference on Distributed Computing in Sensor Systems*, pp. 389–404, 2006.
- [11] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, “Mutual-information-based registration of medical images: a survey,” *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, 2003.
- [12] M. C. Vuran, Ö. B. Akan, and I. F. Akyildiz, “Spatio-temporal correlation: theory and applications for wireless sensor networks,” *Computer Networks*, vol. 45, no. 3, pp. 245–259, 2004.
- [13] F. Zhao, J. Liu, J. Liu, L. Guibas, and J. Reich, “Collaborative signal and information processing: an information-directed approach,” *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1199–1209, 2003.
- [14] M. Chu, H. Haussecker, and F. Zhao, “Scalable information-driven sensor querying and routing for ad hoc heterogeneous sensor networks,” *International Journal of High Performance Computing Applications*, vol. 16, no. 3, pp. 293–313, 2002.
- [15] Y. He and E. K. P. Chong, “Sensor scheduling for target tracking: a Monte Carlo sampling approach,” *Digital Signal Processing*, vol. 16, no. 5, pp. 533–545, 2006.
- [16] C. Kreucher, K. Kastella, and A. Hero, “Sensor management using an active sensing approach,” *IEEE Transactions on Signal Processing*, vol. 85, no. 3, pp. 607–624, 2005.
- [17] Y. K. Toh, W. Xiao, and L. Xie, “A wireless sensor network target tracking system with distributed competition based sensor scheduling,” in *Proceedings of the International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP '07)*, pp. 257–262, December 2007.
- [18] H. Medeiros, J. Park, and A. C. Kak, “Distributed object tracking using a cluster-based Kalman filter in wireless camera networks,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 2, no. 4, pp. 448–463, 2008.
- [19] B. Song, A. T. Kamal, C. Soto et al., “Tracking and activity recognition through consensus in distributed camera networks,” *IEEE Transactions on Image Processing*, vol. 19, no. 10, pp. 2564–2578, 2010.
- [20] 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.
- [21] A. Yilmaz, O. Javed, and M. Shah, “Object tracking: a survey,” *ACM Computing Surveys*, vol. 38, no. 4, article 13, 2006.
- [22] W. Li, J. Portilla, F. Moreno, G. Liang, and T. Riesgo, “Improving target localization accuracy of wireless visual sensor networks,” in *Proceedings of the 37th IEEE Industrial Electronics Conference*, Melbourne, Australia, November 2011.
- [23] T. Massey, R. Kapur, F. Dabiri, L. N. Vu, and M. Sarrafzadeh, “Localization using low-resolution optical sensors,” in *Proceedings of the IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS '07)*, Pisa, Italy, October 2007.
- [24] P. Zhou, *Computational Geometry Algorithm Design and Analysis*, Tsinghua University Press, 2008.

