

# A Statistical Approach for Target Counting in Sensor-Based Surveillance Systems

Dengyuan Wu\*, Dechang Chen<sup>†</sup>, Kai Xing<sup>‡</sup>, Xiuzhen Cheng\*

\*Department of Computer Science, The George Washington University, Washington, DC 20052, USA

<sup>†</sup>Division of Epidemiology and Biostatistics, Uniformed Services University of the Health Sciences, MD 20814, USA

<sup>‡</sup>Department of Computer Science, University of Science and Technology of China, P. R. China

Email: {andrewwu,cheng}@gwu.edu, dchen@usuhs.mil, kxing@ustc.edu.cn

**Abstract**—Target counting in sensor-based surveillance systems is an interesting task that potentially could have many important applications in practice. In such a system, each sensor outputs the number of targets in its sensing region, and the problem is how one can combine all the reported numbers from sensors to provide an estimate of the total number of targets present in the entire monitored area. The main challenge of the problem is how to handle different sensors' outputs that contain some counts of the same targets falling into the overlapped area from these sensors' sensing regions. This paper introduces a statistical approach to estimate the target count in such a surveillance system. Our approach avoids direct handling of the overlapping issue by adopting statistical methods. First, depending on whether or not certain prior knowledge is available regarding the target distribution, the procedure in minimizing the residual sum of squares or kernel regression is used to estimate the distribution of targets. Then the estimated count of the total targets is obtained by the method of likelihood estimation based on a sequence of binomial distributions that are derived from a sampling procedure. Comparisons based on simulations show that our proposed counting approach outperform the state of art counting algorithms. Extensive simulations also show that our proposed approach is very fast and very promising in estimating the target count in sensor-based surveillance systems.

## I. INTRODUCTION

Wireless sensor networks have been widely used to monitor various activities in many different types of environments [1]–[6]. One of important monitoring tasks is the estimation of the total number of targets within a region at a specific time. For example, one may want to estimate the size of a group of sea birds in a certain seashore area at specified times for a given day. A sensor network can be deployed for this purpose to gather relevant data and make inferences periodically. Although it is often difficult to obtain the exact count, target counting algorithms with a high counting accuracy are always sought.

In general, there are two types of errors that lead to inaccurate counting: miss-detection and double-counting [7]. A target may fail to be detected by a sensor even if the target is in the sensor's sensing region. This is often due to imprecision of the hardware. A target in the overlapped area of different sensors' sensing regions may be detected by more than one sensor. This can result in double counting, where a target may be counted for at least two times in the final estimate. Double counting tends to occur as sensors are more densely deployed and more targets are present. Usually, the

issue of miss-detection can be handled by using hardware-based detection algorithms or building more sensitive sensors. However, double-counting is normally resolved or mitigated by analyzing information obtained from sensors.

In this paper, we propose a simple but effective statistical approach to estimate the total number of targets in a sensor-based surveillance system, given the count of targets reported by each sensor within its sensing region. The issue of miss-detection is not considered here. However, the issue of overlapping or double-counting has been avoided with our method. In our approach, we first rebuild the target distribution in the monitored area using regression techniques. Parametric or non-parametric methods are used to estimate the target distribution, depending on whether or not certain prior knowledge is available regarding the distribution. Then a sampling procedure is used to randomly select subsets of sensors, where sensors from the same subset do not have overlapped sensing regions. Finally, the total count of targets is estimated through binomial distributions and the method of maximum likelihood estimation. Based on our knowledge, the work presented in this paper is the first that employs statistical methods to effectively deal with the double counting problem in wireless counting sensor network.

The first and the most relevant research work that deals with double counting in the context of sensing is [7]. The counting method in [7], largely based on the probability theory, mainly works for the scenario where (1) targets are uniformly distributed, (2) the number of targets present in the monitored area is not large, and (3) two sensing regions from two sensors do not overlap heavily.

Compared to the state-of-the-art, our paper has the following novel contributions.

- 1) Regression techniques (minimizing the residual sum of squares and kernel regression), sampling, and the method of likelihood estimation have been integrated into an approach for target counting. It is the first time that a completely statistical approach is introduced to estimate target counts in the field of wireless sensor networks.
- 2) Our approach avoids direct handling of the issue of double counting or overlapping of sensing regions.
- 3) Our approach works for any number of targets present in the monitored area.

- 4) The proposed counting approach works for any distribution of targets. Targets can come from a specific distribution, or they can form clusters.
- 5) The proposed counting approach can provide real time estimates of target counts.
- 6) The proposed approach outperforms the state-of-the-art counting algorithms in terms of counting accuracies.

The rest of the paper is organized as follows. In Section II, we review the most related research in target counting. Our network model and other background information are introduced in Section III. In section IV we discuss how to estimate the distribution of targets by using parametric and non-parametric statistical methods. The sampling procedure used to select subsets of sensors and the method of likelihood estimation used to estimate the target count are provided in Section V. Simulation results are reported in Section VI. And this paper is concluded with a discussion on future research in Section VII.

## II. RELATED WORKS

Research works on target counting are roughly categorized into four types corresponding to minimalistic, complex, binary, and energy-based modeling of wireless sensor network.

The works in [7]–[9] are representatives of the research based on the minimalistic model, where each sensor outputs the number of distinct targets in its sensing region. The work in [7] estimates the target count by probability theory. The authors claim that under the assumption of uniform targets, the number of targets falling into a specific region follows a Poisson distribution. So they first partition the monitored area into non-overlapped subareas. Then the expected value based on these subareas and the Poisson modeling is used to estimate the number of targets in the monitored area. This approach that aims at resolving the double counting issue has several disadvantages. It is impossible to provide an estimate for a big target count within a reasonable period of time. When sensing regions overlap heavily, not only the counting accuracy degrades but also the amount of time in computation increases dramatically. Paper [8] proposes a target counting algorithm to estimate the range of the number of targets in the monitored area. The range is based on the worst case and thus might be too wide to be applicable in real applications. Paper [9] starts from topology integration and uses the expected value as an estimate of the number of targets. Since there are no experimental or simulation results provided, it is difficult to evaluate the proposed method.

In the setting of complex sensor models, [10] also addresses the double counting problems. However, in addition to the number of targets within the sensing region of a sensor, more information such as distances or angles is needed in their algorithms.

Binary-sensing based approaches [11]–[13] estimate the number of targets by assuming that a sensor reports a value ‘1’ if one or more targets are detected in its sensing region and ‘0’ otherwise. The success of these algorithms relies on the assumption of sparse targets.

Energy-based approaches [14]–[16] estimate the count of targets by using the unit target energy volume and the estimated energy volume of the energy landscape. These algorithms assume the same energy level and energy decay model for all individual targets.

Other algorithms in counting targets exist in the literature. For example, [17] relies on probability theory to estimate the number of targets, but the algorithm works only for sparse targets. And in [18], a system is developed to measure the vehicle count in a road network.

Regression is the process of constructing a curve, or mathematical function, which is closest to a series of data points under a predefined metric. In statistics, regression analysis involves the analysis between a single dependent variable and a few independent variables. In wireless sensor networks, polynomial regression has been used in [14] to recover an approximate energy landscape. In [19], polynomial regression is used to compress the data transported in wireless sensor network communications. In our work, we use the regression technique to estimate the target distribution over the two-dimensional monitored area.

In this paper we propose a statistical approach for target counting under a minimalistic model. While most existing counting algorithms focus on estimating the number of targets in each overlapped area of sensing regions of sensors, our approach treats the total number of targets in the system as a parameter to be estimated by statistical methods. With no restriction on the distribution of targets and the number of monitored targets, our approach provides a fast and effective tool in target count estimation.

## III. BACKGROUND INTRODUCTION

We consider the network model where sensors are deployed in a grid or random uniform pattern in a monitored area that is contained in the two-dimensional Euclidean space. These two deployment patterns represent two widely used scenarios in the literature. Suppose there are  $N_s$  sensors deployed in the field:  $s_1, s_2, \dots, s_{N_s}$ . Let  $x_i = (x_1^{(i)}, x_2^{(i)})$  denote the location of sensor  $s_i$  ( $i = 1, 2, \dots, N_s$ ). The set of locations of all sensors is represented by  $L_S = \{x_1, x_2, \dots, x_{N_s}\}$ . Note that sensor locations can be derived based on existing approaches such as those proposed in [20], [21]. The sensing region of a sensor  $s_i$ , denoted by  $A(s_i)$ , is modeled as a circular region with center  $x_i$  and radius  $h$ . Then all  $A(s_i)$  have the same size  $a = \pi h^2$ . Suppose at the time of investigation, sensor  $s_i$  outputs a reading  $r_i$  that equals the number of targets residing in the sensor’s sensing region  $A(s_i)$ . Let  $R_S = \{r_1, r_2, \dots, r_{N_s}\}$  denote the set of all such readings from all sensors. The entire monitored area  $\mathcal{A}$  of the surveillance system can be defined as the union of the sensing regions of all sensors.

Suppose  $N_t$  targets are distributed in the monitored area  $\mathcal{A}$  according to a specific distribution with the probability density function (pdf)  $f(x|\theta)$ . Here,  $x$  represents a two-dimensional vector  $(x_1, x_2)'$  and  $\theta$  is the parameter or set of parameters associated with the distribution. Examples of the target distribution include a uniform or bivariate normal

distribution. In this paper, we assume a general form of the distribution of targets. Due to the use of different statistical approaches, we consider two separate cases: (a) we have some prior knowledge about  $f(x|\theta)$ ; and (b) we do not have any prior knowledge about  $f(x|\theta)$ . Case (a) does not necessarily mean that we know completely about the distribution. For instance, we may know the distribution is bivariate normal but we do not know the exact parameter(s). Case (b) is equivalent to the fact that we know nothing about the target distribution.

This paper focuses on estimating the total number of targets ( $N_t$ ) by using the sensors' locations  $L_S$  and sensors' readings  $R_S$ . This estimation task is challenging since a target could be caught and reported by several sensors simultaneously. An intuitive method to estimate  $N_t$  might be that one simply adds the readings from all sensors and then multiply the total by a factor  $\gamma$  small than 1 to eliminate the effect from repeatedly counting. This method would fail frequently since it is usually difficult to find such a factor  $\gamma$  that could be influenced by a lot of other factors such as the size of the network, the target distribution, and the sensor deployment pattern, etc. Therefore, novel procedures are needed in order to obtain a reliable estimate of  $N_t$ . In the next two sections, we will discuss how to estimate the pdf  $f(x|\theta)$  of the target's distribution for both cases (a) and (b) by regression technique and then discuss how to estimate  $N_t$  by the method of maximum likelihood estimation.

#### IV. TARGET POSITION PROBABILITY DENSITY DISTRIBUTION ESTIMATION

In this section, we will discuss in details how to estimate the pdf  $f(x|\theta)$  of targets' distribution by using sensors' locations  $L_S$  and sensors' readings  $R_S$ . Our approach is based on the observation that the expected count of targets in an unit area centered at  $x$  is approximately proportional to the value of  $f(x|\theta)$  at  $x$ . Below is a derivation leading to this observation.

Suppose targets are independently distributed according to a distribution with the pdf  $f(x|\theta)$ . Consider the sensing region  $A(s_i)$  from sensor  $s_i$ . It is seen that a target falls into  $A(s_i)$  with the probability  $p_i = \int \int_{A(s_i)} f(x|\theta) dx_1 dx_2$ . In general, for  $N_t$  targets, the total number of targets, denoted by  $R_i$ , that reside  $A(s_i)$  follows the binomial distribution  $B(N_t, p_i)$ . Therefore, the expected count of targets in this sensing region is the following conditional expectation

$$E(R_i | x_i) = N_t * \int \int_{A(s_i)} f(x|\theta) dx_1 dx_2 \quad (1)$$

According to Calculus, when  $f(x|\theta)$  is a continuous function, we have the following:

$$\begin{aligned} \int \int_{A(s_i)} f(x|\theta) dx_1 dx_2 &\approx f(x_i|\theta) * \int \int_{A(s_i)} dx_1 dx_2 \\ &= f(x_i|\theta) * a \end{aligned} \quad (2)$$

Set

$$Y_i = \frac{R_i}{a}$$

We see that  $Y_i$  represents the count of targets per unit area around the sensor  $s_i$ . From (1) and (2), we obtain the following

$$E(Y_i | x_i) \approx N_t * f(x_i|\theta) \quad (3)$$

The approximation (3) is derived at sensors' locations. However, it is easy to see that (3) actually holds for any location  $x$  in the monitored area  $\mathcal{A}$ , that is,

$$E(Y | x) \approx N_t * f(x|\theta) \quad (4)$$

where  $Y = \frac{R}{a}$  for  $R$  denoting the count of targets in the circular region with center  $x$  and radius  $h$ . Therefore, the expected count of targets per unit area at  $x$  is approximately proportional to the value of  $f(x|\theta)$  at  $x$ . Though simple, this relationship between  $Y$  and  $f(x|\theta)$  depicted in (4) is the basis for the following two subsections where we estimate  $f(x|\theta)$  by using sensors' location and readings. For simplicity, we treat  $Y$  as a continuous variable.

##### A. Parametric Approach

In this subsection, we discuss how to estimate the pdf  $f(x|\theta)$  of the distribution of targets when the form of  $f(x|\theta)$  is known but  $\theta$  is unknown. This corresponds to estimating  $f(x|\theta)$  for the case (a) stated in Section III.

We will utilize (4) to obtain the estimate of  $\theta$ . From the discussion proceeding (4), we know that the following sequence is available as a given dataset for this estimation problem:

$$(x_1, y_1), (x_2, y_2), \dots, (x_{N_s}, y_{N_s}), \quad (5)$$

where  $x_i = (x_1^{(i)}, x_2^{(i)})$  is the location of sensor  $s_i$  and  $y_i = \frac{r_i}{a}$  is the observed count of targets per unit area around sensor  $s_i$ . Using the idea in regression, we obtain an estimate  $\hat{\theta}$  of  $\theta$  by minimizing the residue sum of squares, i.e., by solving

$$\min_{\theta} \sum_{i=1}^{N_s} (y_i - c * f(x_i|\theta))^2, \quad (6)$$

where  $c$  is an unknown parameter. Thus,  $f(x|\hat{\theta})$ , denoted by  $\hat{f}(x)$ , is an estimate of  $f(x|\theta)$ . In this paper, we used the Levenberg-Marquardt algorithm (LMA) [22] in Matlab 2009a to find a solution  $\hat{\theta}$  to this minimization problem. LMA is a well known iterative procedure used to solve a wide variety of optimization problems.

##### B. Non-parametric Approach

Often in practice, the form of  $f(x|\theta)$  is unknown. This is the case (b) stated in Section III. We know that (4) gives a simple relationship between the conditional expectation  $E(Y | x)$  and  $f(x|\theta)$ . Therefore, to estimate  $f(x|\theta)$ , we only need to find and rescale an estimator of  $E(Y | x)$ . In this subsection, we will discuss how to estimate  $E(Y | x)$  by kernel regression, a non-parametric technique in estimating conditional expectation of a random variable [23].

Let  $g(y, x)$  be the joint pdf of  $Y$  and  $X$  with  $X$  denoting a random location in the monitored area  $\mathcal{A}$ ,  $g_x(x)$  the marginal

pdf of  $X$ , and  $g(y|x)$  the conditional pdf of  $Y$  given  $X = x = (x_1, x_2)'$ . We have

$$\begin{aligned} E(Y | x) &= E(Y | X = x) \\ &= \int yg(y | x)dy \\ &= \frac{\int yg(y, x)dy}{g_x(x)} \end{aligned} \quad (7)$$

If we use (5) to estimate densities  $g_x(x)$  and  $g(y, x)$  by the technique of kernel density estimation, then  $g_x(x)$  and  $\int yg(y, x)dy$  can be estimated by

$$\frac{1}{N_s} \sum_1^{N_s} K_{h_1, h_2}(x - x_i)$$

and

$$\frac{1}{N_s} \sum_1^{N_s} K_{h_1, h_2}(x - x_i)y_i$$

respectively, where by using a multiplicative kernel with the same bandwidth  $h (= h_1 = h_2)$ ,

$$K_{h_1, h_2}(x - x_i) = \frac{1}{h^2} K\left(\frac{x_1 - x_1^{(i)}}{h}\right) K\left(\frac{x_2 - x_2^{(i)}}{h}\right)$$

for a univariate kernel  $K(\cdot)$ . Therefore, from (7) we obtain the following Nadaraya-Watson estimator of  $E(Y | x)$ :

$$\frac{\sum_{i=1}^{N_s} K\left(\frac{x_1 - x_1^{(i)}}{h}\right) K\left(\frac{x_2 - x_2^{(i)}}{h}\right) y_i}{\sum_{i=1}^{N_s} K\left(\frac{x_1 - x_1^{(i)}}{h}\right) K\left(\frac{x_2 - x_2^{(i)}}{h}\right)}$$

Then the above function divided by its integral over the entire monitored area of the sensor filed will give one estimate  $\hat{f}(x)$  of  $f(x|\theta)$ . Thus

$$\hat{f}(x) = C \frac{\sum_{i=1}^{N_s} K\left(\frac{x_1 - x_1^{(i)}}{h}\right) K\left(\frac{x_2 - x_2^{(i)}}{h}\right) y_i}{\sum_{i=1}^{N_s} K\left(\frac{x_1 - x_1^{(i)}}{h}\right) K\left(\frac{x_2 - x_2^{(i)}}{h}\right)} \quad (8)$$

where  $C$  is a constant.

Note that the distribution of sensors in our setting is known, i.e., sensors are uniformly distributed. In this case,  $g_x(x)$  is a constant. Then by (7), the following function divided by its integral over the entire monitored area will also give one estimator of  $f(x | \theta)$ :

$$\sum_{i=1}^{N_s} K\left(\frac{x_1 - x_1^{(i)}}{h}\right) K\left(\frac{x_2 - x_2^{(i)}}{h}\right) y_i \quad (9)$$

That is,

$$\hat{f}(x) = C \sum_{i=1}^{N_s} K\left(\frac{x_1 - x_1^{(i)}}{h}\right) K\left(\frac{x_2 - x_2^{(i)}}{h}\right) y_i \quad (10)$$

for a constant  $C$ .

In this paper, we use the Gaussian kernel  $K(u) = \frac{1}{2\pi} e^{-\frac{u^2}{2}}$ , as it is one of the widely used kernels. And we choose the bandwidth  $h$  to be the radius of the sensing region of a sensor for simplicity. Note that only  $L_S$  and  $R_S$  are needed in computing  $\hat{f}(x)$ .

## V. TARGET COUNT ESTIMATION

For sensors whose sensing regions do not overlap, the sum of their readings is the exact count of targets falling into their sensing regions. This section discusses how to use this observation and the estimated distribution of targets to estimate the total number of targets in the monitored area  $\mathcal{A}$  by the method of maximum likelihood estimation.

### A. Selecting Sensors with Non-Overlapped Sensing Regions

We begin with a brief discussion on selecting sensors with non-overlapped sensing regions. Two sensing regions  $A(s_i)$  and  $A(s_j)$  from two sensors  $s_i$  and  $s_j$  are overlapping if the intersection of  $A(s_i)$  and  $A(s_j)$  is not empty, i.e.,  $A(s_i) \cap A(s_j) \neq \emptyset$ . Clearly,  $A(s_i)$  and  $A(s_j)$  are overlapping if and only if the distance between  $s_i$  and  $s_j$  is less than or equal to  $2h$ , where  $h$  is the radius of a sensing region of a sensor. This simple rule can be used to randomly select a set of sensors that do not contain two sensors whose sensing regions are overlapping. Suppose we repeat this selection process for  $m$  times to obtain  $m$  such sets  $S_1, S_2, \dots, S_m$ , where  $A(s_i) \cap A(s_j) = \emptyset$  for any two sensors from the same set  $S_k$  ( $1 \leq k \leq m$ ).

We can also add one more confinement in constructing these sets of sensors by requiring that each  $S_k$  is the largest in the sense that the sensing region of any sensor outside  $S_k$  overlap with at least one sensing region of a sensor inside  $S_k$ . In other word,  $S_k$  is the largest if it is impossible to add another sensor into  $S_k$  while keeping the non-overlapping status of the subset.

There are several advantages when using  $S_k$ . For example, the integral of  $f(x | \theta)$  over the union of sensing regions of all sensors in  $S_k$  is the sum of the integrals of  $f(x | \theta)$  over all individual sensing regions. As another example, we see easily that the sum of the readings of all sensors in  $S_k$  represents the total number of targets falling into the monitored areas of these sensors.

### B. Target Count Estimation By the Method of Maximum Likelihood

Now we discuss how to estimate the target count in the monitored area  $\mathcal{A}$ . We will use the estimated pdf  $\hat{f}$  from Section IV, the sets  $S_k$  constructed in Section V-A, and the method of maximum likelihood. The method of maximum likelihood [24] is one of the most widely techniques for obtaining estimates of parameters. With this method, the maximum likelihood estimator (MLE) of the parameter(s)  $\theta$  is  $\hat{\theta}$  at which the likelihood function obtains its maximum as a function of  $\theta$ .

Suppose  $N_i$  targets are independent and identically distributed in the monitored area  $\mathcal{A}$  according to a distribution with the pdf  $f(x|\theta)$ . Let  $P_i$  denote the probability that a target falls into the area monitored by the sensors in the set  $S_i$  ( $1 \leq i \leq m$ ). Then using an estimate  $\hat{f}(x)$ , we obtain an estimate  $\hat{P}_i$  of  $P_i$  as follows:

$$\begin{aligned} \hat{P}_i &= \int \int_{\cup A(s_j), s_j \in S_i} \hat{f}(x) dx_1 dx_2 \\ &= \sum_{s_j \in S_i} \int \int_{A(s_j)} \hat{f}(x) dx_1 dx_2 \end{aligned} \quad (11)$$

Let  $U_i$  denote the number of targets that fall into the area monitored by the sensors in the set  $S_i$ . We see that  $U_i$  is a discrete random variable that approximately follows the binomial distribution  $B(N_t, \hat{P}_i)$ , where  $\hat{P}_i$  is known and computed in (11) but  $N_t$  is unknown. Below we proceed to find an MLE for  $N_t$  by using these binomial variables  $U_1, U_2, \dots, U_m$

From the selection process of  $S_1, S_2, \dots, S_m$ , it follows that the random variables  $U_1, U_2, \dots, U_m$  can be treated as independent. By the readings of sensors in  $S_i$ , we obtain an observed value  $u_i$  for  $U_i$ , where

$$u_i = \sum_{s_j \in S_i} r_j \quad (12)$$

Therefore we have the likelihood function

$$l(N_t) = \prod_{i=1}^m \binom{N_t}{u_i} \hat{P}_i^{u_i} (1 - \hat{P}_i)^{N_t - u_i} \quad (13)$$

The MLE of  $N_t$  is obtained by maximizing the likelihood function. However, the traditional method involving differentiation is not easy to be carried out due to the factorials. Therefore a different approach is needed to solve this optimization problem. Since  $l(N_t) = 0$  for  $N_t < \max_i \{u_i\}$ , it is seen that the MLE is an integer  $N_t$  such that

$$N_t \geq \max_i \{u_i\} \quad (14)$$

and

$$\frac{l(N_t)}{l(N_t - 1)} \geq 1, \quad \frac{l(N_t + 1)}{l(N_t)} < 1 \quad (15)$$

Below we show that  $N_t$  satisfying (14) and (15) is unique.

In fact, it follows from (13) that

$$\begin{aligned} \frac{l(N_t)}{l(N_t - 1)} &= \prod_{i=1}^m \frac{\binom{N_t}{u_i} \hat{P}_i^{u_i} (1 - \hat{P}_i)^{N_t - u_i}}{\binom{N_t - 1}{u_i} \hat{P}_i^{u_i} (1 - \hat{P}_i)^{N_t - 1 - u_i}} \\ &= \prod_{i=1}^m \frac{\binom{N_t}{u_i}}{\binom{N_t - 1}{u_i}} (1 - \hat{P}_i) \\ &= \prod_{i=1}^m \frac{N_t}{N_t - u_i} (1 - \hat{P}_i) \\ &= \prod_{i=1}^m (1 - \hat{P}_i) \prod_{i=1}^m \left(1 + \frac{u_i}{N_t - u_i}\right) \end{aligned} \quad (16)$$

Consider the function

$$g(w) = \prod_{i=1}^m (1 - \hat{P}_i) \prod_{i=1}^m \left(1 + \frac{u_i}{w - u_i}\right)$$

defined for  $w \in (\max_i \{u_i\}, +\infty)$ . Clearly,  $g(w)$  is continuous and strictly decreases in its domain. In addition,

$$\lim_{w \rightarrow +\infty} g(w) = \prod_{i=1}^m (1 - \hat{P}_i) < 1,$$

and as  $w$  approaches  $\max_i \{u_i\}$  from the right  $g(w) \rightarrow +\infty$ . Therefore, it follows from (16) that there exists only one integer, denoted  $\hat{N}_t$ , such that (14) and (15) hold. This  $\hat{N}_t$  is the MLE of  $N_t$ .

In practice, it is easy to find  $\hat{N}_t$  by (14) and (15). Starting from  $\max_i \{u_i\}$ , we compute the ratio

$$\frac{l(N_t)}{l(N_t - 1)} = \prod_{i=1}^m \frac{N_t}{N_t - u_i} (1 - \hat{P}_i) \quad (17)$$

for each possible integer larger than or equal to  $\max_i \{u_i\}$ . (Assuming the ratio becomes  $+\infty$  at  $\max_i \{u_i\}$ .) The last integer with which the ratio is larger than or equal to 1 is  $\hat{N}_t$ . Note that only (11) and (12) are needed in computing the ratios.

## VI. PERFORMANCE EVALUATION

In this section, we mainly report our simulation results to demonstrate the strength of our statistical approach for target counting in sensor-based surveillance systems. The wireless sensor network under our study contains  $N_s = 100$  sensors and  $N_t = 10, 50, 100, 500, 1000$  targets are present in the entire monitored area  $\mathcal{A}$ . Sensors are deployed over the region of  $100m * 100m$  in two different patterns: the grid pattern and random uniform pattern. For grid sensor deployment, the center of each  $10m * 10m$  grid cell has one sensor. For random uniform sensor deployment, each  $10m * 10m$  grid cell contains one sensor whose location is random within the grid. The radius of the sensing region of a sensor is set to be  $14.2m (> 10 * \sqrt{2}m)$ , to ensure that the region of  $100m * 100m$  is fully covered by the sensing regions of sensors even if the sensors are randomly and uniformly deployed. Note that sensing regions of neighboring sensors are heavily overlapped in our setting. For targets, we consider two scenarios: targets form one single cluster where only one distribution of the targets is available, and targets form multiple clusters where each cluster corresponds to one distribution of the targets within the cluster. Details about the simulation results are given in the following subsections.

Before presenting our simulation results on target counting, we make a note regarding numerical computation of the probability  $\hat{P}_i$  in (11). Clearly, to obtain  $\hat{P}_i$ , we only need to compute the integral  $\int \int_{A(s_j)} \hat{f}(x) dx_1 dx_2$  in (11). We do this by the traditional approximation procedure. Specifically, we divide the region  $A(s_j)$  into  $1m * 1m$  cells, then approximate the integral over each cell. If the center  $x^* = (x_1^*, x_2^*)$  of a cell lies inside  $A(s_j)$ , the integral over the cell is approximated by  $\hat{f}(x^*) * 1 * 1 = \hat{f}(x^*)$ . And if the center  $x^*$  is outside  $A(s_j)$ , the integral is simply approximated by 0. Then  $\int \int_{A(s_j)} \hat{f}(x) dx_1 dx_2$  is obtained by summing up these approximated integrals over the cells.

### A. Single Cluster

In this part, we report the target counting accuracy when targets form a single cluster over the monitored area. Both uniform and normal distributions of targets are considered. We assume that the nature (uniform or normal) of the distribution is known to the investigator *a priori* but the exact form of the normal distribution is unknown. Therefore, if the distribution of the targets needs to be estimated, the parametric approach in Section IV-A should be used. However, for a comparison,

TABLE I  
RELATIVE ERRORS FOR COUNTING UNIFORM TARGETS.

		$N_t = 10$	$N_t = 50$	$N_t = 100$	$N_t = 500$	$N_t = 1000$
Grid Sensors	No Estimation	-0.2%	-0.2%	-0.2%	0.0%	-0.1%
	Non-Parametric Estimation	0.1%	-0.3%	0.3%	0.0%	-0.1%
Random Uniform Sensors	No Estimation	-0.3%	-0.1%	-0.2%	-0.1%	-0.1%
	Non-Parametric Estimation	0.4%	0.3%	-0.1%	0.0%	-0.1%

TABLE II  
RELATIVE ERRORS FOR COUNTING NORMALLY DISTRIBUTED TARGETS.

		$N_t = 10$	$N_t = 50$	$N_t = 100$	$N_t = 500$	$N_t = 1000$
Grid Sensors	Parametric Estimation	0.5%	1.1%	1.2%	1.4%	1.1%
	Non-Parametric Estimation	3.0%	2.8%	3.2%	3.2%	3.0%
Random Uniform Sensors	Parametric Estimation	0.4%	1.0%	0.8%	1.2%	1.4%
	Non-Parametric Estimation	2.6%	2.5%	2.0%	2.7%	2.5%

we also report the results when the non-parametric approach in Section IV-B is used to estimate the distribution. To assess the accuracy of our approach in target counting, we run the simulation 200 times, and a final value of a performance metric is reported. Details are given below.

For uniformly distributed targets, no parametric estimation of the probability density function is needed in each simulation. Specifically, we first generate  $N_t$  (fixed number) targets that are uniformly distributed over the entire monitored area. Though parametric estimation of the distribution of the targets is not needed, we record the corresponding 100 pairs as listed in (5) and use Nadaraya-Watson estimator (8) to provide a non-parametric estimation of the distribution simply for a comparison. Then we randomly select  $m = 100$  largest non-overlapped subsets  $S_1, S_2, \dots, S_{100}$ , as described in Section V-A. Compute  $\hat{P}_i$  in (11) (with the true  $f$  replacing  $\hat{f}$  if no estimation is required). And compute  $u_i$  (the observed number of targets falling into the sensing regions of the sensors in the set  $S_i$ ) in (12). Finally, we use (17) to find the MLE  $\hat{N}_t$ . After repeating the above process for 200 times, we use the following *Relative Error (RE)* based on 200 runs as a performance metric to evaluate the counting accuracy of our approach:

$$RE = \frac{\sum_{i=1}^{200} \hat{N}_t^{(i)} - N_t}{N_t} \quad (18)$$

with  $\hat{N}_t^{(i)}$  denoting the MLE of the target count at the  $i$ th run.  $RE$  is the ratio of the difference between the averaged estimated count and the true count to the true count.  $RE$  can be positive (meaning overestimating) or negative (meaning underestimating). The simulation results are reported in Table I.

For normally distributed targets, we estimate the distribution and the target count in each simulation. Specifically, we first generate  $N_t$  targets according to a bivariate normal distribution  $\mathcal{N}(\mu_p, \Sigma)$ , where  $\mu_p$  is fixed at the center of the monitored area and  $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$  with  $\sigma_1$  and  $\sigma_2$  randomly chosen from  $[10, 20]$  and  $\rho$  randomly chosen from  $(-1, 1)$ . We choose the interval  $[10, 20]$  to assure that the generated targets are virtually within the monitored area. Then the corresponding 100 pairs as listed in (5) are recorded, so that

(6) from the parametric approach can be used to estimate parameters  $\mu_p$ ,  $\sigma_1$ ,  $\sigma_2$ , and  $\rho$ . For a comparison, Nadaraya-Watson estimator (8) from the non-parametric approach is also used to estimate the pdf of the normal distribution. And then we randomly select 100 largest non-overlapped subsets  $S_1, S_2, \dots, S_{100}$ , as described in Section V-A. Compute  $\hat{P}_i$  in (11) and  $u_i$  and use (17) to find the MLE  $\hat{N}_t$ . After repeating the above process for 200 times, the final values based on the performance metric  $RE$  in (18) is reported in Table II.

From Tables I and II we observe the following. First, the relative error under grid sensor deployment is similar to that under the random uniform sensor deployment. Though it could be anticipated, this observation indicates that our approach is robust against spatial noise in sensor positions. Second, for uniformly distributed targets, it is difficult to see if the relative error from no estimation of the distribution is better than that from non-parametric estimation. However, for normally distributed targets, the relative error from the parametric estimation of the target distribution is smaller than that resulting from the non-parametric estimation. This observation shows that the prior knowledge in the target distribution can help us better rebuild the distribution and thus better estimate the target count. The difference between parametric and non-parametric methods is that in parametric estimation, we only need to estimate the parameter(s) of the distribution, while in non-parametric estimation the 3-dimensional surface of the distribution has to be estimated.

## B. Multiple Clusters

In this subsection, we evaluate the performance of our counting method on multiple clusters of targets. Three cases are considered, where the distribution of targets over the entire monitored area  $\mathcal{A}$  is much more complicated than a single distribution. Case 1 (Figure 1(a)) involves piecewise uniformity of targets. Specifically, the monitored area is divided into four subareas with approximately the same size. Targets within each subarea is uniformly distributed, and the ratio of the counts of targets in four subareas (in the order of lower left, upper left, lower right, and upper right) is 1 : 2 : 3 : 4. Clearly targets form four clusters.

Case 2 (Figure 1(b)) contains two clusters of targets cor-

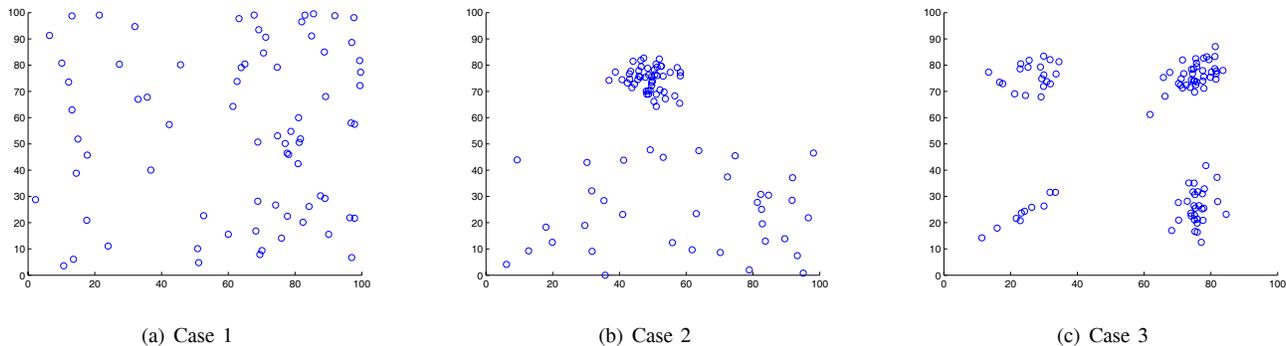


Fig. 1. Clustered Targets over the Monitored Area.

TABLE III  
RELATIVE ERRORS FOR COUNTING TARGETS IN MULTIPLE CLUSTERS.

		$N_t = 10$	$N_t = 50$	$N_t = 100$	$N_t = 500$	$N_t = 1000$
Grid Sensors	Case 1	-0.2%	-0.5%	0.1%	0.0%	0.2%
	Case 2	2.5%	1.8%	1.8%	2.4%	2.1%
	Case 3	1.2%	2.2%	2.0%	2.3%	2.4%
Random Uniform Sensors	Case 1	-1.0%	0.3%	-0.3%	0.1%	-0.1%
	Case 2	3.0%	3.4%	3.4%	4.0%	3.5%
	Case 3	0.9%	1.8%	1.9%	2.4%	2.2%

responding to one uniform distribution and one normal distribution. Specifically, the monitored area is divided into two halves with approximately the same size. Over the lower half targets are uniformly distributed. Over the upper half targets are distributed according to a normal distribution as described in Section VI-A but with some exception that now  $\mu$  is taken to be the center of the upper half and  $\sigma_1$  and  $\sigma_2$  are chosen from  $[3, 8]$ . (This interval  $[3, 8]$  is chosen so that all generated targets will virtually fall on the monitored area.) We assume the ratio of the counts of targets in two halves is 1 : 1.

Case 3 (Figure 1(c)) is similar to case 1 but deals with normal distributions. The monitored area is divided into four subareas with approximately the same size. Targets within each subarea has a normal distribution centered at the subarea. All four normal distributions have the same parameter setting as in case 2. The ratio of the counts of targets in four subareas (in the order of lower left, upper left, lower right, and upper right) is 1 : 2 : 3 : 4.

We assume no prior knowledge on any distributions of the targets in the above three cases. Therefore, (8) in the non-parametric procedure of Section IV-B is applied to estimate the target distribution over the entire monitored area. Again 100 largest non-overlapped subsets  $S_1, S_2, \dots, S_{100}$  are randomly selected for the use of the maximum likelihood estimation. And for comparison, both grid sensor and uniform sensor patterns are considered. The relative errors based on 200 runs of simulations are provided in Table III. A couple of observations follow immediately from the table. First, the results in the table show that the relative error under grid sensor deployment is similar to that under the random uniform sensor deployment. Second, for any specific distribution pattern of the targets, there is no much variation on the relative error as the number

of targets changes.

### C. Comparison with Other Algorithms

This section provides a comparative study where we compare our counting approach with the PC+ and PC- algorithms proposed in [14]. We use the same basic setting as in [14]:  $N_s = 100$  sensors are deployed uniformly and randomly over the  $100m * 100m$  region. The radius of the sensing region of a sensor is set to be  $7m$  and  $10m$ . (Note that due to use of such values of the radius, the monitored area may not cover the  $100m * 100m$  region entirely.) Two types of distributions of targets are considered: uniform distribution over the monitored area and normal distribution as used in Section VI-A. And the true number of targets  $N_t$  is set to be 10, 15, 20. We randomly select  $m = 100$  largest non-overlapped subsets  $S_1, S_2, \dots, S_{100}$  and follow the steps in Section VI-A to conduct target counting with our approach. The results based on 200 runs of simulations are shown in Figures 2(a), 2(b), 3(a), and 3(b). In the figures, NE, PA, and NPA represents our method when no estimation, parametric estimation, and non-parametric estimation of the target distribution is made, respectively.

From these figures, the following observations are immediate:

- 1) All four methods achieve similar results for uniformly distributed targets with the radius of the sensing region  $h = 7$ . However, for all the other three cases, the performance of PC+ and PC- is worse than that of our approach.
- 2) The performance of PC+ and PC- on uniformly distributed targets is much better than that on normally

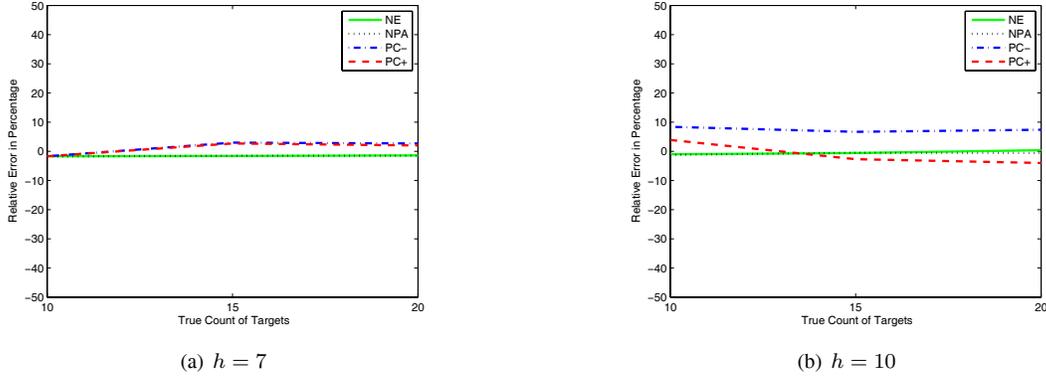


Fig. 2. Comparison Results for Uniform Distributed Targets.

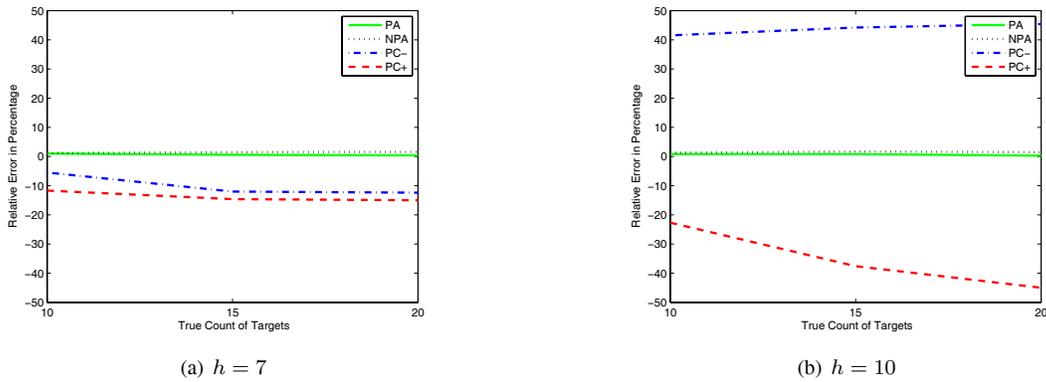


Fig. 3. Comparison Results for Normally Distributed Targets.

distributed targets. This difference is not significant for our counting approach.

Many other simulations have also been conducted to compare PC+ and PC- with our approach. We have found that the performance of PC+ and PC- is good and comparable to ours if targets are uniformly distributed and sensing regions of sensors slightly overlap. If the target distribution is not uniform, the performance of PC+ and PC- will degrade in general. If the sensing regions of sensors heavily overlap, the amount of time needed for PC+ and PC- to reach an estimate of target count would become huge and the estimated counts of targets by PC+ and PC- would look unrealistic. In addition, PC+ and PC- work slowly when  $N_t$  becomes large. These are the reasons that we do not include large values (e.g., 14) of the radius of sensing regions and large number of targets (e.g.,  $N_t = 50, 100$ , etc.) in the above comparison.

#### D. Further Discussion

In addition to the above simulations, we have also conducted many other simulations to evaluate the proposed counting approach. We summarize the major findings as follows.

1) Grid sensor deployment is almost equivalent to random uniform sensor deployment in terms of relative errors of the proposed counting approach. This finding is expected

and has been shown in Sections VI-A and VI-B.

- 2) In general, if the distribution of targets can be obtained through estimating the parameter(s), the parametric procedure should be used instead of using the non-parametric procedure to estimate the entire distribution. An example of this is provided in Section VI-A.
- 3) The proposed approach works for a small number of targets as well as a large number of targets. The relative error does not seem to change significantly as the number of targets changes. See Tables I II, and III for a demonstration.
- 4) For the use of the maximum likelihood estimation, one can randomly select sets of sensors  $S_1, S_2, \dots, S_m$ , each of which do not contain two sensors whose sensing regions are overlapping. However, for a fixed  $m$ , largest such sets should be used, since largest sets contain more information on targets.
- 5) For non-parametric estimation of the target distribution, both (8) and (10) can be used. However, in our simulations, we have found that in general, (8) actually works better. One reason for this is that in our setting (8) usually provides a better estimation of the target's distribution.
- 6) Our proposed target counting approach works best if

targets are spread out over the entire monitored area. If targets are clustered in a small area, a good way to estimate the target count is applying our approach to an appropriate small area that is larger than or equal to the union of the sensing regions each containing at least one target.

- 7) Our proposed approach in target counting is fast in practice. For example, in all the experiments in Sections VI-A VI-B and VI-C, it takes 1 ~ 3 seconds for each run of simulation.

## VII. CONCLUSION AND FUTURE WORK

This paper proposes a solid statistical approach for target counting in sensor-based surveillance systems. In this approach, regression techniques are first used to estimate the distribution of target positions for two cases. If the parametric form of the distribution is known, the parameter(s) is estimated by minimizing the residual sum of squares. If no prior information is available for the distribution, the kernel regression method is used to estimate the distribution. Using the estimated distribution of the targets, we can estimate the probability that a target falls into a specific subarea of the entire monitored area in the system. The estimated count of the targets is then obtained by the method of likelihood estimation based on a sequence of binomial distributions derived from the estimated distribution of targets and a sampling procedure.

A large number of simulations have been performed and the results show that our approach is effective in estimating the actual number of targets present in the surveillance system. Simulations also show that our counting approach is in general superior to the most recent algorithms in the literature in terms of execution time and counting accuracies.

Our future research would lie in the following directions: i) examining alternative procedures in estimating the target's distribution; ii) analyzing the effect of noise in sensors' reading on the estimated count of targets; iii) seeking possibly better metrics and methods to measure the performance in target counting; iv) investigating the effect of various bandwidths used in kernel regression on the counting accuracies; v) examining in details the computational complexity of the proposed approach; and vi) studying applications of target count estimation to real life detection problems.

## ACKNOWLEDGMENT

This work is partially supported by the NSF of the US under grants CNS-0964060, CNS-0963957, CNS-1017662, and CNS-0721669, the NSF of China under grant 61170267, Jiangsu Natural Science Foundation under grant BK2011358, the National Basic Research Program of China (973 Program) under grant 2011CB302905, and RFDP under grant 20113402120008.

## REFERENCES

- [1] M. Ding, D. Chen, K. Xing, and X. Cheng, "Localized fault-tolerant event boundary detection in sensor networks," in *IEEE INFOCOM 2005*, 2005, pp. 902–913.
- [2] A. Y. Teymorian, W. Cheng, L. Ma, X. Cheng, X. Lu, and Z. Lu, "3d underwater sensor network localization," *IEEE Transactions on Mobile Computing*, vol. 8, no. 12, pp. 1610–1621, 2009.
- [3] M. Ding and X. Cheng, "Fault-tolerant target tracking in sensor networks," in *Proceedings of the ACM MobiHoc*, 2009, pp. 125–134.
- [4] X. Wang, W. Gu, K. Schosek, S. Chellappan, and D. Xuan, "Lifetime optimization of sensor networks under physical attacks," in *Proc of IEEE International Conference on Communications (ICC)*, May 2005, pp. 3295–3301.
- [5] S. Ren, Q. Li, H. Wang, X. Chen, and X. Zhang, "Analyzing object detection quality under probabilistic coverage in sensor networks," in *Thirteenth International Workshop on Quality of Service (IWQoS)*, Passau, Germany, June 2005, pp. 107–122.
- [6] H. Wang, C. C. Tan, and Q. Li, "Snoogle: A search engine for physical world," in *IEEE Infocom*, Phoenix, AZ, April 2008, pp. 1382–1390.
- [7] S. Guo, T. He, M. Mokbel, J. Stankovic, and T. Abdelzaher, "On accurate and efficient statistical counting in sensor-based surveillance systems," *Pervasive and Mobile Computing*, vol. 6, no. 1, pp. 74 – 92, 2010.
- [8] S. Gandhi, R. Kumar, and S. Suri, "Target counting under minimal sensing: Complexity and approximations," in *Proceedings of the 4th international workshop Algosensors*, 2008, pp. 30–42.
- [9] Y. Baryshnikov and R. Ghrist, "Target enumeration via euler characteristic integrals," in *SIAM Journal on Applied Mathematics*, 2009.
- [10] M. Karakaya and H. Qi, "Target detection and counting using a progressive certainty map in distributed visual sensor networks," in *Proceedings of the 3rd ACM/IEEE International Conference on Distributed Smart Cameras*, 2009., 2009, pp. 1 –8.
- [11] W. Kim, K. Mechitov, J.-Y. Choi, and S. Ham, "On target tracking with binary proximity sensors," in *IPSN '05: Proceedings of the 4th international symposium on Information processing in sensor networks*, 2005, pp. 301–308.
- [12] N. Shrivastava, R. Mudumbai, U. Madhow, and S. Suri, "Target tracking with binary proximity sensors," *ACM Transactions On Sensor Networks*, vol. 5, no. 4, pp. 1–33, 2009.
- [13] J. Singh, U. Madhow, R. Kumar, S. Suri, and R. Cagley, "Tracking multiple targets using binary proximity sensors," in *Proceedings of the 6th international conference on Information processing in sensor networks*, 2007, pp. 529–538.
- [14] Y. Guo, B. Hua, and L. Yue, "Energy-based target numeration in wireless sensor networks," in *FGCN '07: Proceedings of the Future Generation Communication and Networking*, 2007, pp. 380–385.
- [15] Q. Fang, F. Zhao, and L. Guibas, "Lightweight sensing and communication protocols for target enumeration and aggregation," in *In Proceedings of the 4th ACM International Symposium on Mobile ad hoc networking and computing*, 2003, pp. 165–176.
- [16] D. Wu, X. Cheng, D. Chen, W. Cheng, B. Chen, and W. Zhao, "A monte carlo method for target counting," in *Proceedings of the 31th IEEE international conference on Distributed Computing Systems*, 2011, pp. 750–758.
- [17] B. Zhang, X. Cheng, N. Zhang, Y. Cui, Y. Li, and Q. Liang, "Sparse target counting and localization in sensor networks based on compressive sensing," in *IEEE INFOCOM*, 2010.
- [18] K. Kwong, R. Kavalier, R. Rajagopal, and P. Varaiya, "Real-time measurement of link vehicle count and travel time in a road network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 4, pp. 814–825, 2010.
- [19] T. Banerjee, K. R. Chowdhury, and D. P. Agrawal, "Using polynomial regression for data representation in wireless sensor networks," *International Journal of Communication Systems*, vol. 20, pp. 829–856, 2007.
- [20] X. Cheng, A. Thaler, G. Xue, and D. Chen, "TPS: A time-based positioning scheme for outdoor wireless sensor networks," in *INFOCOM*, vol. 4, March 7-11 2004, pp. 2685–2696.
- [21] W. Cheng, A. Y. Teymorian, L. Ma, X. Cheng, X. Lu, and Z. Lu, "Underwater localization in sparse 3d acoustic sensor networks," in *INFOCOM 2008*, Phoenix, AZ, April 2008, pp. 798–806.
- [22] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *The Quarterly of Applied Mathematics*, vol. 2, pp. 64–68, 1944.
- [23] A. W. Bowman and A. Azzalini, *Applied Smoothing Techniques for Data Analysis, The Kernel Approach with S-Plus Illustrations*, 1st ed. Oxford University Press, 1997.
- [24] P. J. Bickel and K. A. Doksum, *Mathematical Statistics Basic Ideas and Selected Topics, Volume 1*, 2nd ed. Prentice Hall, 2001.