FOCUS: A Cost-Effective Approach for Large-Scale Crop Monitoring with Sensor Networks

Yuan Yuan*†, Shanshan Li*, Kui Wu‡ Weijia Jia† and Yuxing Peng*

*Dept. of Computer Science, National University of Defense Technology, Changsha, China
†Dept. of Computer Science, City University of Hong Kong, Hong Kong SAR, China
‡Dept. of Computer Science, University of Victoria, B.C.Canada
Email: {yuyuan, wei.jia}@cityu.edu.hk, {ShanshanLi,Yuxing Peng}@nudt.edu.cn, WKui@cs.uvic.ca

Abstract

Current investment in crop monitoring consumes a large amount of financial cost, and how to reduce this cost has been a long-standing problem in agriculture. Traditional crop monitoring approaches are not cost-effective, because they rely on either heavy human labor or intensive computation with expensive instruments. In this paper, we explore the possibility of deploying networked sensor nodes for low-cost crop monitoring. As an example, we compute an important agricultural metric called global leaf area index (LAI) to illustrate the benefit of using sensor networks. We propose an approach called FOCUS that incrementally deploys sensor nodes into farmland to improve the accuracy of global LAI measurements. We design and implement a novel algorithm that calculates the total size of crop leaves with light intensity readings captured by the sensors under the crop canopies. FOCUS not only lowers the deployment cost considerably but also reduces the number of sensors for the long-term monitoring. Through a small-scale field test and large-scale simulations, we validate our design and show its effectiveness in crop monitoring.

1. Introduction

As reported by the U.S. Department of Agriculture, the national financial cost for agriculture rises continuously these years. It came up to a new record of 279.2 billion dollars in 2008, and counted for 75% of the gross income of agriculture. Crop monitoring is an integral part of agriculture and plays an important role for resources saving and yields increasing. Traditional crop monitoring uses machines or human resources to collect data from crop ecosystem to guide farmers in irrigation and fertilization. These methods are very costly and reducing the cost of crop monitoring has become an urgent problem in agriculture. With the availability of cheap sensor nodes and the progress of wireless technology, sensor networks have been widely deployed in many large-scale applications, such as environment monitoring [1] and surveillance [2]. In this paper, as shown in Figure 1, we explore the possibility of deploying networked sensor nodes in crop monitoring and use an important agricultural metric called global leaf area index (LAI) as an example to illustrate the benefit of sensor networks for cost reduction in agriculture.

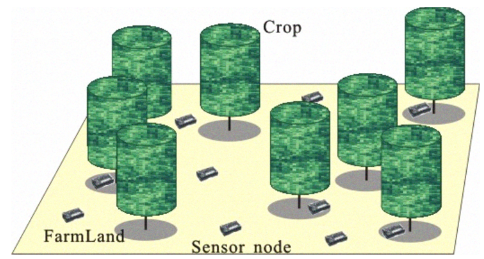


Figure 1. Crop monitoring with a sensor network

Global LAI is the most important metric in crop monitoring. *It is defined as the amount of upper surface area of leaves in crop canopy per unit land area* [3]. It can be used to assess various agricultural issues, such as canopy coverage, soil property, diseases, and yields. There are various approaches to estimating this metric. But most of them rely on either manual checking with prohibitive man-hours or intensive computation with expensive instruments, which impose unacceptable cost on farmers. In our work, we aim to utilize cheap sensor nodes to acquire accurate global LAI for the long-term and large-scale crop monitoring. In the rest of this paper, the term *global LAI* refers to the LAI in the whole farmland, while *local LAI* or simply *LAI*, refers to the LAI value at a specific local area.

Using sensor networks to obtain Global LAI poses several challenges. First, a sensor node may only capture light, temperature and humidity readings. How can it measure leaf area? Second, the seeds may be broadcast unevenly, and as such the branches and leaves from different grown crops interlace with each other in one place but may not in another. How can we compute the total area size of leaves in the farmland with irregular three-dimensional leaf distribution? Third, how to reduce the cost both for the deployment process and the long term monitoring? Finally, diverse weather and environment conditions may affect the accuracy of crop monitoring.

In this paper, we design a cost-effective approach named *F*ast *O*ptimal *C*ost red*U*cing deployment *S*cheme (FOCUS)

978-1-4244-5113-5/09/$25.00 ©2009 IEEE

that handles the aforementioned challenges. Based on Monte Carlo theory, FOCUS employs multiple iterative steps to randomly deploy sensor nodes into the farmland. In each step, according to the sensors' readings, FOCUS groups the sensor nodes to represent different canopy thickness, and decides the number of nodes needed in the next step to incrementally achieve certain precision requirement. After being deployed, the system collects the light intensity data regularly and uses a quasi-integral method to compute the global LAI of the whole land. Due to diverse weather conditions and changeable canopy figures, we consider multiple methods to enhance the accuracy and the reliability of FOCUS. We make a field test in a test field near Hunan Agricultural University. Using the readings from the test, we conduct a comprehensive simulation to validate our design in large farmlands. The results show that our approach achieves desirable performance in large-scale crop monitoring. The contributions of this work are as follows:

- To the best of our knowledge, we are the first to utilize sensor networks to compute global LAI in large farmlands.
- We propose FOCUS which only uses 2 or 3 iterative steps to incrementally deploy sensor nodes needed into the farmland to satisfy the global LAI precision requirement, and considerably reduces the cost both from the deployment process and the long-term crop monitoring.
- We design a novel quasi-integral algorithm to reduce the complexity of the global LAI computation. It utilizes the light intensity readings to describe the various thickness of the canopies and transforms the whole canopies into a frustum with uniform leaf density for the final LAI computation.
- We also consider the impact of changeable canopy figures and diverse weather conditions to enhance the accuracy and the reliability of FOCUS.

The rest of the paper is organized as follows. We survey the related work in Section 2. The models and the basic approaches are described in Section 3. In Section 4, we present our design of FOCUS. Section 5 introduces refinement methods for some practical issues. The performance evaluation is conducted in Section 6. We conclude the paper in Section 7.

2. Related Work

Existing LAI measurements employ either "direct" or "indirect" approaches. Direct approaches [3] harvest crops destructively or collect defoliations, and then measure the leaf area directly. These approaches may achieve the highest accuracy, but they can be performed only once in a crop's lifecycle and cost too many man-hours. Indirect approaches utilize optical properties to observe other metrics which can be translated into LAI. Remote sensing [4, 5] is the most well-known means which conduct reflection spectrum analysis on high resolution images captured by satellites. Although it can monitor large area for long term, the achievable accuracy and required computation are not desirable. Other indirect approaches based on gap fraction analysis have been successfully integrated into industrial instruments [6, 7], which measure LAI in limited area by using fisheye lenses or complex camera sensors. However, since the unit price of these instruments generally exceeds $2000, they cannot be widely deployed.

Recently, wireless sensor networks have been used in agriculture. Most of the work pays attention to the generic architecture of sensor networks in farmlands. For instance, Burel *et al.* propose a switchable architecture that can self-configure according to temporal factors [8]. During certain months, the proactive sensor system monitors the land. Other times of year, the system would use the data mule approach. Kabashi *et al.* introduce a capable decision support system [9], which improves agricultural practice. They study a zone-based joint topology control and power scheduling mechanism, and a multi-sink architecture. Hirafuji *et al.* develop a field monitoring architecture that includes access points embedded with large solar cells and ad-hoc sensor networks [10]. Their approach considerably reduces the energy consumption using sensors' sleep mode and solar cell's driving mode. Damas *et al.* use seven sensor nodes to develop a distributed, remotely controlled, automatic irrigation system in Spain [11]. Cugati *et al.* develop an automated fertilizer applicator for tree crops [12]. However, all of these work do not address the measurements of certain agricultural metrics and also provide no close guidance on specific farming issues. Some only utilize crop monitoring as a scenario to test their hardware or their networking protocols.

3. Models and Basic Approaches

In this section, we first present two models closely related to our work, including the light interception model and the crop canopy model. Then, we provide a brief description on the LAI computation and our basic ideas.

3.1. Light Interception Model

The light interception property of crop canopy can be described as follows: when a beam passes through the canopy, it will be partly intercepted by leaves and branches. The thicker the canopy, the less the light captured by the sensor nodes beneath. It has been proved in [13] that the interception property of crop canopy follows the famous model, Beer-Lambert law, which describes the quantitative relationship between the light intensity un-intercepted and the light intensity intercepted by canopy with certain thickness. The law is as follows:

$$Q_i = Q_0 e^{-kLAI_i} \tag{1}$$

Where Q_i and Q_0 are the vertical light intensity at height i in the canopy and that uncovered (i.e., on the top), respectively. k denotes the extinction coefficient, which means when light penetrates through each unit leaf area the rest of light intensity will be proportional to k. LAI_i is the local LAI from the top of the canopy to the height i at a fixed point in the farmland. The practical meaning of (1) is that if we use two sensors to monitor a crop, one on the top and the other under the canopy, we could estimate the local LAI value if the extinction coefficient is known. Using the local LAI computed from Equation (1), we can calculate global LAI. The calculation details will be introduced in Section 4.

In our work, we only consider the condition that the zenith angle[1] is $0°$, which means the captured light is vertical, and the Beer-Lambert law can be used directly. Other zenith angles can be extended by computing the vertical fraction of the light intensity. We assume that the light dispersion and reflection in canopy can be ignored.

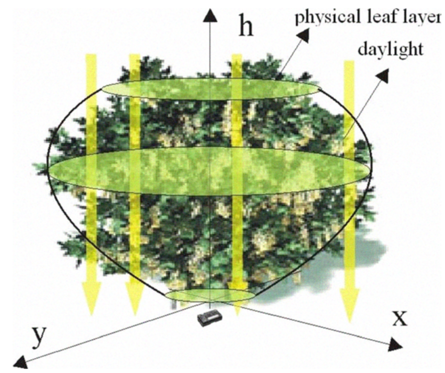


Figure 2. Canopy Model

3.2. Crop Canopy Model

The canopy structure model and the leaf distribution are essential for understanding the light interception property and our LAI computation as well. In botany, lots of literatures [14, 15] focus on these topics. Briefly, according to the height, the canopy of a single crop can be vertically classified into several *physical leaf layers* with a certain interval, such as $0.1m$. And those leaves which cross two layers would be clipped and put into the two layers respectively. From the experiments, the vertical distribution of the number of leaves can be approximated by Gaussian distribution, meaning that most leaves assemble at the middle of the plant. So a crop's canopy can be generally modeled as an irregular shuttle with a stem axis, as shown in Figure 2. In this model, layer j of crop i is defined by a 4-tuple $(p_i, h_{ij}, s_{ij}, F_{ij})$, where p_i stands for the coordinates of crop i, h_{ij} and s_{ij} denote the height and the total area size of leaves at layer j of crop i, respectively. F_{ij} is a set of functions which describe the shape of the vertical projection of layer j. It is well known that a beam more close to the axis needs to penetrate more physical leaf layers, hence the corresponding readings from the underneath sensors may be lower. We stress that our FOCUS does not need any specific information about the canopy structure in the target farmland, except the size of the farmland and the extinction coefficient of this certain type of crops.

3.3. The Objective

In this problem, the cost involves two aspects, one is the number of sensor nodes deployed which represents the long-term cost of crop monitoring. The other is the cost in deployment process. the reason that we choose random deployment scheme instead of grid-based scheme is to reduce the deployment cost. Grid-based deployment scheme (GDS) costs heavy human labor to locate each of the accurate grid points in a large farmland. In some cases, farmers use helicopters to deploy sensors, the same way as they use helicopter to sprinkle seeds. So in our work, it is reasonable to use the times of helicopter takeoffs to approximate the deployment cost.

In a large farmland with an area size G, a set of sensor nodes are randomly deployed on the ground for crop monitoring. They capture the light intensities under the canopies and transmit their readings to the base station for the global LAI computation. Our objective is to reduce the cost as much as possible but also to achieve the user's LAI precision requirement ε (defined later).

3.4. Basic Ideas of FOCUS

The main challenge of the global LAI computation is to count the total leaf area of the canopies with various leaf densities at different points of the field. We regard all the individual canopies as a whole called *global canopy* and define the important terminology *logical leaf layer* as a virtual layer of the global canopy. It is composed of leaves from different crops with the same thickness of leaves underneath, or, in other words, with the same number of physical leaf layers underneath. For example, as shown in Figure 3, the leaves at position A and B are at the same physical layer, but they belong to different logical leaf layers; the leaves at position A and C are at the different physical layers, but at the same logical layer, because there are many leaves overlapping with each other under position A, and the leaf area density [2] under position A is equal to that of position C, but much higher than that of position B. Based on this concept, we can find these logical leaf layers and reconstruct the global canopy model for the LAI computation.

The design of FOCUS is based on Monte Carlo theory [16], which guarantees the precision of using randomly selected points (sensor nodes) to measure the area size. we use R to stand for the ratio of the size of the area vertically projected by points at a particular logical leaf layer to the size of the farmland. We call R *the projection area ratio.*

1. zenith angle is the angle between directly overhead and a line through the sun.

2. The leaf area density is defined as the square meters of leaves per cubic meters [3].

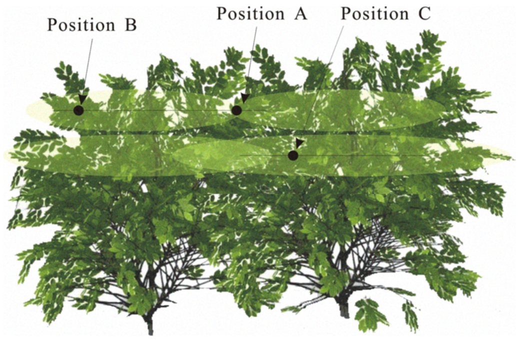


Figure 3. Logical leaf layer vs. physical leaf layer

According to [17], if we regard the R as a random variable, using the Central Limit Theorem and the Large Number Theorem, the area measurement accuracy is calculated as:

$$\varepsilon = \frac{c \cdot \sigma}{\sqrt{N}} = \frac{c \cdot \sqrt{(1-R) \cdot R}}{\sqrt{N}} \tag{2}$$

Where c denotes the confidential interval, σ is the standard deviation of the projection area distribution and N is the number of sensor nodes deployed in the farmland. Obviously, the more sensor nodes deployed, the higher accuracy we can achieve. However, to reduce the cost, we use a few iterative steps to gradually deploy sensor nodes into the farmland, until the precision requirement is achieved.

4. Design of FOCUS

In this section, we introduce the design of FOCUS in great detail. The outline is as follows: first, the number of sensor nodes needed is unknown at the beginning. We predict the initial number of nodes by using some empirical information. Second, in each deployment steps, we collect the light intensity readings, and group the sensor nodes by the extended Jaccard coefficients of these readings to find the logical leaf layers of the global canopy. Third, A heuristic rule is given to decide which logical layer's precision should be satisfied first, based on which we compute the extra number of sensor nodes needed in the next step. Finally, we propose a quasi-integral method to compute the global LAI.

4.1. Initial Prediction and Grouping Process

Before deployment, it is necessary to derive a general estimation on the number of sensor nodes needed in the farmland. In our work, we utilize some loose empirical information to do this estimation. We stress that the empirical information does not have to be very accurate, since we would use iterative steps to adjust the precision.

We first predict the projection area size of the global canopy through some probabilistic analysis as an approximation of area size of the largest logical layer. When seeds are sprinkled, they actually fall into the land in sequence. We use s to stand for projection area size of a single grown crop, which comes from empirical data. Suppose we could count the seeds one by one, we use A_i to denote the size of the projection area formed by up to i seeds, then:

$$\begin{aligned} A_1 &= s \\ A_2 &= s + (1 - s/G) \cdot s \\ &\dots \\ A_n &= A_{n-1} + (1 - (A_{n-1}/G)) \cdot s \end{aligned} \tag{3}$$

Transform (3) as follows:

$$A_k = ((G-s)/G)A_{k-1} + s \tag{4}$$

We set $\alpha = (G-s)/G$, then:

$$\begin{aligned} A_n &= \alpha A_{n-1} + s \\ &= \alpha(\alpha A_{n-2} + s) + s \\ &= \alpha^2 A_{n-2} + \alpha s + s \\ &= \dots \\ &= \alpha^{n-1} A_1 + s \cdot \sum_{i=0}^{n-2} \alpha^i \\ &= \alpha^{n-1} A_1 + s.\alpha(1-\alpha^n)/(1-\alpha), \end{aligned}$$

Assume that M seeds are sprinkled into the land and will grow into M crops. The expected projection area size of the global canopy can be estimated as $\alpha^{M-1}A_1 + s.\alpha(1-\alpha^M)/(1-\alpha)$. Then the projection area ratio, which also can be considered as the expected probability of a sensor node fallen into the projection in the farmland is:

$$R_{init} = \frac{\alpha^{M-1}A_1 + s.\alpha(1-\alpha^M)/(1-\alpha)}{G} \tag{5}$$

Apply (2) and (5), we can compute the number of nodes for the initial deployment. Suppose the number is N_{init}.

After N_{init} sensor nodes are randomly deployed into the farmland, FOCUS gathers N_{init} series of feedback readings for a certain time interval, then conducts the grouping process. In the grouping process, FOCUS first computes the Similarity Matrix $[r_{kl}]$.

$$\begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ & & \dots & \\ r_{n1} & r_{n2} & \dots & r_{nn} \end{bmatrix}$$

Where r_{kl} stands for the extended Jaccard coefficient of nodes k and l, and is defined as:

$$r_{kl} = \frac{X_k \cdot X_l}{\|X_k\|^2 + \|X_l\|^2 - X_k \cdot X_l} \tag{6}$$

In the above formula, we consider the readings of sensor nodes k and l as vectors X_k and X_l, respectively and use $X_k \cdot X_l$ to denote their dot product, and $\|X_k\|$ to denote the modules of X_k. The extended Jaccard coefficient takes values in the interval $[0, 1]$. Given a threshold θ, the similarity test succeeds when $\theta \geq 0$. Then we transform $[r_{kl}]$ into a 0-1 matrix, where 1 represents $r_{kl} \geq \theta$, and 0 is used otherwise. Actually, the Similarity Matrix $[r_{kl}]$ can be considered as an Adjacency Matrix of an undirected graph, whose nodes are the sensors and edges represent the data similarity between two sensors. Using approximation algorithm in [18] to solve the clique cover problem[3], we group the nodes in the adjacency matrix. Suppose that the result is denoted by $C_i = \{s_{i1}, s_{i2}, ..., s_{in}\}$, where i is the group number, and n is the number of nodes in group i. Therefore, FOCUS classifies the deployed sensor nodes into several groups. In each group, the readings of any two nodes are similar according to the threshold θ. Clearly, different groups represent different canopy thicknesses. In the perspective of the canopy structure, the sensors belonging to the same group have the same number of logical layers above them.

4.2. Incremental Deployment

After the initial deployment, the actual values of global LAI may not satisfy the precision requirement, since the projection area size predicted initially is based on some empirical information and only an approximation of the area size of one logical leaf layer. Intuitively, there are some naive schemes for the later steps. One is to deploy sensor nodes one by one till the precision requirement is satisfied. However, this scheme brings too many iterative steps. A better choice is to deploy a bigger fixed number of nodes in each step, like 20, 50. However, this may cause excessive cost of sensor nodes and some undesired deployment cost.

In FOCUS, We propose an iterative scheme to gradually improve the precision. In each later step, using the information supplied by the formal steps, we adaptively compute a certain number of complementary nodes, which can reduce the cost as much as possible.

To compute the number of complementary nodes, We first count the projection area ratio of each logical layer. Based on Monte Carlo theory, the projection area ratio can be accurately approximated by the ratio of the number of sensor nodes deployed in the projection of this logical layer to the total number of sensor nodes in the farmland. So the issue is transformed to calculate the number of sensor nodes at each logical layer. This calculation is based on the following observation: when a beam penetrates thicker canopies, it passes through more intermediate logical layers. *Statistically,* this means the logical layers have an embedding relationship, i.e., a logical layer i at which sensors have smaller light intensity readings is embedded in another logical layer $j (j > i)$ at which sensors' light intensity readings are higher. By "embedding", we mean that the sensor nodes at logical layer i are included in logical layer j. Suppose that the groups sorted by ascending order of their readings are $C_1, C_2, ..., C_n$. We use $|C_i|$ to denote the number of nodes in group i. According to the above reasoning, the number of nodes in logical layer i can be counted as:$N_i = \sum_{k=1}^{i} |C_k|$. So the projection area ratio of the ith logical layer R_i can be computed as follows:

$$R_i = \frac{N_i}{N_{total}} = \frac{\sum_{k=1}^{i} |C_k|}{\sum_{k=1}^{n} |C_k|} \tag{7}$$

Where N_{total} stands for the total number of nodes. From (7) and (2), we can easily compute the complementary number of nodes needed to meet the precision requirement of each logical layer.

However, there are still two challenges about which logical layer should be satisfied first and how many nodes we should deploy in the next step. We utilize an important feature of Monte Carlo theory that more nodes are needed to satisfy the precision requirement at a logical layer whose projection area ratio is closer to fifty percent [16], which brings us a heuristic rule to decide the extra number of nodes in the next step:

Theorem 1: In each iterative step of FOCUS, if we guarantee the number of nodes needed in a certain logical layer whose projection area ratio is the closest to fifty percent, then the precision requirement of other existing logical layers is also satisfied.

Proof: Suppose $\triangle_i$ denotes the difference between fifty percent to the projection area ratio of layer i, therefore, R_i can be represented as $0.5 \pm \triangle_i$. From (2), we obtain the precision formula:

$$\varepsilon_i = \frac{c\sqrt{((1-(0.5+\Delta_i))(0.5+\Delta_i)}}{\sqrt{N_{total}}} = \frac{c\sqrt{0.5^2 - |\Delta_i|^2}}{\sqrt{N_{total}}}$$

Since we assume that the projection area ratio of layer i is more adjacent to fifty percent than layer j, so $|\Delta_i| \leq |\Delta_j|$. Hence we get $\varepsilon_i > \varepsilon_j$, which ends the proof. □

By Theorem 1, FOCUS always first satisfies the precision requirement of the layer whose projection area ratio is the closest to fifty percent. Our simulation shows that Theorem 1 largely decreases the iterative steps and brings no excessive nodes.

4.3. Quasi-integral LAI Computation

The ultimate goal of our work is to compute the global LAI periodically. Based on Equation (1) and the explanation in Section 3, we can obtain the local LAI value at each sensor's location based on the sensor's light intensity readings. The challenge is that we cannot simply add up all local LAI values to obtain the global LAI value, because such calculation may count a local LAI value multiple times due to the overlap of crop leaves.

3. The clique cover is to use the minimum number of cliques to cover all vertices in the graph.

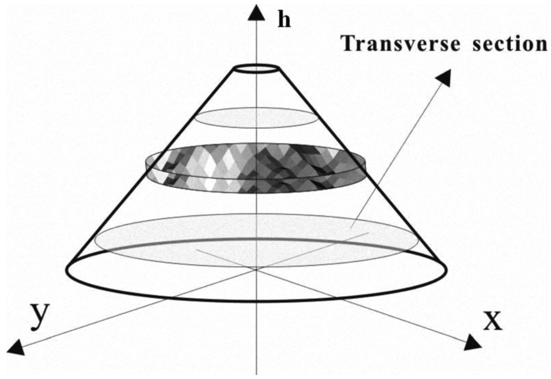


Figure 4. A cone type frustum built by the leaves of the whole farmland: each transverse section denotes a logical layer whose size is known

Thus far, we have formed logical layers based on sensors' light intensity readings after incremental deployment process. By the "embedding" relationship mentioned in Section 4.2 and Formula (7), the projection area size of logical layer i is included in the projection area of logical layer j. This relationship is determined by the way we calculate the projection area ratio of a logical layer based on Equation (7). With this embedding relationship, the leaves of the whole farmland can be built into a *virtual* cone type frustum to facilitate the global LAI calculation. As shown in Figure 4, the height of the virtual frustum is equal to the maximum local LAI values. Each logical layer can be imagined as a horizontal transverse section of the frustum. The size of each transverse section is equal to the projection area size of the logical layer, which has been obtained with the method in Section 4.2. In order to calculate the global LAI, we thus need to calculate the total size of leaves in the frustum.

To calculate the total size of leaves in the virtual frustum, we utilize the following important fact: The leaf area density in the virtual frustum is uniform. This is because the heights of these logical layers in the virtual frustum are proportional to their local LAIs. If h and a denote the variables of height in the virtual frustum and the leaf area density, respectively, we have the following integral to calculate the total size of leaves:

$$A_{total} = \int a \cdot dV = \int_0^H a \cdot A(h)dh \tag{8}$$

Where dV denotes the volume element of the integral, H is the total height of the frustum and $A(h)$ is the area function. To compute this integral, we have to get the three undefined parameters a, H and $A(h)$.

We first consider how to determine a and H. In fact, the Beer-Lambert law has its original form as:

$$T = \frac{Q_i}{Q_0} = e^{-klN} \tag{9}$$

Where l is the distance the light travels through the material and N is the density of absorbers. Comparing (9) with (1), the local LAI can be computed as the product of distance and density. So formula (8) can be transformed as:

$$A_{total} = \int_0^H A(h)d(a \cdot h) = \int_0^{LAI'} A(LAI_{layer})dLAI_{layer} \tag{10}$$

Where LAI_{layer} and LAI' stand for the integral variable and the integral upper bound of LAI, respectively. According to (1), LAI_{layer} and LAI' can be computed by the light intensity readings without a and H.

Formula (10) is in the form of continuous integral, while what we have found so far are the discrete logical layers. Therefore, we approximate Formula (10) with the volume cumulative formula of several discontinuous cylinders, that is,

$$\begin{aligned} A_{total} &= \sum_{i=1}^{n} A(LAI_i) \cdot \Delta LAI_i \\ &\approx \sum_{i=1}^{n-1} (LAI_i - LAI_{i+1}) \cdot A_i \\ &= \sum_{i=1}^{n-1} (LAI_i - LAI_{i+1}) \cdot R_i \cdot G, \end{aligned} \tag{11}$$

Where A_i and R_i $(i = 1, 2n - 1)$ express the size of the projection area and the projection area ratio of logical layer i, respectively, LAI_i denotes the average local LAI of layer i, and G denotes the size of the farmland.

Finally, we have the equation to calculate the global LAI, denoted as LAI_{global}, with the data that we have already obtained:

$$LAI_{global} = A_{total}/G = \sum_{i=1}^{n-1} (LAI_i - LAI_{i+1}) \cdot R_i \tag{12}$$

The quasi-integral method implies that the more logical layers found, the higher precisions achieved. However, the tradeoff between the cost and the precision and the limited accuracy of the sensor readings make the error inevitable. We evaluate the errors in Section 6.

5. Refinement

Weather conditions and changeable canopy figures may impact the global LAI results. Fortunately, the usage of the global LAI is to detect concerned events in a relatively large-time scale, e.g. diseases, pests, and coverage. The impacts of weather conditions and outliers are usually temporary. Nevertheless, we introduce some methods to reduce these impacts to further enhance accuracy and reliability as seen in a small time scale.

5.1. Non-Uniform Clouds Coverage

In cloudy days, sunlight may first be intercepted by the clouds and then penetrates the crops' canopies. Based on Beer-Lambert Law, the ratio Q_i/Q_0 can be regarded as a descriptor of the local LAI or canopy thickness. That's to say, if we utilize Q_i/Q_0 to group the sensor nodes, the condition that the whole farmland is covered by the clouds uniformly is equivalent to the condition that no clouds above. Note that Q_0 is measured by sensors above the canopies or in the gap between crops. To consider partial clouds coverage, it is not appropriate to use one reading sequence Q_0 to stand for the light intensity above the canopies in the whole farmland, since different places in the field may have different Q_0 values. Intuitively, to handle this problem, we can measure Q_0 for each node, but this would cost too much. The reasonable method is to use the readings captured by the sensor node at the nearest gap between crops to approximate Q_0 for each covered node.

5.2. Outliers

Due to some temporary node failures or incidents (e.g. some insect stays on the sensing board for a while), sensor nodes may momentarily capture outliers, whose values are much smaller or larger than the historical readings. In FOCUS, we take Median filter [19] as a component to cut unpredictable and temporary outliers. In our real implementation, given the sampling rate of the sensor as once per minute, each sensor node uses a FIFO buffer to cache the latest 60 readings. Before transmitting the data directly, the Median filter sorts them according to ascending order, then transmit the average value of $30th$ and $31st$ to the base station. By using Median filter, the temporary outliers will be replaced by Median values in half an hour. What's more, some concerned events (e.g. disease or pests) which last more than half an hour will be reflected to the base station. Another reason that we choose Median filter as our component is because it has lower time and space complexity, and it is very easy to implement.

5.3. Errors from Rains and Winds

Winds and rains also bring errors into the light intensity readings, since they may slightly change the leaves' positions and vary the foliage inclination angles. These two factors can hardly be modeled. In this case, FOCUS utilizes Kalman filter to control the impacts. Kalman filter is a recursive estimator and takes optimal estimates only from the previous time interval. Its low resource requirement inherently suits sensor nodes. In our work, it is easy to construct the measurement equation and the state equation from the light intensity readings. With Kalman filter equations flushed into the sensor nodes, the captured readings have been calibrated and the impact of rains and winds can be alleviated.

5.4. Changeable Canopy Figures

In crop's lifecycle, the canopy figure may change seasonally. From spouting to maturity and to death of the crops, the sensor nodes under the canopies may capture diverse light intensity readings. Since crop growth is a long time process, FOCUS can periodically check the validation of the layered model and slowly adapt to the new canopy figures. In some cases, FOCUS may re-conduct iterative steps to deploy complementary nodes into the farmland to improve the layered model. However, we stress that it's not necessary to reclaim the excessive nodes deployed in the farmland, because they may contribute higher precision to the final results.

6. Performance Evaluation

We first implement and test our approach in a small testing cotton field. To save cost, we conduct simulation for large farmlands based on the measurements from the field test.

6.1. Field Test

To better understand the light interception model and to validate our design, we conducted a field test in a $50m \times 50m$ testing cotton field near Hunan Agricultural University in the second week of September 2008. In the test, the network was consisted of 11 Crossbow Micaz motes with 10 XMTS310 sensing boards. One Mote without sensing board worked as a sink node. We tied one of the 10 sensor nodes on a wood shelf for the light intensity in the uncovered area as a reference (i.e.,Q_0 in Formula (1)). This shelf was above the canopy and was in the center of the field. We set the sampling rate as once per minute. The everyday test was operated from 11 am. to 15 pm., and the whole test lasted for one week. Figure 5 exhibits our field test.

The field test has two parts. One is to measure individual crops. We first randomly deployed the rest of the 9 sensor nodes in a single crop's projection on the ground and collected their readings for about 3 hours. As illustrated in Figure 6, the 10 light intensity sequences can be divided into 5 groups, and the highest values come from the node on the shelf. Based on these measurements, we can construct a model for the large-scale simulation. In Figure 5, we can also find that when the readings from the node on the shelf fluctuate, other readings also fluctuate with similar wave shapes. It proves that apparent correlations exist between different sensors' readings, since they all originate from the sunlight. That is the reason why we choose extended Jaccard coefficient to test the similarity.

The other is to measure the whole field. The 9 sensor nodes were randomly deployed into the whole field in multiple rounds to represent more nodes on the ground at one time[4]. In each round, the 9 nodes gathered readings to the base station for half an hour. This lasts about 3 hour to simulate a test bed of 54 nodes. To use this method,

4. This practice of course has less accurate results compared to using hundreds of sensors at the same time. But in this way, we can use much smaller system cost to obtain good approximation because the global LAI results are not very time sensitive as seen in the first part of the test.

Figure 5. Field Survey

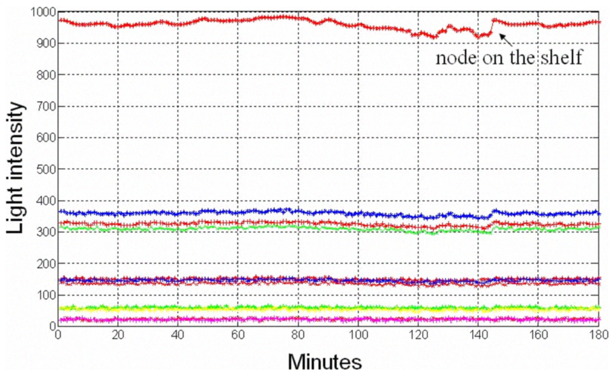


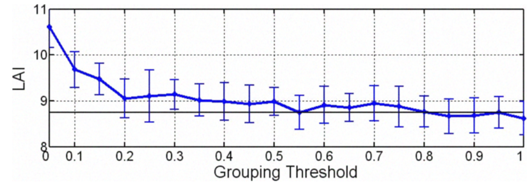
Figure 6. Light Intensity Readings captured in a single crop's Projection

we roughly verified the validity of FOCUS under lower LAI precision requirement and collect the raw data of light intensity for the later large-scale simulation. Our test results also confirm the validity of Beer-Lambert law with k around 0.5 for the tested cotton plants.

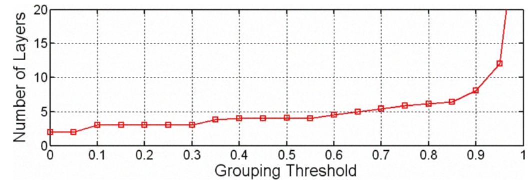
6.2. Large-Scale Simulation

We do large-scale simulation in a C-based simulator to evaluate FOCUS. We first generate a $1,000m \times 1,000m$ square farmland region with various number of crops on it. We utilize the canopy model mentioned in Section 3 to generate diversity crops, each of which has 8 discontinuous physical leaf layers but with random layer shape and random area size. Suppose that the vertical distribution of the area size of each physical layer in a single crop subjects to Gaussian distribution $N(0.6, 0.1)$. We record the large amount of data deliberately for the later computation. Suppose that the extinction coefficient k is set to 0.5. FOCUS is conducted to compute the global LAI in this farmland.

In the simulation, we first analyze how to choose the grouping threshold θ , then evaluate the performance using three metrics, which are the number of sensor nodes needed n_{node}, the number of deployment steps n_{step}, and the relative error of LAI computation $\hat{e}$, which is defined as $\hat{e} = e/LAI_{actual}$, where LAI_{actual} is the actual global LAI,



(a) LAI



(b) Number of layers

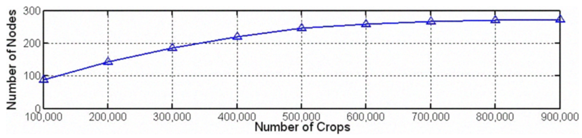
Figure 7. Grouping threshold analysis

which can be computed from crop canopy models mentioned above. And e denotes the error of LAI computation, which is defined as $e = |LAI_{actual} - LAI_{global}|$.

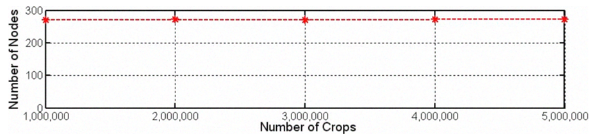
6.2.1. Grouping Threshold. To study the grouping threshold θ in FOCUS, we set the number of crops as a fixed value $3,000,000$, the precision requirement as 0.05, and change θ from 0 to 1. For each threshold, FOCUS runs 50 times to get the average results. As shown in Figure 7(a), when θ is small, the LAI computation results deviate largely from the actual global LAI (denoted by the horizontal straight line), because lots of the sensor nodes with distinct readings would be put into the same group. When θ becomes larger than 0.7, the results approach the requirement and the gain becomes trivial. Figure 7(b) plots the relationship between the number of logical layers and θ. With the threshold increasing, the number of layers increases steadily. However, the curve achieves a rapid ascend after 0.85, since at this point a slightly difference between two nodes' readings will lead them to different groups. So 0.7 to 0.8 might be good choices for θ considering the tradeoff between the accuracy and the grouping overhead. In the later simulation, we use 0.75 as a default grouping threshold for other evaluations.

6.2.2. Number of Nodes & Number of Steps Needed. Both n_{node} and n_{step} are evaluated under two varied parameters: the number of crops in the farmland and the precision requirement.

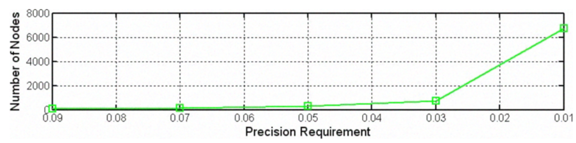
Figure 8(a) shows that with the number of crops increasing from 0.1 million to 0.9 million, n_{node} rises gradually, since the predicted projection area size of the global canopy becomes more close to fifty percent of the farmland. However, as shown in Figure 8(b), when the number of crops exceeds 0.9 million, n_{node} becomes stable round 270. Because after the coverage reaches a certain degree, FOCUS can always find a logical layer whose projection area ratio is very close to fifty percent. It proves that FOCUS is very scalable for large-scale crop monitoring, since the number



(a) Under small number of crops



(b) Under large number of crops



(c) Under varied precision requirements

Figure 8. Evaluation of the number of nodes needed

of nodes has its upper-bound under different crop densities. Figure 8(c) represents the relationship between the precision requirements and n_{node}. We can see that n_{node} increases rapidly after the precision requirement exceeds 0.03. When the precision requirement reaches 0.01, we need more than $7,000$ sensor nodes to monitor the farmland. So, 0.05 to 0.04 is a good choice for the precision requirement, considering the tradeoff between the cost and the reliability of the global LAI results.

Figure 9(a) and Figure 9(b) show the probability of the number of required steps under different number of crops and the precision requirements, respectively. No matter how these two parameters change, we have a high probability to use only 2 or 3 iterative steps to finish the deployment process in FOCUS. We compare FOCUS with the heuristic rule with random layer chosen scheme. In each deployment step, the later scheme randomly choose a logical layer to satisfy the precision requirement. The results are plotted in Figure 9(c). As we can see that by using FOCUS n_{step} is only about $1/3$ to that in the random layer scheme.

6.2.3. Relative Errors of global LAI. Given the rate of LAI computation is once per hour, we set the same number of sensor nodes deployed with FOCUS and grid-based deployment scheme(GDS). Figure 10(a) shows that FOCUS and GDS have similar relative errors of LAI results $\hat{e}$, which are below 0.06. However, GDS cannot achieve the exact number of nodes needed in such a few steps and it has to spend more human resources to locate the grid points for nodes deployment.

To evaluate the effects of outlier filter, several random outliers are injected in the simulation. Each outlier would last a random time but no more than half an hour. As shown in Figure 10(b), the relative errors of LAI $\hat{e}$ with outlier filter are much lower than that without outlier filter, since the unfiltered outliers affect the grouping process. Figure 10(c) illustrates the relative errors of LAI with or without Kalman filter. We use the readings and LAI measurements collected in field test to test the effects. We can see that the LAI precision values with Kalman filter will be slightly higher than that without.

7. Conclusion

In this paper, we present a novel approach called FOCUS for cost-effective, large-scale crop monitoring with sensor networks. In FOCUS, we use iterative steps to randomly deploy sensor nodes into a large farmland in order to save cost both for deployment process and long-term monitoring. We use a grouping process to transfer the original canopies with various thicknesses into a frustum with uniform leaf area density and compute global LAI by quasi-integral method. Both our field test and simulation show that FOCUS achieves desirable precision of global LAI measurements with a low cost.

Acknowledgment

The work is supported by the Research Grants Council of Hong Kong SAR, China No. (CityU 114908), by the CityU Applied R & D Funding (ARD-(Ctr)) No. 9681001 and 9678002, by the National Basic Research Program of China (973 Program) (2005CB321801), and by the National High-tech Research and Development Program of China (863 Program) (2006AA01Z332).

References

[1] G. Tolle, J. Polastre, R. Szewczyk, and D. Culler, “A macroscope in the redwoods,” in *Proceedings of SenSys*, November 2005.

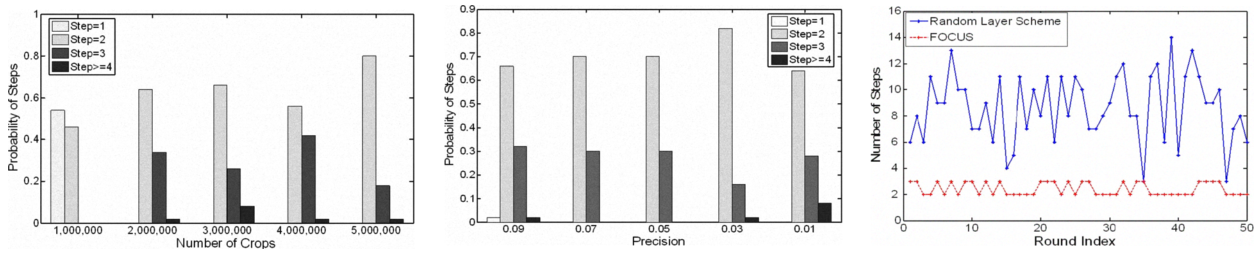
[2] S. Gandhi, S. Suri, and E. Welzl, “Catching elephants with mice: sparse sampling for monitoring sensor networks,” in *Proceedings of SenSys*, November 2007.

[3] I. Jonckheere, S. Fleck, K. Nackaerts, and B. Muys, “Methods for leaf area index determination,” *Agricultural and Forest Meteorology*, vol. 121, pp. 19–35, 2004.

[4] P. Gong, R. Pu, and G. S. Biging, “Estimate of forest leaf area index using vegetation indices derived from hyperion hyperspectral data,” *IEEE Transcation on Geoscience and Remote Sensing*, vol. 41, no. 6, pp. 1355–1362, June 2003.

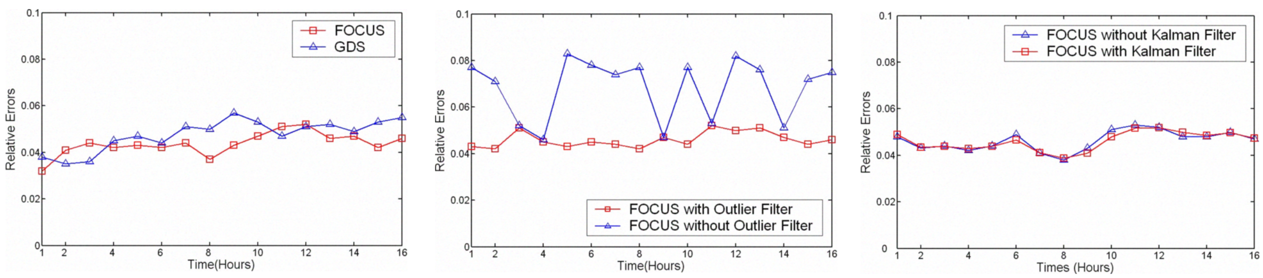
[5] R. Liu, J. Chen, J. Liu, F. Deng, and R. Sun, “Application of a new leaf area index algorithm to chinas landmass using modis data for carbon cycle research,” *Journal of Environmental Management*, vol. 85, no. 3, pp. 649–658, 2007.

[6] S. Martens, S. Ustin, and R. Rousseau, “Estimate of tree canopy leaf area index by gap fraction analysis,” *Forest Ecology and Management*, vol. 61, pp. 91–108, 1993.



(a) Under different number of crops (b) Under varied precision requirements (c) Comparison with random layer scheme

Figure 9. Evaluation of the number of steps needed



(a) Comparison with grid-based deployment scheme (b) FOCUS with and without outlier filter (c) FOCUS with and without Kalman filter

Figure 10. Evaluation of the relative errors

[7] M. Wijk and M. Williams, "Optical instruments for measuring leaf area index in low vegetation: application in arctic ecosystems," *Ecological Applications*, vol. 15, no. 4, pp. 1462–1470, 2005.

[8] J. Burrell, T. Brooke, and R. Beckwith, "Vineyard computing: sensor networks in agricultural production," *IEEE Journal of Pervasive Computing*, vol. 3, no. 1, pp. 38–45, Jan-Mar 2004.

[9] A. H. Kabashi and J. Elmirghani, "Full-wireless field monitoring server for advanced sensor-network," in *Proceedings of NSDR*, August 2008.

[10] M. Hirafuji, T. Fukatsu, and H. Haoming, "Full-wireless field monitoring server for advanced sensor-network," in *Proceedings of AFITA/WCCA Joint Congress on IT in Agriculture*, 2004.

[11] M. Damas, A. Prados, F.Gomez, and G. Olivares, "Hidrobus system: fieldbus for integrated management of extensive areas of irrigated land," *Microprocessors Microsyst*, vol. 25, no. 1, pp. 174–184, 2001.

[12] S. Cugati, W. Miller, and J. Schueller, "Automation concepts for the variable rate fertilizer applicator for tree farming," in *Proceedings of European Conference in Precision Agriculture*, june 2003.

[13] J. Ross, *Radiation Regime and Architecture of Plant Stands (Tasks for Vegetation Science)*. Dr W. Junk Publishers, 1981.

[14] N. Boedhram, T. Arkebauer, and W. Batchelor, "Season-long characterization of vertical distribution of leaf area in corn," *Agronomy Journal*, vol. 93, no. 4, pp. 1235–1242, 2001.

[15] J. Vose, N. Sullivan, B. Clinton, and P. Bolstad, "Vertical leaf area distribution, light transmittance, and application of the beer-lambert law in four mature hardwood stands in the southern appalachians," *Canadian Journal of Forest Research*, vol. 25, pp. 1036–1043, 1995.

[16] R. Kawai, "Adaptive monte carlo variance reduction with two-time-scale stochastic approximation," *Monte Carlo Methods and Applications*, vol. 13, no. 3, pp. 197–217, 2007.

[17] P. Wijayatunga and B. Cory, "Sample size reduction in monte carlo based use-of-system costing," in *Proceedings of IEEE International Conference on Advances in Power System Control, Operation and Management*, November 1991.

[18] C. Liu, K. Wu, and J. Pei, "An energy efficient data collection framework for wireless sensor networks by exploiting spatiotemporal correlations," *IEEE Transcation on Parallel and Distributed System*, vol. 18, no. 7, pp. 1036–1043, 2007.

[19] O. Yli-Harja, J. Astola, and Y. Neuvo, "Analysis of the properties of median and weighted median filters using threshold logic and stack filter representation," *IEEE Transcation on Signal Processing*, vol. 39, pp. 395–410, 1991.